

TABLE OF CONTENTS

ARTICLES

- 1 **Mission, Margin and Trust in the Nonprofit Health Care Enterprise**
Thomas L. Greaney, J.D. and Kathleen M. Boozang, J.D., LL.M.
- 89 **The Battle over Self-Insured Health Plans, or “One Good Loophole Deserves Another”**
Russell Korobkin, J.D.
- 137 **Breathing Life into the Framework Convention on Tobacco Control: Smoking Cessation and the Right to Health**
Benjamin Mason Meier, J.D., LL.M.
- 193 **Pharmaceutical Arbitrage: Balancing Access and Innovation in International Prescription Drug Markets**
Kevin Outterson, J.D., LL.M.

COMMENTARY

- 293 **Using Reciprocity To Motivate Organ Donations**
Marc S. Nadel, J.D. and Carolina A. Nadel, M.D.

BOOK REVIEW

- 327 **Sex, Fear, and Public Health Policy**
John G. Culhane, J.D.

CASE STUDY

- 341 **How can the current state of medical malpractice insurance be improved?**
- 343 **Responses by:**
Richard E. Anderson, M.D.; Geoff Boehm, J.D.; Donald J. Palmisano M.D., J.D.; and Kathryn Zeiler, M.S., J.D., Ph.D.

ANNIVERSARY ESSAYS

- 399 **Essays by:**
Lori B. Andrews, J.D. and Jordan Paradise, J.D.; Lawrence O. Gostin, J.D., LL.D.; Christine Grady, R.N., Ph.D.; Timothy Stoltzfus Jost, J.D.; Aaron S. Kesselheim, M.D., J.D. and Troyen A. Brennan, M.D., J.D., M.P.H.; Evan Lyon, M.D. and Paul Farmer, M.D., Ph.D.; Theodore R. Marmor, Ph.D. and Jacob S. Hacker, Ph.D.; and O. Carter Snead, J.D.

ARTICLES

Mission, Margin, and Trust in the Nonprofit Health Care Enterprise

Thomas L. Greaney, J.D.* and Kathleen M. Boozang, J.D., LL.M.†

INTRODUCTION

Lost in the recent flurry of legal activity occasioned by corporate integration, disintegration, and scandalous episodes of managerial abuse, the law governing charitable corporations remains neglected and thoroughly muddled. Still unsettled are central issues regarding the accountability of directors and management, legal standards governing organic changes by nonprofit institutions, and mechanisms to ensure fidelity to the organization's charitable mission. For nonprofit corporations in the health care sector, which represent a large proportion of all health services supplied nationwide, particularly charity care, these shortcomings have had serious repercussions.

The adaptation of for-profit corporate law to charitable corporations¹

* Professor of Law and Director, Center for Health Law Studies, Saint Louis University.

† Associate Dean for Academic Affairs and Professor of Law, Seton Hall University. We gratefully acknowledge the comments of James Fishman and Evelyn Brody, and the outstanding research assistance of Allana Holub, Patricia McManus, and Jeanie Montrey. Our Yale editors were superb, for which we thank them.

1. The structural hallmark of the nonprofit firm is the absence of owners, or shareholders, who share in its profits. Professor Henry Hansmann famously characterized the legal regime governing nonprofits as imposing a "nondistribution constraint," requiring nonprofits to reinvest net earnings in the entity and precluding any distribution among individuals who control the organization. Henry Hansmann, *The Role of the Nonprofit Enterprise*, 89 YALE L.J. 837, 840 (1980). As used in this article, "charitable corporations" are a

has been clumsy and ineffective at best; in its worst moments, it has proved perverse. Legal doctrine has never adequately addressed the accountability void that results from charitable corporations' lack of shareholders and market for corporate control.² Nor has it confronted squarely the *raison d'être* of nonprofits—that they exist not primarily to make money but to pursue charitable objectives. When dealing with transactions that implicate the nonprofit enterprise's purpose, such as conversions, closures, and abandonment of mission, courts and regulators are essentially left to their own devices. The law has failed to furnish guidance on the bedrock questions surrounding accountability and mission.

Confronted with ambiguous law governing oversight of the nonprofit enterprise, state attorneys general have resurrected charitable trust principles to facilitate more aggressive intervention in the managerial decisions of nonprofit boards. This activism by attorneys general, which predominantly focuses on hospitals and health insurers, addresses two broad categories of activities: alleged mismanagement by the nonprofit's board or its officers and organic changes that alter the status of the community hospital or nonprofit health plan. In both instances, the attorneys general quite properly serve as surrogate stakeholders for the societal and charitable interests inevitably implicated in such matters.³ Yet

subset of nonprofit corporations that have as their purpose charitable activities as required by the Internal Revenue Code, I.R.C. § 501(c)(3) (2004).

2. Academic accounts diverge sharply over whether nonprofit corporations can be thought of as having owners, and if so, who those owners are. *See, e.g.*, David M. Cutler & Jill R. Horwitz, *Converting Hospitals from Not-for-Profit to For-Profit Status: Why and What Effects?*, in *THE CHANGING HOSPITAL INDUSTRY* 45 (David M. Cutler ed., 2000) (asserting that the public does not own nonprofits); Jennifer Kuan, *The Phantom Profits of the Opera: Nonprofit Ownership in the Arts As a Make-Buy Decision*, 17 J.L. ECON. & ORG. 507, 517 (2001) (arguing that nonprofits have an owner—the board); Denise Lee Ping, Note, *The Business Judgment Rule: Should It Protect Nonprofit Directors?*, 103 COLUM. L. REV. 925, 931 (2003) (suggesting that nonprofits have no real owners); *see also* Lawrence Singer, *Realigning Catholic Healthcare: Bridging Legal and Church Control in a Consolidating Market*, 72 TUL. L. REV. 159, 162 (1997) (raising the question of whether a Catholic hospital is owned by the religious institute sponsor or the community being served).

3. Governmental enforcement actions against charities go back to fifteenth century England when the attorney general represented the Crown as *parens patriae*. NAT'L ASS'N OF ATTORNEYS GEN., STATE ATTORNEYS GENERAL: POWER AND RESPONSIBILITIES 184 (1990). In the United States, the authority of attorneys general to enforce charitable trusts was originally found in the common law; gradually, however, states enacted a variety of statutes that vested expanded powers in attorneys general to regulate charitable trusts and charitable corporations. *Id.* at 185; *see also* MARION R. FREMONT-SMITH, *GOVERNING NONPROFIT ORGANIZATIONS: FEDERAL LAW AND STATE REGULATION* 54-55 (2004).

their actions increasingly run squarely into two other important values: nonprofit managers' need for the autonomy, discretion, and flexibility essential to fulfilling their charitable missions; and the need to foster coordinated public policies governing the provision of safety net health care resources.

Our focus in this Article is on the legal oversight of the dominant species of nonprofit health care organizations as measured by revenues and public policy concerns: the "commercial" nonprofit corporation,⁴ specifically nonprofit hospitals and health plans. These nonprofit hospitals constitute a large proportion of the nation's hospital capacity,⁵ representing billions of dollars of charitable assets. As nonprofit health care enterprises also constitute a substantial percentage of the nation's nursing homes⁶ and comprise many of the nation's largest health insurers and managed care entities,⁷ these firms play a central role in providing much of the nation's safety net services; as a result, they take on added significance (and earn regulatory scrutiny).

The modern nonprofit health care enterprise faces a rapidly evolving

4. As Henry Hansmann's typology suggested some twenty years ago, the commercial nonprofit uniquely receives most of its funds from the sale of services with an expectation that it will return societal benefits in the form of charitable services or "community benefits" from its profits. See Henry B. Hansmann, *Reforming Nonprofit Corporation Law*, 129 U. PA. L. REV. 497 (1981) [hereinafter Hansmann, *Reforming Nonprofit Corporation Law*]. For more recent analysis, see HENRY HANSMANN ET AL., OWNERSHIP FORM AND TRAPPED CAPITAL IN THE HOSPITAL INDUSTRY (Yale Law & Econ. Research Paper No. 266, 2002), <http://ssrn.com/abstract=313827> [hereinafter HANSMANN ET AL., OWNERSHIP].

5. See Jill R. Horwitz, *Why We Need the Independent Sector: The Behavior, Law, and Ethics of Not-for-Profit Hospitals*, 50 UCLA L. REV. 1345, 1352 (2003) ("Of the nearly 2800 urban acute care hospitals, slightly fewer than 20 percent are government hospitals run by state, local, and federal governments, slightly fewer than 20 percent are for-profit hospitals, and the remainder are not-for-profit corporations.").

6. Approximately 28.6% of nursing homes are owned by not-for-profit corporations. See *id.*

7. See generally JACK NEEDLEMAN, NON-PROFIT TO FOR-PROFIT CONVERSIONS BY HOSPITALS AND HEALTH PLANS: A REVIEW (1996), <http://www.pioneerinstitute.org/research/whitepapers/wp5.cfm>. Dr. Needleman concludes that it is impossible to accurately estimate health plan conversions, which generally occur as changes in corporate form rather than acquisitions. *Id.* "Many of the converted HMOs have since merged with one another or with historically for-profit insurers. Six firms now dominate the national HMO market." *Id.* Importantly, Blue Cross and Blue Shield plans (the Blues), which were established during the depression to provide expansive hospital and physician coverage and were historically nonprofit in their orientation, changed their requirements in 1994 by eliminating the requirement that their licensees be organized as nonprofit corporations. *Id.*

economic and technological environment—as well as well-capitalized for-profit rivals.⁸ Owing to its charitable and tax-exempt status, it must also undergo close scrutiny from community and regulatory overseers. Some of the most controversial legal questions arise from hospitals' efforts to adapt to ensure their continued relevance and financial stability. Prominent examples include shifting acute to out-patient services,⁹ relocating or closing a hospital facility,¹⁰ affiliating with multi-state systems,¹¹ and joint venturing with for-profit entities or with religious groups that require changes in services.¹² Nonprofit health plans fit uncomfortably in this legal landscape—some now claim that they are not charitable entities, and indeed, abandoned their original “mission” decades ago. Congress recognized this when it began taxing health insurers, and the IRS generally resists according charitable status to HMOs.¹³ Nonetheless, attorneys general and other regulators have intervened aggressively in many instances in which health plans sought to convert to for-profit status.

Although in most states it is unquestionably the responsibility of attorneys general to ensure the preservation and appropriate disposition of charitable assets,¹⁴ we question whether in its current unsettled and ambiguous state, the law can adequately guide their actions. It is also questionable whether attorneys general have the resources or expertise to engage in the detailed assessments of the business and health policy issues surrounding the appropriate deployment of charitable assets that such decisions implicate. Frequently presented in a politically charged

8. Several factors contribute to the changing landscape of health care and the increasing need to compete with for-profits. With governmental regulation of the health field receding and market forces becoming dominant, medicine has taken on a primarily business (rather than service) orientation, and the line between the standards governing for-profit and nonprofit enterprises has blurred. David B. Starkweather, *Profit Making by Nonprofit Hospitals*, in NONPROFIT ORGANIZATIONS IN A MARKET ECONOMY 105 (David C. Hammack & Dennis R. Young eds., 1993).

9. See *Manhattan Eye, Ear & Throat Hosp. v. Spitzer*, 715 N.Y.S.2d 575 (Sup. Ct. 1999).

10. See *id.*; *Paterson v. Paterson Gen. Hosp.*, 235 A.2d 487 (N.J. Super. Ct. 1967). See generally N.H. DEP'T OF JUSTICE, NEW HAMPSHIRE ATTORNEY GENERAL'S REPORT ON OPTIMA HEALTH (1998), at <http://doj.nh.gov/publications/optimal.html>.

11. See *Banner Health Sys. v. Long*, 663 N.W.2d 242, 245-46 (S.D. 2003).

12. See *Nathan Littauer Hosp. Ass'n v. Spitzer*, 287 A.D.2d 202 (N.Y. App. Div. 2001). See generally N.H. DEP'T OF JUSTICE, *supra* note 10.

13. I.R.C. § 501(m) (2004); see also Evelyn Brody, *Whose Public? Parochialism and Paternalism in State Charity Law Enforcement*, 79 IND. L.J. 937, 1017 (2004) [hereinafter Brody, *Whose Public?*].

14. See *infra* note 277 and accompanying text.

atmosphere, these enforcement decisions may reflect policy judgments and preferences that go beyond the attorneys generals' competence or mandate.

The central issue addressed in this Article is how fidelity to the mission of the charitable health care corporation should be monitored. Part I sets the stage, providing a brief overview of the economic underpinnings of the regulation of nonprofit health care players. It surveys the economic literature, concluding that ownership form is not the decisive factor in the cost, quality, or efficiency of hospital services. However, the record is a mixed one and many benefits associated with the nonprofit sector are not readily quantified; others may be enhanced by a more supportive regulatory environment. Part II introduces some of the most notorious interventions by attorneys general in nonprofit health care and explains the legal means by which regulators attempt to accomplish their goals. Part III begins the analysis of the legal framework in which nonprofit governance is analyzed, finding corporate fiduciary law muddled and too permissive in its oversight of nonprofit corporate governance. Part IV turns to charitable trust law, which it concludes is doctrinally inapposite and pragmatically unsuited to govern business conduct in the contemporary health care market. The consequence has been to enable attorneys general and charitable enforcers to inappropriately stretch legal doctrine, thereby exacerbating confusion for nonprofit boards over the boundaries of their discretion and the role of charitable mission in decision-making. Finally, Part V offers guidance for the future direction of law and policy governing nonprofit health care firms. It advances the normative perspective that the law should maximize opportunities for nonprofits to fulfill their charitable missions, but should insist on more than nebulous assurances that society will receive tangible benefits. For nonprofit corporate doctrine, this Article proposes that nonprofit corporate law incorporate a principle of "mission primacy"—a doctrinal recognition that the nonprofit corporation's articulated charitable mission is its central objective. Further, nonprofit directors should enjoy a presumption of deference to define and, within limits, alter that mission to serve the public's interest and preserve the relevance and financial stability of the charitable entity. Judges and regulators should read mission-centered values into interpretations of the traditional fiduciary duties of care and loyalty. This approach should preserve managerial discretion to balance the various constituents of the nonprofit firm, including donors, consumers, and the community.

I. THE NATURE OF THE COMMERCIAL NONPROFIT ENTERPRISE IN HEALTH CARE

Before considering state regulation of the nonprofit health care enterprise as a charitable entity, it is essential to first understand the role these “commercial nonprofits” play in health care delivery and coverage in the United States. Theoretically, charitable corporations are mission-driven institutions established to benefit the communities that they ostensibly serve. By all relevant indicia, nonprofits in the health industry are significant, profit-seeking enterprises that compete vigorously (and, for the most part, successfully) against for-profit rivals. Empirical studies reach varying conclusions on the question of whether nonprofits in the health care industry satisfactorily fulfill their purpose of supplying essential public goods and substituting for government in maintaining the health services safety net.¹⁵ However, these appraisals exhibit strong normative disagreements about what society expects from nonprofits. Also unclear is whether the vast array of laws affecting nonprofit entities enhances the sector’s provision of benefits and accountability to the community or merely establishes minimal standards that encourage a “race to the bottom.”

This Part examines the economic underpinnings for the public policies and legal doctrine that govern the nonprofit health care sector. It first provides, as background, a brief overview and critique of the theoretical justifications for the existence of the nonprofit firm. We find in this account no grounds for confidence that the nonprofit sector will automatically supply promised public benefits. Next we examine the economic literature, which paints a decidedly mixed picture. The nonprofit form currently plays a modest role in helping the hospital sector to achieve the ends of cost, quality, and access, but appears to have little if any similar salutary role with respect to health plans. We caution, however, that historical evidence may not provide an accurate assessment of the potential of the nonprofit sector if, as suggested by our analysis of legal doctrine, those firms are not given sufficient flexibility or incentives to achieve those goals.

A. *Agency Cost, Trust, and Mission in Nonprofit Organizations*

The explanation of why nonprofit firms exist provides the foundation

15. To qualify for exempt status as a charitable 501(c)(3) organization, they must be operated “exclusively” for charitable or other exempt purposes. *See generally* St. David’s Health Care Sys. v. United States, 349 F.3d 232 (5th Cir. 2003).

for all discussions about their legal characteristics. In his seminal work, Henry Hansmann suggested that the prohibition on nonprofits disbursing their profits, denominated the “nondistribution constraint,” provides a mechanism for overcoming the significant information asymmetries in the services those firms provide.¹⁶ Hansmann claimed that the institutional commitment not to distribute profits to private parties helps overcome agency costs by inducing patrons (customers and donors) to trust nonprofits. The theory suggests that for “commercial nonprofits” like hospitals and third party payors, the constraint ameliorates consumers’ inability to accurately gauge the quality of services.¹⁷ The nondistribution constraint does double duty: It not only explains the existence of the nonprofit firm, but, in the words of Professor Evelyn Brody, it “keeps [them] honest, ensuring the dedication of assets and effort towards performing good deeds.”¹⁸ Consumers do not have to undertake the costly and perhaps impossible task of monitoring nonprofits’ delivery of services, thereby further reducing agency costs.

On closer examination, however, this rosy scenario collapses. First, multiple layers of informational and transaction cost problems are associated with the complex services provided by nonprofits. Even if the nondistribution constraint fosters trust, it does not solve the principal-agent problem between managers and directors of nonprofit firms. Board members of nonprofits are typically unpaid volunteers,¹⁹ many of whom are recruited for services other than providing supervision or assisting management.²⁰ Most students of nonprofit boards question their capacity to effectively supervise management.²¹

16. Hansmann, *Reforming Nonprofit Corporation Law*, *supra* note 4.

17. *Id.* at 505.

18. Evelyn Brody, *Agents Without Principals: The Economic Convergence of the Nonprofit and For-Profit Organizational Forms*, 40 N.Y.L. SCH. L. REV. 457, 459 (1996) [hereinafter Brody, *Agents Without Principals*].

19. See Cutler & Horwitz, *supra* note 2, at 63.

20. See Peggy Sasso, *Searching for Trust in the Not-for-Profit Boardroom: Looking Beyond the Duty of Obedience to Ensure Accountability*, 50 UCLA L. REV. 1485, 1539-40 (2003) (arguing that boards should include more insiders to increase trust between directors and management and to enable education of lay trustees who are generally not selected for their expertise in the nonprofit’s enterprise).

21. See Brody, *Agents Without Principals*, *supra* note 18, at 499-500 (summarizing Richard Heimovics & Robert D. Herman, *The Salient Management Skills: A Conceptual Framework for a Curriculum for Managers in Nonprofit Organizations*, 19 AM. REV. PUB. ADMIN. 295, 307-08, 309 n.13 (1989)) (“We were unprepared for the fact that both actors and observers in our research found the [nonprofit chief executive] as responsible for all nonprofit

In addition, the nonprofit firm justifies its existence by reference to a “mission” that includes subsidization of worthy causes with the proceeds from commercial sales. The nondistribution constraint cannot meliorate contract failure given management’s objective (indeed “mission”) to accomplish charitable goals through revenue shifting and its unsupervised discretion to do so; in short, despite nondistribution, the patron of the nonprofit firm has no assurance that the nonprofit will fulfill her aspirations.²² Finally, the Hansmann analysis leaves unanswered the question of how, given information asymmetry, consumers can distinguish one nonprofit from another. In the end, market failure cannot by itself explain the continued existence of the nonprofit hospital.²³ A more plausible account may be found in the complex agency arrangements that pervade health-purchasing decisions. First, health care decisions are the product of multi-tiered agency relationships. Consumers’ “choice” of hospitals is strongly influenced by intermediaries, namely their physicians and insurance plans. In turn, employers typically select health plans.²⁴ At each stage of the decision-making process, agents are operating with highly imperfect information about the services they are selecting and about the preferences of their principals (the patient/consumer).²⁵

Physician intermediaries may have multiple reasons for preferring nonprofit hospitals, including their own autonomy and self-interests as well as quality of care considerations peculiarly within their expertise.²⁶ To the

organizational outcomes, both successes and failures.”).

22. See Brody, *Agents Without Principals*, *supra* note 18, at 508-09 (“No matter how meritorious the cross-subsidization, how can a donor or patron be sure that her money is being used to provide the service that *she* wants? This pattern illustrates that the nondistribution constraint, while perhaps helpful, is not a sufficient bond to align the interests of management with the interests of patrons.”).

23. Hansmann conceded as much in later writings, contending that information asymmetry with a “lag effect” caused nonprofit hospitals’ predominance. See HENRY HANSMANN, *THE OWNERSHIP OF ENTERPRISE* 236 (1996). Nonprofit hospitals gained an initial foothold as donative institutions prior to the advent of widespread private insurance and public payment programs. Hansmann argues that “forces of inertia” have kept consumers from switching to for-profits despite their superior efficiency. *Id.*

24. Catherine Hoffman et al., *Holes in the Insurance System—Who Lacks Coverage and Why*, 32 J.L. MED. & ETHICS 390, 391 (2004).

25. Thomas L. Greaney, *Night Landings on an Aircraft Carrier: Hospital Mergers and Antitrust Law*, 23 AM. J.L. & MED. 191, 203 (1997) (describing market imperfections and agency relationships in health care services).

26. See generally MARK V. PAULY, *DOCTORS AND THEIR WORKSHOPS: ECONOMIC MODELS OF PHYSICIAN BEHAVIOR* (1980); Jerry Cromwell, *Barriers to Achieving a Cost-Effective Workforce Mix:*

extent that doctors prefer hospitals for selfish reasons, agency failure is the root cause for the steady predominance of the nonprofit form. This explanation is obviously inconsistent with an efficient market and militates against public policies and legal doctrines that favor the form. If, on the other hand, physicians' election to affiliate with and steer patients to nonprofit institutions is an exercise of professional judgment that helps overcome their patients' information deficits as to quality and other salient non-price factors, the nonprofit form is efficiency-enhancing and should be encouraged. Unfortunately, empirical evidence is lacking as to which scenario most plausibly explains physicians' hospital preferences.

B. Economic Analyses of the Nonprofit Enterprise in the Health Care Industry

Few contemporary hospitals and virtually no nonprofit health plans reflect the popular image of a charity—an institution selflessly dedicated to all comers, irrespective of ability to pay. Quantitatively measured solely in terms of providing health services to the poor,²⁷ hospitals offer at best marginal returns to society on its “investment,” while nonprofit payors offer negligible direct subsidies to the needy and only slight benefits

Lessons from Anesthesiology, 24 J. HEALTH POL. POL'Y & L. 1331, 1354 (1999) (claiming hospitals remain, as much as ever, “doctors' workshops”).

27. Scholars and public policy makers disagree about what comprises the community benefit that should be uniquely contributed by nonprofit hospitals. Uncompensated care is frequently cited because it is presumably measurable. In 2001, acute-care hospitals spent \$21.5 billion on uncompensated care, or six percent of total expenses, which is the lowest percentage recorded since 1983. Patrick Reilly, *Charitable Dropoff: Uncompensated Care Drops to Lowest Level in Years*, MOD. HEALTHCARE, Feb. 17, 2003, at 4. However, an exclusive focus on uncompensated care discounts the important value of the maintenance of “loss leader” services, community education, and research. Further, controversies and data collection problems surround the issue of defining and calculating the amount of uncompensated care provided by nonprofits. Charity care rendered is not synonymous with accounting measures such as bad debt. In addition, calculations must include offset for payments received from government sources and other forms of support received. Comparisons across sectors require resolving the role to be afforded tax payments by for-profits. A public good framework would reflect uncompensated care, uncompensated community services, medical research, and taxes, and potentially includes federal health plan shortfalls, price discounts on private pay patients, and losses on medical education. Sean Nicholson et al., *Measuring Community Benefits Provided by For-Profit and Nonprofit Hospitals*, 19 HEALTH AFF. 168, 169 (2000); see also Jill A. Marsteller et al., *Nonprofit Conversion: Theory, Evidence, and State Policy Options*, 33 HEALTH AFF. 1495, 1523 (1998); Ramesh K. Shukla, et al., *A Comparative Analysis of Revenue and Cost-Management Strategies of Not-for-Profit and For-Profit Hospitals*, 42 HOSP. & HEALTH SERVS. ADMIN. 117, 131 (1997).

through their rating and underwriting practices. But appreciating the impact of the nonprofit health care sector under the current legal regime requires an examination of both nonquantifiable elements of the safety net and the societal framework within which nonprofits operate. As Jill Horwitz put it, besides “function[ing] as safety nets where government fails[,] [nonprofit hospitals] provide avenues of civic participation that generate social capital, and allow for the expression and promotion of diverse values or world views that sustain democracy.”²⁸ Additionally, economic studies reveal the chameleon-like character of nonprofit organizations: Their performance is strongly influenced by the degree to which they compete with for-profit counterparts and by the regulatory and payment environment in which they operate.

1. Hospitals

The economic literature concerning the nonprofit hospital sector is vast and in some respects indeterminate. One cannot confidently conclude that the nonprofit form does or does not “make a difference” in terms of its net “payback” for tax exemption and other benefits it enjoys. At the same time, a close examination of these studies reveals intriguing patterns that can guide legal and policy analysis. Moreover, uncertainty about performance of nonprofits is itself an important finding that should inform doctrinal analysis.

To start with the bottom line, measures of price,²⁹ cost,³⁰ profit

28. Horwitz, *supra* note 5, at 1350 (footnotes omitted).

29. Older studies pretty consistently showed that for-profits charged their patients more. See, e.g., Marsteller et al., *supra* note 27, at 1503. One recent study, focusing exclusively on Medicare data, found that in 1989, 1992, and 1995, “per capita Medicare spending in areas served by for-profit hospitals was higher than in areas served by not-for-profit hospitals.” Elaine M. Silverman et al., *The Association Between For-Profit Hospital Ownership and Increased Medicare Spending*, 341 NEW ENG. J. MED. 420, 424 (1999). Specifically, the study found that spending growth increased after conversion to for-profit status. *Id.* at 423. Many explanations are offered for why for-profits charge more, including price gouging, greater costs, and the economic disadvantage of for-profits’ obligation to pay taxes. Uwe E. Reinhardt, *The Economics of For-Profit and Not-for-Profit Hospitals*, 18 HEALTH AFF. 178, 183 (2000); see Shukla et al., *supra* note 27, at 129 (suggesting that only about thirty percent of for-profits’ higher costs can be attributed to higher taxes). The most recent data on hospital pricing is mixed, suggesting that pricing is more sensitive to market factors. See, e.g., Cutler & Horwitz, *supra* note 2, at 71.

30. Older studies consistently showed for-profit expenses per day or admission to be greater than nonprofits. See Marsteller et al., *supra* note 27, at 1506. One study using 1993 data from Virginia hospitals found that for-profits’ revenue margins were attributable to

margin,³¹ efficiency,³² quality,³³ and access give modest support to the claim

pricing strategies rather than cost savings: “[For-profit] hospitals charged 24.8 percent more for outpatient procedures and 28 percent more for inpatient procedures.” Shukla et al., *supra* note 27, at 128. For-profit hospitals’ administrative costs in 1994 averaged twenty-three percent more than those of nonprofit hospitals, and thirty-four percent more than those of public hospitals. Steffie Woolhandler & David U. Himmelstein, *Costs of Care and Administration at For-Profit and Other Hospitals in the United States*, 336 NEW ENG. J. MED. 769, 772 (1997). In 1994, both the per discharge and day in-patient costs were higher in for-profit than either not-for-profit or public hospitals, despite the lower wage and salary costs in for-profit facilities. See Woolhandler & Himmelstein, *supra*, at 772. Cutler and Horwitz have questioned the extent to which the accuracy of for-profit cost reports has been affected by behavior such as that engaged in by Columbia/HCA, which consistently overestimated costs to Medicare. Cutler & Horwitz, *supra* note 2, at 64.

31. For-profits unquestionably generate a healthier profit margin than other hospitals, hovering around nine percent, while not-for-profit margins come in at around four percent with public hospitals falling in slightly behind. Richard G. Frank & David S. Salkever, *Market Forces, Diversification of Activity, and the Mission of Not-for-Profit Hospitals*, in THE CHANGING HOSPITAL INDUSTRY, *supra* note 2, at 198. But see James B. Rebitzer, *Comments on Chapters 1 and 2*, in THE CHANGING HOSPITAL INDUSTRY, *supra* note 2, at 87 (citing data from Tennessee that conversion did not improve profitability). Cutler and Horwitz suggest that one of the primary reasons for-profits more successfully generate revenue is because they more effectively game the loopholes in Medicare reimbursement. Cutler & Horwitz, *supra* note 2, at 64. They further found, however, that nonprofits in the same market, after discovering the billing practices of the for-profit, soon followed suit. *Id.*

32. In more competitive regions, for-profit hospitals increase investment in hotel, as opposed to clinical services, presumably focusing on the aspects of care observable to patients. By comparison, nonprofit investment in clinical services increases with increased competition. DANA B. MUKAMEL ET AL., BMC HEALTH SERVICES RESEARCH, HOSPITAL COMPETITION, RESOURCE ALLOCATION AND QUALITY OF CARE 58 (2002), <http://www.pubmedcentral.gov/articlerender.fcgi?tool=pubmed&pubmedid=12052>; PAUL GERTLER & JENNIFER KUAN, ARE NONPROFITS EFFICIENT? A TEST USING HOSPITAL MARKET VALUES (SSRN Elec. Library, Working Paper No. 323922, 2002), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=323922. Nonprofit and government entities are perceived as having no or diffuse owners, which results in weak governance with ill-defined, or at least not wealth maximizing, goals. *Id.*

33. Studies from the 1990s suggest that nonprofits perform more favorably than for-profits on many of the benchmarks of quality. One study focusing on quality of care in Utah and Colorado hospitals, as measured by the occurrence of preventable adverse events, found a lower frequency of these events at nonprofit hospitals as compared with for-profit hospitals and minor teaching or non-teaching public hospitals. Eric J. Thomas et al., *Hospital Ownership and Preventable Adverse Events*, 15 J. GEN. INTERNAL MED. 211, 215 (2000). A recent meta-analysis comparing mortality rates of for-profit and not-for-profit hospitals concluded that for-profits are “associated with a statistically significant increase in the risk of

that nonprofit hospitals historically have returned benefits to society. While some suggest that broader conceptions of “community benefit” (that include charity care, bad debt, losses from community programs, teaching, and research) yield convincing evidence that nonprofit hospitals contribute significantly more benefits than the cost of their tax exemption,³⁴ others observe that for-profit hospitals’ “contribution” to society is at least as great when one counts their tax payments as a community benefit.³⁵ Evidence further suggests that characteristics of the local market, such as the presence of other hospitals, managed care penetration, and socio-economic status of the community, are far more

death.” P.J. Devereaux et al., *A Systematic Review and Meta-Analysis of Studies Comparing Mortality Rates of Private For-Profit and Private Not-For-Profit Hospitals*, 166 CAN. MED. ASS’N J. 1399, 1402 (2002). The authors suggest that their results may underestimate the relative rate of mortality in for-profit facilities because of a possibility that nonprofits serve patients with greater disease severity, and that for-profits serve a greater proportion of private pay patients. *Id.* at 1404. Further, even if not-for-profits do set the bar in a market for quality, for-profits co-existing in the same market will be compelled to meet that bar, at least with respect to aspects of quality that are measurable and marketable. However, more sophisticated analysis suggests a more positive outcome for the for-profit entity:

On average, we find that for-profit hospitals have higher mortality among elderly patients with heart disease, and that this difference has grown over the last decade. *However, much of the difference appears to be associated with the location of for-profit hospitals: When we compare hospital quality within specific markets, for-profit ownership appears, if anything, to be associated with better quality care.* Moreover, the small average difference in mortality between for-profit and not-for-profit hospitals masks an enormous amount of variation in mortality within each of these ownership types. Overall, these results suggest that factors other than for-profit status per se may be the main determinants of quality of care in hospitals.

Mark McClellan & Douglas Staiger, *Comparing Hospital Quality at For-Profit and Not-for-Profit Hospitals*, in *THE CHANGING HOSPITAL INDUSTRY*, *supra* note 2, at 93, 94-95 (emphasis added). This outcome may be explained by the fact that higher quality hospitals tend to attract more difficult cases. *Id.* at 96. McClellan and Staiger confirmed others’ findings that higher volume hospitals tended to have lower mortality rates. *Id.* at 100. McClellan and Staiger further hypothesize that for-profit hospitals might be attracted to markets with lower quality care if low quality is a signal of poor management, making the hospital an attractive takeover target. *Id.* at 110.

34. Gary Claxton et al., *Public Policy Issues in Nonprofit Conversions*, 16 HEALTH AFF. 9, 18 (1997) (summarizing over twenty studies and concluding “the evidence indicates that there is a substantial difference between nonprofit and for-profit hospitals in terms of the [broadly defined] community benefits they provide.”).

35. *See id.* at 18; *see also* Jack Needleman, *The Role of Nonprofits in Healthcare*, 26 J. HEALTH POL. POL’Y & L. 1113, 1122-130 (2001) (summarizing the literature comparing for-profit and nonprofit hospitals).

powerful predictors of performance than the nonprofit form. Nevertheless, there can be little question that the nonprofit sector contributes to society free care and other measurable community benefits. Whether these benefits are less than or greater than the sum of societal expenditures (via foregone taxes, volunteer labor and other sources) remains a hotly disputed question.³⁶

This empirical record must be approached with caution, however. Most importantly, the economic literature does not enable one to draw conclusions about a “but for” world, i.e., one without nonprofit hospitals. A number of studies have attempted to compare performance between for-profits and nonprofits, finding generally that for-profits provide considerable charity care, perhaps approaching that of nonprofits, though certainly not at the level provided by government hospitals or academic medical centers.³⁷ Notably, for the most part these studies do not account for the dynamics that drive both sectors. Left unanswered are questions as to whether for-profits would be more or less willing to offer charity care in the absence of nonprofits in their markets, and whether nonprofits would generally adopt more aggressive pricing policies in response to competitive pressures of their counterparts.³⁸ The few studies that do tackle the issue depict a highly interactive relationship.³⁹

Furthermore, these studies cannot inform us about the potential of nonprofit firms to fulfill their goals if legal and regulatory constraints were removed. Indeed, across a number of characteristics and behavior, nonprofit status does appear to have significance in ways highly relevant to public policy analysis. Most importantly, ownership form correlates with market entry and exit with product line. Studies show that for-profits tend to locate in more affluent areas;⁴⁰ are quicker to enter new markets;⁴¹ and

36. See Robert C. Clark, *Does the Nonprofit Form Fit the Hospital Industry?*, 93 HARV. L. REV. 1417, 1434 (1980) (questioning whether nonprofits provide societal benefits commensurate with the advantages offered to them).

37. Many nonprofit to for-profit conversion transactions involve contract provisions requiring maintenance of current levels of charity care for a fixed period of time. Only time will tell whether the expiration of these contract requirements will affect for-profits' provision of uncompensated care.

38. Horwitz, *supra* note 5, at 1361-62.

39. See Cutler & Horwitz, *supra* note 2, at 71-73 (citing studies that depict the highly interactive relationship); Horwitz, *supra* note 5, at 1361 (hypothesizing that “for-profit hospitals often move first in markets and that not-for-profit and governmental hospitals copy the behavior of for-profit hospitals.”).

40. H. Shelton Brown, *Income, Location, and the Demand for Health Care from Public, Nonprofit, and For-Profit Hospitals*, 27 J. HEALTH CARE FIN. 24, 24 (2001).

more readily exit if the community experiences economic deterioration.⁴² Growing evidence also suggests that for-profit and not-for-profit hospitals diverge along product market dimensions, with not-for-profit hospitals more likely to offer unprofitable services⁴³ and less inclined to drop services.⁴⁴

2. *Nonprofit Health Plans*

There is also a large literature analyzing differences between nonprofit and for-profit health plans. Deriving generalized conclusions from these studies is difficult because much depends on how one defines “community benefits” and “health plans.” However, as a general matter, they illustrate some significant differences between nonprofit and for-profit HMOs in the extent to which they provide broadly-defined community benefits. As to health insurers, there is little evidence that the nonprofit form makes a positive difference for the communities in which they operate.

As with the hospital sector, the meaning of “community benefits” for health plans lies in the eye of the beholder. Health insurers are not providers of care and do not supply charity health services; in addition few plans provide significant amounts of free insurance, though some subsidize premiums for those who cannot afford to pay. Community rating, which spreads risk broadly across populations, has largely disappeared as competitive market pressures have caused nonprofit Blue Cross plans (which were once required by regulation to community rate) to emulate for-profit counterparts and adopt experience rating. More subtle community benefits may be found in the underwriting and risk selection practices of these organizations. That is, nonprofits may eschew practices associated with favorable risk selection such as seeking to attract healthier subscribers through underwriting or product design and marketing. Such practices diminish the benefits of broad pooling of risk and thus deprive the less healthy segments of society the implicit subsidy they receive from healthier citizens. Even here, however, the picture is not one-sided: More accurate risk underwriting increases the number of people who will be able to afford health insurance. Finally, there are a host of other, somewhat inchoate benefits that may be associated with nonprofit health plans. For example, they may be more responsive to community needs, more active in

41. See HANSMANN ET AL., OWNERSHIP, *supra* note 4.

42. See Brown, *supra* note 40, at 36.

43. Horwitz, *supra* note 5, at 1364.

44. *Id.* at 1373.

advocating public policies that serve the community, or more inclined to provide coverage for services that have public goods characteristics such as immunization and health education programs.

Studies of HMOs, which integrate insurance and delivery of health care, provide fairly persuasive evidence of differences between for-profit and nonprofit firms in the non-price dimensions of their services. One important recent study comparing HMOs using fifty-three measures representing eight distinctive dimensions of community impact reports that nonprofits provide more community benefits than their for-profit counterparts.⁴⁵ It found statistically significant evidence that nonprofit HMOs were more likely to provide subsidies for medical services, support safety net health care agencies, target community benefit programs to low income neighborhoods, and provide general philanthropy.⁴⁶ In addition, studies of consumer satisfaction and consumer evaluations of quality generally, but not uniformly, reflect favorably on nonprofit HMOs.⁴⁷ Such findings may be the result of the public's perception that the for-profit HMO owners' financial stake and ability to make a profit results in the limiting of services to patients.⁴⁸

Turning from nonprofit HMOs to nonprofit companies primarily engaged in selling health insurance and network packages such as Blue Cross and Blue Shield plans (the Blues), there is far less evidence of community benefit, however defined. For example, a large number of studies examining health plans that converted from nonprofit to for-profit status show that the conversion had little or no impact on customer service

45. See generally Mark Schlesinger et al., *Measuring Community Benefits Provided by Nonprofit and For-Profit HMOs*, 40 INQUIRY 114 (2003).

46. *Id.* at 125.

47. Bruce E. Landon et al., *Health Plan Characteristics and Consumers' Assessments of Quality: For the First Time, the Characteristics of Health Plans Are Linked with Consumer Feedback in a Nationwide Survey*, 20 HEALTH AFF. 274, 281 (2001); see also Mark A. Hall & Christopher J. Conover, *The Impact of Blue Cross Conversions on Accessibility, Affordability, and the Public Interest*, 81 MILBANK Q. 509, 520 (2003) (summarizing studies and concluding that "although the evidence is mixed, it suggests that members of nonprofit HMOs are more satisfied and receive better service and a somewhat higher quality of care"); Robert Kuttner, *Must Good HMOs Good Bad?*, 21 NEW ENG. J. MED. 1558, 1562 (1998) ("[N]onprofit [health] plans as a group tend to score better on many objective indicators and in surveys of consumers.").

48. See Bradford H. Gray, *Conversion of HMOs and Hospitals: What's At Stake?; The Pros and Cons of Nonprofit Conversions Through the Lens of Public Policy*, 16 HEALTH AFF. 29, 40 (1997). Another important qualification of statistical comparisons between for profit and nonprofit HMOs is that they may not adjust adequately to reflect significant differences in the populations they serve. See Hall & Conover, *supra* note 47, at 520.

or consumer satisfaction; evidence regarding recent Blue Cross Plans which have converted show that customer satisfaction scores have actually increased post-conversion.⁴⁹ Conversion studies also examine relative profitability, pricing, and access: Here too there is no persuasive evidence that nonprofits offer significant benefits. While it is clear that moving from not-for-profit to for-profit status impels organizations to generate more profits,⁵⁰ the change neither generates significant gains in efficiency nor improvements in terms of the firm's overall financial condition.⁵¹ Although some claim that for-profit health plans in general engage in aggressive risk selection in underwriting practices,⁵² the evidence on this score is at best mixed.⁵³

Of course when one addresses the conversion issue from a policy standpoint, it is necessary to consider offsetting benefits that may accrue. Weighing in favor of conversions are factors such as enhanced efficiency and lower costs resulting from more aggressive negotiating with providers and tax payments that will flow to the public sector.⁵⁴ Finally, and perhaps most important is putting resources to their best use. As Hall and Conover put it, "The largest potential benefit [of conversions of nonprofit plans] is to unlock considerable wealth that can be devoted to explicitly health related charitable purposes."⁵⁵

49. Hall & Conover, *supra* note 47, at 531 (noting that Blue Cross plans in California have improved customer satisfaction scores).

50. Hall & Conover, *supra* note 47, at 515.

51. See Robert Cunningham & Douglas Sherlock, *Bounceback: Blues Thrive as Markets Cool Toward HMOs*, 21 HEALTH AFF. 24, 30 (2002) (noting that while all Blue Cross plans have become more profitable in recent years, the for-profit Blue Cross plans may have been profitable even if they had remained nonprofit).

52. See Kuttner, *supra* note 47, at 1561 ("[E]ntrepreneurial commercial HMOs . . . tend to engage in more aggressive risk selection, use more stringent systems of approval and denial of care, and put a higher fraction of the physicians' income at risk.").

53. Hall & Conover, *supra* note 47, at 530 (studies indicate that "the time has passed when [Blue Cross] plans were much more lenient underwriters than other insurers, and underwriting practices and policies at nonprofit [Blue Cross] plans are now broadly consistent with those of for-profit insurers."). Interviews conducted by Hall and Conover with a broad array of individuals familiar with the effects of Blue Cross conversions in their states indicate divergent outcomes. In some states, interviewees thought that the underwriting practices of the converted Blues were similar in comparison to other insurers, if not more lenient. However, respondents in California and Missouri thought that conversion had adversely impacted the risk selection in these states. *Id.* at 530-31.

54. See *id.* at 521-23, 532-33.

55. *Id.* at 538.

This generalized description of the nonprofit health care sector provides background for evaluating legal doctrine in specific contexts. It suggests that theoretical accounts purporting to explain the persistence of the nonprofit sector do not provide a convincing argument that it will automatically supply desired public benefits. While the empirical literature confirms that the sector has not fulfilled society's goals, our interpretation of this evidence views the glass as half full. We find ample reason to believe that, properly incentivized, nonprofits could supply public goods efficiently and creatively. We turn next to explaining why the legal regime does not satisfy the conditions necessary to promote the sector achieving its goals.

II. ILLUSTRATIVE CASES

The 1990s witnessed a sharp increase in the number of cases involving breaches of fiduciary duties by directors and officers of nonprofit corporations that have prompted aggressive review by state attorneys general.⁵⁶ We identify in subsequent Sections of this Article two central flaws in the law regulating nonprofit governance: an insufficiently stringent standard of conduct for directors, which has countenanced neglect and abuse, and a failure to afford directors leeway to take into account the charitable mission in their business decisions. As a prelude to our doctrinal analysis and recommendations, this Part presents a handful of prototypical cases that illustrate these problems.

In the view of many academic commentators, the experience of recent years in the nonprofit sector involving well-publicized directorial conflicts of interest and lax oversight confirm theoretical claims that fiduciary standards are set "too low" and inadequately constrain the behavior of nonprofit management tempted by opportunities for abuse.⁵⁷ While it is hazardous to generalize from a few episodes of abuse,⁵⁸ the "too low"

56. See *Aramony v. United Way*, 949 F. Supp. 1080 (S.D.N.Y. 1996); *Adelphi Univ. v. Bd. of Regents*, 647 N.Y.S.2d 678 (1996); *In re Found. for New Era Philanthropy Litig.*, 175 F.R.D. 202 (E.D. Pa. 1997); Honolulu Star Bulletin, *Kamehameha Schools/Bishop Estate Archives*, <http://starbulletin.com/specials/bishop1997.html> (last visited Nov. 19, 2004). See generally MARION R. FREMONT-SMITH & ANDRAS KOSARAS, *WRONGDOING BY OFFICERS AND DIRECTORS OF CHARITIES: A SURVEY OF PRESS REPORTS 1995-2002* (Hauser Center for Nonprofit Organizations, Working Paper No. 20, 2003), <http://ssrn.com/abstract=451240> (identifying 104 criminal cases and fifty-four breach of duty cases, the majority of which involved human service agencies).

57. Harvey J. Goldschmid, *The Fiduciary Duties of Nonprofit Directors and Officers: Paradoxes, Problems, and Proposed Reforms*, 23 IOWA J. CORP. L. 631, 642 (1998).

58. See, e.g., Sasso, *supra* note 20, at 1519 ("[E]xtrapolating from a few outrageous

hypothesis merits close attention and has received implicit endorsement from legislative actions targeting directorial oversight abuses through federal tax,⁵⁹ Sarbanes-Oxley,⁶⁰ House of Representative hearings on the tax exempt status of hospitals,⁶¹ Senate Finance Committee oversight hearings regarding nonprofit governance,⁶² state laws targeting governance in specific circumstances⁶³ and regulatory actions taken by the Internal Revenue Service⁶⁴ and national exchange regulators.⁶⁵ The second

scandals to conclude that there is a pervasive problem plaguing the entire not-for-profit industry is a misguided leap in logic.”).

59. A significant recent change in federal tax policy targeting self-dealing abuses was the enactment of an excise tax penalizing so-called excess benefit transactions. 26 U.S.C. § 4958 (2000).

60. Pub. L. No. 107-204, 116 Stat. 745 (2002). While not directly altering fiduciary obligations, Sarbanes-Oxley contains a number of provisions that affect the conduct of fiduciaries and composition of important committees. For example, section 301 requires that audit committee members be independent; section 402 forbids personal loans to directors and executive officers; and section 407 mandates rules requiring public companies to disclose whether the audit committee is comprised of at least one member who is a financial expert. *Id.* See generally Lyman P.Q. Johnson & Mark A. Sides, *The Sarbanes-Oxley Act and Fiduciary Duties*, 30 WM. MITCHELL L. REV. 1149 (2004).

61. See *First Hearing in a Series on Tax Exemption: Pricing Practices of Hospitals Before the House Comm. on Ways & Means, Subcomm. on Oversight*, 108th Cong. (2004).

62. In June 2004, the Senate Finance Committee held hearings concerning a variety of abuses and failures of governance in charitable organizations. *Charity Oversight and Reform: Keeping Bad Things from Happening to Charities: Hearing Before the Senate Comm. on Fin.*, 108th Cong. (2004). The committee also issued a white paper, STAFF OF SENATE COMM. ON FIN., 108TH CONG., STAFF DISCUSSION DRAFT, <http://finance.senate.gov/hearings/testimony/2004test/062204stfdis.pdf>, outlining possible reforms, many of which deal with the mechanisms of accountability in nonprofit organizations. Among the proposals contained in the white paper are limitations on the size of boards of directors, specific standards for fulfilling fiduciary duties, improved disclosures of financial matters, standards and enhanced penalties for self dealing, and a required five-year review of exempt status of all exempt organizations by the IRS. *Id.*

63. See, e.g., MODEL ACT FOR NONPROFIT HEALTHCARE CONVERSIONS (1998), reprinted in THOMAS L. GREANEY & ROBERT SCHWARTZ, *HEALTH LAW: SELECTED STATUTES AND REGULATIONS* 300 (2003).

64. Responding to widespread concerns that charities were awarding excessive compensation and benefits to officers and insiders, the IRS recently announced a new enforcement effort that will examine levels of compensation, insider loans, and the exchange and sale of property to officers and others. Kurt Ritterpusch, *IRS Launches Enforcement Effort Targeting Compensation in Tax-Exempt Organizations*, 13 BNA HEALTH L. REP. 1183, 1183 (2004). The heightened attention to compensation issues appears to have been prompted in part by Congressional oversight hearings concerning nonprofit organizational

important challenge inadequately met by state law governing fiduciaries is the need to ensure nonprofit agents' fidelity to their institutions' charitable purposes. State law is curiously silent on how mission—the central precept guiding the nonprofit charity—should inform directors' interpretations of their responsibilities under nonprofit corporate law. Wielding considerable leverage over nonprofit boards, some attorneys general have through their enforcement actions implicitly assumed de facto powers over a broad spectrum of business decisions and health policies.

A. *Attorneys General's Attempts To Police Conflicts of Interest and Laxity*

The widespread conversions to for-profit status by nonprofit health plans and hospitals in the nineties⁶⁶ served as a wake-up call to attorneys general, most of whom had not previously actively monitored that sector. These transactions, which in many cases the attorney general learned of after the fact, gave rise to numerous allegations of breaches of fiduciary duties by directors and officers. In some instances, overt conflicts of interest were present in which insiders took jobs⁶⁷ or ownership interests in the for-profit acquirer with which they had negotiated sales on favorable terms.⁶⁸ In *Butterworth v. Anclote Manor Hospital*,⁶⁹ for example, Florida's

governance. *Id.* ("The closer we look at charities in our Finance Committee, the stronger the case gets for meaningful legislative reforms that shut down exorbitant pay for charity executives and sweetheart deals for insiders . . .") (quoting Senator Grassley).

65. See Self-Regulatory Organizations; New York Stock Exchange, Inc. and National Association of Securities Dealers, Inc., Exchange Act Release No. 34-48745, 68 Fed. Reg. 64154 (Nov. 4, 2003) (approving NYSE and NASDAQ self governance regulations).

66. See James J. Fishman, *Checkpoints on the Conversion Highway: Some Trouble Spots in the Conversion of Nonprofit Health Care Organizations to For-Profit Status*, 23 IOWA J. CORP. L. 701, 702 (1998) (describing the conversion phenomenon of the 1990s as "the largest redeployment of charitable assets in the Anglo-American world since Henry VII [sic] closed the monasteries in 1536-1540"). See generally Eleanor Hamburger et al., *The Pot of Gold: Monitoring Health Care Conversions Can Yield Billions of Dollars for Health Care*, 29 CLEARINGHOUSE REV. 473 (1995).

67. See Andrea Gerlin, *Hospital in Florida Is Focus of Probes Tied to Scuttled Bid by Columbia/HCA*, WALL ST. J., May 8, 1995, at B10 (reporting allegations that the president of a Florida hospital who intentionally devalued the hospital in an attempt to sell it at an attractive price to a proprietary chain subsequently took a management position with that chain after being terminated by the hospital).

68. For example, when Health Net, a nonprofit HMO, converted to for-profit form, thirty-three executives were able to purchase twenty percent of the stock of the new entity for \$1.5 million; four years later those shares were worth approximately \$315 million.

Attorney General challenged the conversion of a nonprofit hospital whose assets were purchased by a for-profit, the sole shareholders of which were the directors and corporate members of the nonprofit. The assets were purchased for \$6.3 million; two years later, the converted, for-profit hospital was sold for more than \$29 million.⁷⁰

While conversions and closures of health systems fueled concerns among attorneys general about managerial abuse,⁷¹ the rapid vertical integration occurring throughout the health care sector also gives rise to instances of self-dealing and lax directorial supervision. The collapse of the Allegheny Health, Education, and Research Foundation (AHERF) in the nation's largest nonprofit health care bankruptcy case provides the paradigm example of unsupervised management excess. Under the leadership of its Chief Executive Officer, Sherif Abdelhak, AHERF grew rapidly, borrowed heavily, and collapsed precipitously. As several careful studies of AHERF business operations reveal, the over-arching problem was the structure and performance of its corporate governance system.⁷² Over

Lawrence E. Singer, *The Conversion Conundrum: State and Federal Responses to Hospitals' Changes in Charitable Status*, 23 AM. J.L. & MED. 221, 231 n.52 (1997).

69. 566 So. 2d 296 (Fla. Dist. Ct. App. 1990); *see also* Fair Care Found. v. D.C. Dep't of Ins. & Secs. Regulation, 716 A.2d 987 (D.C. 1998) (rejecting claims that the board's decision was infected by conflicts of interest and issues going to members integrity).

70. *Butterworth*, 566 So. 2d at 297.

71. In an interesting twist, in October 2003 the Santa Paula, California City Council voted to ask the California Attorney General to compel a local nonprofit hospital to complete a merger deal with the public health care system, which, the board claimed, offered a better chance than the nonprofit alternative to save the cash strapped rural facility. Laura B. Benko, *California Attorney General Asked to Force Merger Meant to Save Hospital*, MOD. HEALTHCARE, Oct. 27, 2003, at 14. The City Council alleged that the nonprofit board has been dilatory in taking the necessary actions to save the hospital, in violation of the state code governing nonprofit facilities. Amanda Covarrubias, *Hospital Merger May Get a Nudge*, L.A. TIMES, Oct. 22, 2003, at B1. Santa Paula ended up closing and declaring bankruptcy. Lynne Barnes, *Clinics To Extend Medical Services*, L.A. TIMES, Jan. 14, 2004, at B3.

72. The complex AHERF organization was governed by a parent board consisting of no fewer than thirty-five members. Ten other boards, having little overlapping membership, governed fifty-five corporations; each board was generally unaware of what other parts of the system were doing. Directors were chosen and dominated by Mr. Abdelhak and board meetings were, according to one analysis, "scripted affairs, intentionally staged to limit oversight and participation by board members [M]embers . . . receive as many as 1,000 pages of paper to be discussed at board meetings. . . . As one former member explained, 'Half of the people didn't even open the book. They didn't have the time.'" Lawton R. Burns et al., *The Fall of the House of AHERF: The Allegheny Bankruptcy*, 19 HEALTH AFF. 7, 21 (2000). Although the AHERF boards consisted of top-notch executives, all were extremely

sixty lawsuits were filed after AHERF's collapse, most alleging breaches of the duty of care and duty of loyalty by directors.⁷³ The Pennsylvania Attorney General's prosecution and resulting recovery stressed the role of nonprofit directors in safeguarding assets and their legal responsibilities when oversight is lacking. The ultimate AHERF settlement resulted in a distribution of \$93.7 million.⁷⁴ Criminal prosecution also resulted in confinement for Mr. Abdelhak.⁷⁵

B. Attorneys General's Attempts To Regulate Mission

As we discuss in Part III, the law is virtually silent on the question of when, why, and how a charitable corporation may alter its purpose or redeploy its assets to fulfill a re-envisioned sense of its mission.⁷⁶ This Section samples a few instances in which attorneys general have challenged nonprofit boards' strategic plans. Several have used mission-protective concepts from charitable trust law or invoked corporate fiduciary principles to enjoin the board's execution of its plans or to replace board members. Other attorneys general have used similar legal arguments in attempts to bar movement of charitable assets out of state.

1. Whose Mission?

Frequently, challenges made by attorneys general to actions by

busy and unable to perform a broad oversight responsibility over the organization. In addition, the bylaws permitted many key decisions to be made by Mr. Abdelhak. *Id.*

73. See *infra* Subsection II.A.2.

74. The settlement "represent[ed] payments of \$48 million from the insurers, \$28.5 million from Mellon Bank, \$1 million from Allegheny General Hospital and \$7.75 million from funds held by bankruptcy trustees." FREMONT-SMITH & KOSARAS, *supra* note 56, at 20 (citing the settlement agreement at http://www.attorneygeneral.gov/ppd/PDF/AHERF_Settlement_Agreement.pdf). "More than \$49 million of the total was paid to creditors, \$22 million was paid to the Attorney General for distribution to the surviving charitable foundation, \$13 million was paid for legal fees, and \$4.5 million was paid to settle a class action lawsuit doctors brought against the Foundation." *Id.* at 20.

75. The AHERF CEO was sentenced to eleven and a half months but served three and a half months. See Cinda Becker, *Early Release: Abdelhak Wins Parole after Serving Three Months*, MOD. HEALTHCARE, Feb. 3, 2003, at 18; Editorial, *AHERF Whimper*, PITTSBURGH POST-GAZETTE, Sept. 8, 2002, available at <http://www.post-gazette.com/forum/20020908edsharif0908p1.asp>; *Anatomy of a Bankruptcy* (pts. 1-6), PITTSBURGH POST-GAZETTE, Jan. 17-24, 1999, collected at <http://www.postgazette.com/aherf/>.

76. Evelyn Brody calls this the "front-end *cy pres* issue." Brody, *Whose Public?*, *supra* note 13, at 962.

nonprofit boards implicate the organization's mission. These cases typically arise in the context of disputes over attempts by boards to change the corporate purpose or to undertake "organic" changes, e.g. mergers, joint ventures, conversions, and closures that ultimately impact the institution's mission. Underlying these legal disputes is an issue going to the heart of the nonprofit governance debate: Who ultimately controls a charitable corporation's mission?

Two New York cases illustrate the uncertainty attending judicial (or prosecutorial) attempts to monitor mission fidelity under the current state of the law. First, *Littauer v. Spitzer* involved a merger, driven by financial exigencies, between a secular and a Catholic hospital, each of which were controlled by parents; the merger was accomplished by transferring control of both hospitals to a common parent, which itself became a joint subsidiary of the original parents.⁷⁷ A major point of contention was the hospitals' agreement that the Catholic Ethical and Religious Directives would apply to all corporate entities, thereby eliminating access to certain reproductive health services that had previously been provided by the secular hospital.⁷⁸ Positing that the transaction essentially constituted a change in the purposes and ownership of the two facilities, the Attorney General contended that his approval was required under New York's nonprofit statute. An appellate court concluded that the state's nonprofit law was not implicated and that the attorney general had no role in approving the transaction. In reaching this result, it held that a change in corporate membership of the respective hospital corporations neither added, eliminated, or changed a corporate purpose or power⁷⁹ nor constituted the "functional equivalent of a sale, lease, exchange or other disposition of corporate assets."⁸⁰ Responding to the concerns expressed about the elimination of reproductive health services, the court in *Littauer* distinguished between a change to a corporate power and a change to services, holding that the latter falls squarely within the business discretion of the board of directors and should not be subject to judicial second-

77. *Littauer Hosp. Assoc. v. Spitzer*, 287 A.D.2d 202 (N.Y. App. Div. 2001); see also Robert P. Borsody, *The Parent-Subsidiary Structure in Not-For-Profit Hospital Transfers*, N.Y. L.J., Jan. 2, 2004, at 4.

78. Before proceeding, the parties secured a Department of Health ruling that no regulatory approval of the transaction was required. The State Department of Health declines oversight of nonprofit hospital affiliations under a "passive parent rule." William Josephson, *Charities Law: Guidance for Practitioners and Fiduciaries*, N.Y. L.J., Feb. 10, 2003, at 4 n.9.

79. 287 A.D.2d at 204-06.

80. *Id.* at 207.

guessing.⁸¹

The *Littauer* decision stands in marked contrast to the decision two years earlier of another New York court in the *Manhattan Eye, Ear and Throat Hospital (MEETH)* case that had suggested much broader attorney general authority over nonprofit board decision-making.⁸² MEETH, a fixture on the upper-east side of Manhattan for almost a century, is a world-renowned, acute care specialty hospital in ophthalmology, otolaryngology, and plastic surgery.⁸³ In the face of continuing declines in operating revenues resulting from reductions in third party reimbursements and a general shift from in-patient to out-patient admissions, the board decided that its mission would be best actualized by “monetizing” MEETH’s principal asset—real estate—and investing the proceeds in free-standing diagnostic and treatment centers in underserved areas of the city.⁸⁴ The court characterized MEETH’s strategy as abandoning the “acute care, teaching and research hospital component of its mission,” and analogized it to a conversion.⁸⁵ Because the hospital sale constituted a fundamental change to its business purposes, the court concluded, that the attorney general did indeed have standing to review the transaction. “While it is certainly correct that the definition of ‘hospital’ . . . includes a diagnostic and treatment center, as MEETH now argues, it is sophistry to contend that this means that MEETH is not seeking a new and fundamentally different purpose.”⁸⁶ Thus, in contrast to *Littauer*, the court performed its own “de novo” analysis of the nonprofit’s mission and reached a conclusion that virtually ignored the board’s assessment of how to respond to a significantly changed financial environment while remaining true to its original mission.

The elusive legal status of mission is also illustrated in cases involving integrated delivery systems which bring multiple actors in the health care system under one corporate parent, sometimes including both providers and payors.⁸⁷ The unique invocation of mission principles by the

81. *Id.* at 206-07.

82. *Manhattan Eye, Ear & Throat Hosp. v. Spitzer*, 715 N.Y.S.2d 575, 592-93 (Sup. Ct. 1999).

83. *Id.* at 577.

84. *Id.* at 577-79.

85. *Id.* at 594-95. “[I]n both there is a charitable organization which alleges that it is incapable of continuing its primary mission of operating a hospital, seeks approval of the sale of all its assets, and plans to apply the sale proceeds towards a newly revised mission.” *Id.*

86. *Id.* at 595.

87. These integrated delivery systems are generally formed precisely for the purpose of

Minnesota Attorney General in his investigation of the Allina Health System suggests the protean nature of the doctrine as currently applied.⁸⁸ Allina's multi-corporate structure included entities that provided health services and health insurance.⁸⁹ Although this organizational structure is quite common, the Minnesota Attorney General took the position that the structure is impermissible because it is impossible for related organizations to pursue the missions of both its nonprofit HMO and its hospitals.⁹⁰ He claimed that the HMO's mission—to manage health costs and control premiums—conflicted with the hospitals' "different," "broader," and "sometimes conflicting" mission "to act as caregivers to patients."⁹¹ Following extensive and sometimes bitter negotiations, Allina agreed to

capitalizing on the benefits that can be achieved from horizontal and vertical integration. Thomas L. Greaney, *Managed Competition, Integrated Delivery Systems and Antitrust*, 79 CORNELL L. REV. 1507, 1516 (1994).

88. See MINN. ATT'Y GEN., MEDICA: CONFLICTS OF INTEREST WITH ALLINA 7-11, http://www.ag.state.mn.us/consumer/PDF/Allina/Allina_Medica.PDF (last visited Mar. 17, 2004).

89. Health Systems and Medica Health Plans had interlocking directorates—seven Allina board members served as Medica directors. *Id.* at 3.

90. Allina Health System entered into a Memorandum of Understanding (MOU) that required Allina to spin-off its HMO affiliate, Medica Health Plans, and adopt a variety of new policies dealing with problems arising out of conflicts of interest, expense reimbursement, executive compensation, third party contracting, and other matters. See Memorandum of Understanding Between Allina Health System and Attorney General of Minnesota, <http://www.ag.state.mn.us/consumer/PDF/allina/MemUnder.pdf> (last visited Mar. 18, 2003).

91. The report accompanying the memorandum of understanding between the state and Allina, MIN. ATT'Y GEN. OFFICE, ALLINA HEALTH SYSTEM REPORTS, *collected at* http://www.ag.state.mn.us/consumer/PR/pr_allina_mou_92401.htm (last visited Nov. 12, 2004), enumerated several instances of Medica board decisions that benefited the Allina Health System—by favoring other Allina entities—to the potential detriment of Medica. For example, Medica resolved to undertake a number of changes designed to reduce the unfavorable sector of its Medicare risk pool. Medica reported its plan to Allina Health System, which then studied the profitability of seniors to its hospitals. After Allina concluded that the Medicare population was an important revenue base for its hospitals, Medica reversed course, rejected its conclusions of a year prior, and re-entered the senior Medicare managed care market. This led the Attorney General to conclude:

While it would serve Medica's interest to charge a fee that included a profit for such services, it generally operates the PPO function as a 'channeling' vehicle for Allina. Medica basically charges health plans and TPAs a fee less than competitors for PPO work in order to build up patient volume for Allina.

Id.

spin the HMO off from its integrated delivery system. The outcome was more than a bit startling. Neither before nor after this case have commentators or policy experts seriously entertained the thought that common ownership of providers and insurance subsidiaries gave rise to a disabling conflict of interest. Perhaps equally notable was the Attorney General's ability to assert direct control over the nonprofit plan. The settlement agreement empowered Attorney General Hatch to appoint eight "special administrators,"⁹² itself creating something of a conflict of interest since the new fiduciaries appointed by the Attorney General were also subject to his supervision.

2. *Whose Money?*

Recent interventions by attorneys general and state insurance regulators in multi-state transactions reveal what Evelyn Brody has aptly characterized a growing "parochialism" that often seems more related to political ends than public policy goals.⁹³ As noted above, conversions of not-for-profit to for-profit hospitals and health plans resulted in the disappearance of millions of dollars in charitable assets due to undervaluation, laxity, and in some cases, management self dealing.⁹⁴ Attorneys general and state legislatures finally reacted to ensure that boards were making conversion decisions in the interests of the corporation rather than themselves,⁹⁵ that the assets of the corporation

92. See Stephanie Strom, *Strong-Arm Shaking of Charities Raises Ethics Qualms*, N.Y. TIMES, May 11, 2003; see also Brody, *Whose Public?*, *supra* note 13, at 1007. Perhaps not coincidentally, Minnesota Blue Cross and Blue Shield decided in 2003 to get out of the hospital business, selling its Fargo hospital to a Catholic health care system. Patrick Reilly, *Back to Basics; Minn. Blues To Abandon Hospital Ownership*, MOD. HEALTHCARE, Sept. 15, 2003, at 12.

93. Brody, *Whose Public?*, *supra* note 13.

94. See *supra* notes 72-75 and accompanying text; see also John F. Coverdale, *Preventing Insider Misappropriation of Not-For-Profit Health Care Provider Assets: A Federal Tax Law Prescription*, 73 WASH. L. REV. 1, 3-6 (1998) (describing conversions and attendant abuses); Shelley A. Sackett, *Conversion of Not-for-Profit Health Care Providers: A Proposal for Federal Guidelines on Mandated Charitable Foundations*, 10 STAN. L. POL'Y REV. 247, 250 (1999) (describing how and why the 1990s saw so many health care conversions); James D. Standish, *Hospital Conversion Revenue: A Critical Analysis of Present Law and Future Proposals To Reform the Manner in Which Revenue Generated from Hospital Conversions Is Employed*, 15 J. CONTEMP. HEALTH L. & POL'Y 131 (1998) (explaining the impetus for so many conversions).

95. See Sackett, *supra* note 94, at 252-53, 254-55 (surveying successful enactment of state legislation governing conversions).

were being appropriately valued,⁹⁶ and that the proceeds resulting from the conversions were being dedicated to suitable ends.⁹⁷ In what may at first blush appear to be a natural extension of these concerns, attorneys general have sought vigorously to capture the proceeds of transactions involving *nonprofit* health care enterprises. As we shall see, however, there are serious reasons to question the doctrinal and policy foundations for these enforcement actions.

i. Banner

A common reason that nonprofit health care systems have been disposing of some of their health care facilities only to turn around and pick up new ones is regionalization. Systems whose holdings were scattered across disparate states have been attempting to consolidate in fewer contiguous states where resources can be more effectively (and more profitably) deployed.⁹⁸ In 2001, Banner Health System, a nonprofit corporation based in Arizona, began doing precisely this—funding expansions in Arizona and Colorado with the proceeds from sales of ten of its twenty-seven hospitals and seventeen long-term-care facilities in seven other states.⁹⁹ Concerned about the exodus of charitable assets from their states resulting from these sales, the attorneys general of North Dakota, South Dakota, and New Mexico¹⁰⁰ attempted to prevent Banner from removing the proceeds from the facilities within the borders of their respective states.¹⁰¹ The attorneys general posited that because the facilities

96. See John Colombo, *A Proposal for an Exit Tax on Non-Profit Conversion Transactions*, 23 IOWA J. CORP. L. 779, 785-86 (1998) (describing “horror stories” of excessive enrichment and undervaluation); Sackett, *supra* note 94, at 250-51 (describing pre-regulatory intervention valuation abuses); Standish, *supra* note 94, at 136, 138-40 (describing how under-valuation was accomplished).

97. See Standish, *supra* note 94, at 144-64 (categorizing the different approaches states have taken in legislating post-conversion foundations).

98. Since 1998, the large mergers emblematic of the preceding decade have fallen off. Most mergers and acquisitions in 2002 involved community hospitals acquiring nearby facilities, so that they could expand their local market. Patrick Reilly, *Mergers Minus the Mania*, MOD. HEALTHCARE, Jan. 20, 2003, at 36.

99. Patrick Reilly, *Trust Challenged: AHA Considers Involvement in Charitable Trust Fight*, MOD. HEALTHCARE, Apr. 21, 2003, at 6.

100. Banner’s sale of its forty-seven bed New Mexico facility to Province Healthcare Co., a Tennessee-based for-profit company, prompted that state’s attorney general to threaten a lawsuit for breach of trust; Banner paid a \$4 million settlement to New Mexico. *Id.*

101. Barbara Gorham, *Opinions/Commentary, Banner’s End Run Must End: Company Plays Chess with Assets It Inherited While Communities Pay the Price*, MOD. HEALTHCARE, Mar. 3,

had benefited from the support of their local communities, which enhanced the value of each entity's assets, Banner would be unjustly enriched if allowed to transfer those assets out of state.¹⁰²

On notice of the South Dakota Attorney General's plans, Banner Health System filed a declaratory judgment action¹⁰³ to preclude the Attorney General from imposing a constructive charitable trust on Banner's South Dakota facilities.¹⁰⁴ Although nonprofit corporate law would plainly permit sales and transfers within a multicorporate structure, the South Dakota State Supreme Court was unpersuaded by Banner's argument that the state's nonprofit corporate statute exclusively controlled the transaction. Rather, it held that in enacting the state's nonprofit corporate law, "there is nothing in the code to indicate that the Legislature intended to abrogate common law and statutory trust provisions with regard to nonprofit corporations."¹⁰⁵ And even though Banner was not

2003, at 21.

102. Patrick Coffey et al., *The "Charitable Trust" Controversy Confronting Banner Health and Other Nonprofit Healthcare Systems*, 16 HEALTH L. 1, 3 (2003). Banner's consolidation resulted in several settlements and court decisions. A trial court in North Dakota dismissed the Attorney General's complaint against Banner, concluding that community donations to local hospitals do not satisfy the elements of a constructive trust; the court also rejected the unjust enrichment argument. *Id.* Banner and the North Dakota Attorney General eventually settled their differences when Banner agreed to a \$1 million payment to the state. State Roundup, GRAND FORKS HERALD, Dec. 16, 2003. Banner settled with New Mexico for \$8.5 million, which would be paid to charities dedicated to health care selected by the Attorney General. *Briefly: Hospital Deals*, MOD. HEALTHCARE, Dec. 23, 2003, at 10; *New Mexico: Banner Health Systems, State AG Settle on Sale of Medical Center to For-Profit Firm*, 11 BNA HEALTH L. REP. 831 (2002).

103. Banner was attempting to sell its hospital to Catholic Health Initiatives, a Denver-based nonprofit health care system, and its nursing home to Sisters of Mary of the Presentation Health System. *Banner Health Sys. v. Stenehjem*, No. A3-02-121, 2003 WL 501821 (N.D. Dist. Ct. Feb. 25, 2003).

104. The history of the several facilities, each of which changed hands several times, is detailed in the state Supreme Court decision. Although certain donations to at least a couple of the facilities clearly created trusts (e.g., The Dorsett Home), the facilities were established or supported by a combination of unrestricted donations, fundraisers, and government support. *Banner Health Sys. v. Long*, 663 N.W.2d 242, 245-46 (S.D. 2003).

105. *Id.* at 247. The court specifically sought to preserve the relevance of the following statutory language preserving a court's ability to employ the implied trust device when equity so requires:

The enumeration in §§ 55-1-7 to 55-1-10, inclusive, of cases wherein an implied trust arises does not exclude or prevent the arising of an implied trust in other cases nor prevent a court of equity from establishing and declaring an implied,

obligated under any express trust, the court remanded the case on the theory that an “implied trust” might be applied as a remedial construct to preserve the status quo when “a person owning title to property is under an equitable duty to convey it to another because he would be unjustly enriched if he were permitted to retain it.”¹⁰⁶ If other states adopt this rather freewheeling approach, nonprofit corporations could find their business plans completely thwarted by the imposition of trust-based responsibilities that have little grounding in trust doctrine.

ii. Health Midwest

Another prominent case involving claims of trust-based duties arose from the \$1.125 billion acquisition of nonprofit Health Midwest hospital system by the for-profit corporation HCA, Inc. This transaction provoked a renewed Missouri-Kansas “border war,” pitting the Attorney General of Missouri against the Attorney General of Kansas in a dispute over the legality of the transaction and, more importantly, where the charitable proceeds would land. Although similar to *Banner*, in that it involved an attorney general asserting charitable trust law to extract concessions from the nonprofit entity, the contention met with less success.

Health Midwest was a Kansas City-based integrated delivery system whose various constituent corporations straddled the borders of Kansas and Missouri. After initially threatening to dissolve Health Midwest and remove its board, the Missouri Attorney General settled its side of the case for an agreement that would create a conversion foundation (whose directors would be chosen by the Missouri Attorney General) and which would devote a minimum of ten percent of the conversion proceeds for

resulting, or constructive trust in other cases and instances pursuant to the custom and practice of such courts.

S.D. CODIFIED LAWS § 55-1-11 (Michie 2004).

106. *Long*, 663 N.W.2d at 247 (quoting *Knock v. Knock*, 120 N.W.2d 572 (S.D. 1963)). The court left open the possibility that an implied trust might be appropriately imposed if the Attorney General could establish that *Banner* had engaged in behavior which created unjust enrichment, constituted a breach of fiduciary duties, or improperly amended *Banner*’s articles of incorporation. *Id.* at 248-49. Further, if *Banner* was in a fiduciary relationship with the communities in which its facilities were located, pursuant either to trust law or the general common law governing fiduciary relationships, *Banner* may have breached its duties as a fiduciary if, as alleged by the Attorney General, its actions were premised on the best interests of *Banner*, rather than the local communities, who are the beneficiaries of the relationship. *Id.* at 249.

the benefit of Kansas.¹⁰⁷ Kansas, finding itself on the short end of the distribution of sales proceeds (Health Midwest's internal estimate placed Kansas' share of assets at twenty percent), unleashed arguments grounded in charitable trust and corporate law to oust the board members who had approved the transaction with HCA and settled with the Missouri Attorney General. Fanning the flames, the Kansas legislature attempted to intercede as well.¹⁰⁸

Relying on charitable trust theory, the Kansas Attorney General asked for a judicial *cy pres* proceeding, removal of Health Midwest's directors and the appointment of a fifteen person board (appointed by her) to run the resulting charitable foundation.¹⁰⁹ The court rejected almost all of the Kansas claims, squarely holding that the corporate standard, not the charitable trust standard, governed decision-making in nonprofit corporations.¹¹⁰ Further, contrary to the Attorney General's assertions, the

107. *Health Midwest v. Kline*, No. 02-CV-08043, 2003 WL 328845, at *16-17 (D. Kan. Feb. 6, 2003).

108. Literally days before the Health Midwest trial began, the Kansas legislature enacted a bill, designed to apply to Health Midwest's Kansas' assets, which requires a Kansas nonprofit corporation to forfeit its assets to a foundation rather than to any third party. In the course of declaring the statute unconstitutional, *id.* at *24, the court criticized the state's charitable trust theories as unsupported by Kansas law. Finally, the court observed that the state's compulsion that all charitable assets remain within Kansas' borders could result in the withdrawal from charitable activity any enterprise, foreign or domestic, seeking to protect its assets from seizure by the state. *Id.*

109. The Attorney General claimed that the board was influenced by overly generous compensation packages, failed to exercise due diligence, and failed to exercise reasonable business judgment as to price, process, and use of proceeds in approving their mergers into Health Midwest. See Brody, *Whose Public?*, *supra* note 13, at 1008-17 (summarizing the pleadings in the *Health Midwest* litigation).

110. The Kansas District Court held that application of the charitable trust doctrine in *cy pres* proceedings applied only to changes in restricted gifts and refused to apply it to changes to a corporation's purposes. The Kansas court explained:

The Kansas *cy pres* statute governs changes to the purposes of charitable trusts, devises and bequests. The *cy pres* statute does not apply to changes to the purposes of nonprofit corporations. The *cy pres* statute applies only to any restricted gifts and not the entity as a whole. No restricted gifts have been identified herein and therefore the *cy pres* statute does not apply.

Health Midwest, 2003 WL 328845, at *19 (citation omitted). The court further rejected the Attorney General's attempt to assert the business judgment rule where there was simply a "disagreement over contract terms, *id.* at *18, and reasoned that "a court can not second guess the wisdom of facially valid decisions" of the board of a charitable corporation, *id.* at *17.

court held that *cy pres* does not apply to changes to purposes of charitable corporations.¹¹¹ In sum, the court believed that it was required to uphold the Health Midwest board's decision¹¹² "unless the directors are guilty of 'willful abuse of their discretionary power or of bad faith, neglect of duty, perversion of corporate purpose, or when fraud or breach of trust are involved.'"¹¹³ At the same time the Kansas Court found that under nonprofit corporate law, mission obligations should have compelled the Health Midwest directors to strike a better balance for Kansas: It found that the proposed post-merger Missouri foundation would have insufficient "Kansas participation in its governance" and that the plan offered "nebulous spending commitments to benefit the citizens of Kansas in Health Midwest's Kansas service area."¹¹⁴

iii. *CareFirst*

Finally, some organic changes by nonprofit third party payors have encountered objections from insurance commissioners invoking a mix of corporate, trust and insurance law. Although fourteen Blues plans have converted to for-profit status since 1994,¹¹⁵ such conversions increasingly face stiff opposition, and several have been abandoned, apparently out of concern about the approval process.¹¹⁶ The legal standard applied by state insurance agencies is, if anything, less clear than that invoked by the judiciary.

In 2003, the Maryland Insurance Administration (MIA) rejected the

111. *Id.* at *19. The court specifically observed that the assets represent "proceeds of the sale of corporate assets and not assets of a trust, therefore the *cy pres* statute does not apply." *Id.*

112. *Id.* at *18. The court held that the Attorney General's authority over a nonprofit was limited to determining whether the board's business decisions satisfied the business judgment rule. *Id.* at *17. Calling upon Delaware law, the court recognized its authority to "enjoin the 'transaction of unauthorized business'" if the Attorney General establishes that the board's decision was "ultra vires or a perversion of corporate purpose." *Id.* at *18.

113. *Id.* at *26.

114. *Id.*

115. Laura B. Benko, *Curtain Falls: CareFirst Settlement Dims Hope for Blues Conversions*, MOD. HEALTHCARE, June 16, 2003, at 14. The next conversion battleground is Washington state, where the Washington Hospital Association is attempting to block the proposed conversion of Premera Blue Cross. *Id.*

116. For example, North Carolina Blues withdrew its plan to convert in the face of "a process with no end in sight." *Plan To Convert North Carolina Blues Withdrawn, in Face of Regulatory Risks*, BNA HEALTH CARE DAILY REP., July 9, 2003.

application of CareFirst BlueCross BlueShield to convert and be acquired by for-profit WellPoint Health Networks, Inc.¹¹⁷ In a report exceeding 350 pages, the Maryland Insurance Commissioner concluded that the proposed transaction did not satisfy the public interest standard set forth in the state's conversion statute. The report recounted a number of procedural derelictions, concluding that the bidding process was "flawed and did not produce fair market value."¹¹⁸

The Commissioner's report relies on a mix of corporate law and regulatory criteria from the state conversion statute.¹¹⁹ The result is something of a hodge-podge, with selective application of corporate law principles,¹²⁰ leavened by invocation of various open-ended statutory criteria that the Commissioner concluded justified departure from straightforward corporate analysis.¹²¹ Much of the report is written in the language of corporate fiduciary duties, evaluating the board's diligence and weighing conflicts of interest. Further, the report imposes an obligation "to obey the articulated mission of the corporation,"¹²² and sweepingly concludes that CareFirst's nonprofit status conferred special

117. CAREFIRST CONVERSION INFORMATION, MARYLAND INSURANCE COMMISSIONER, <http://www.mdinsurance.state.md.us/jsp/CareFirst.jsp10?divisionName=CareFirst+Conversion+Information&pageName=/jsp/CareFirst.jsp10> (last visited Mar. 22, 2004) (on file with author) [hereinafter CAREFIRST CONVERSION INFORMATION]. In June 2003, a federal judge approved a settlement between CareFirst and the Insurance Administration that precludes CareFirst from considering a conversion to for-profit status for five years. Benko, *supra* note 115.

118. The auction "appeared designed to, and did, end in a tie on price," while assets were undervalued, the transaction did not protect against private inurement of Blue Cross directors. Overall, the report found that the CareFirst board did not exercise due diligence in deciding to sell, selecting the purchaser, and negotiating the deal; further, it did not sufficiently protect against conflicts of interest. CAREFIRST CONVERSION INFORMATION, *supra* note 117.

119. MD. CODE ANN., STATE GOV'T § 6.5-301 (2004).

120. The *CareFirst* opinion specifically refers to MD. CODE ANN., CORPS. & ASS'NS § 2-405.1 (1999), dealing with the corporate directors generally and codifying the business judgment rule, and to MD. CODE ANN., INS. § 14-115(c) (2002) for the directors of nonprofit health service plans. CAREFIRST CONVERSION INFORMATION, *supra* note 117, at 66.

121. While finding that the state insurance statute "codifies the traditional fiduciary duties of care and loyalty that historically govern the conduct of directors of both for-profit and nonprofit corporations," the report further states that certain entities vested with a public trust have "a higher degree of care than the directors of a general corporation." *Id.* at 68, 69.

122. *Id.* at 75.

obligations on its board.¹²³ At the same time, the opinion expressly declines to apply some bedrock corporate law standards like the business judgment rule¹²⁴ or standards applicable to corporate takeovers.¹²⁵

III. STATE FIDUCIARY LAW

As is the case with for-profit businesses (and probably more so), agency problems make the issue of accountability the central problem that must be addressed by nonprofit organizational law.¹²⁶ Until recently, however, courts and charitable regulators have paid remarkably little attention to the key mechanisms affecting accountability. As Part II describes, state attorneys general have brought dozens of cases in recent years that implicate these issues in contexts ranging from unvarnished corruption to business reorganizations necessitated by changing economic conditions. But the glare of the spotlight has only highlighted the manifold inadequacies of legal doctrine regulating governance of nonprofit organizations.

This Part summarizes and criticizes nonprofit corporation law regarding fiduciary duties, which has been a principal tool used by attorneys general in their cases involving the accountability of nonprofit boards. The Part first concludes that corporate fiduciary law is too permissive and uncertain to protect against opportunistic or lax business

123. The Court stated:

CareFirst is a nonprofit corporation. Its [sic] was formed for a public purpose. Its economic "value" constitutes a public asset. The CareFirst Board is, therefore, entrusted with an enterprise whose assets belong to the public. The CareFirst Board was, therefore, required to act with the highest degree a [sic] care

Id. at 75.

124. The report observes:

The business judgment rule was designed to limit judicial interference in corporate affairs. . . . *The "rule," as such, has no place in this regulatory proceeding. . . .* [O]versight of the Insurance Administration over insurance regulatory matters without exception involve evaluation of the substantive outcomes rather than the process through which those outcomes were derived *Application of the business judgment rule in that type of setting would simply emasculate the role of the MIA in evaluating whether or not the company had complied with the statutory standards that govern financial transactions and financial condition.*

Id. at 71-72 (emphasis added).

125. *Id.* at 70; see also *infra* note 200 (discussing corporate directors' obligation in takeover contests under the "Revlon Rule" to accept the highest bid in certain circumstances).

126. See discussion *supra* Part I.

practices. As is true in the for-profit sector, where market discipline and the possibility of a takeover exerts some pressures, nonprofit corporate law cannot be relied upon to police the activities of nonprofit managers and directors. Second, there is reason to doubt that fiduciary law can ensure that managers and directors remain faithful to the nonprofit's corporate mission or will be effective in vetting decisions to alter the mission.

A. Fiduciary Theory and the Nonprofit Commercial Enterprise: An Uneasy Fit

Fiduciary law, embodied in common law duties, statutory standards, and equitable principles, is the primary legal mechanism for assuring accountability in American corporations.¹²⁷ The chief significance of these duties lies in their capacity to address principal-agent problems inherent in the corporate form.¹²⁸ In the for-profit context, agency costs, principally those arising from information asymmetries, limit the ability of residual claimants to monitor the activities of corporate managers in all forms of business association. For nonprofit corporations, the principal-agent problem is magnified in at least two ways: first, that the principal may be an indefinite class (e.g., donors, public beneficiaries of charity, governmental entities, etc.), whose interests may diverge, and second, that the relationship between the (uncertain) principal and agent is not specified

127. Fiduciaries are those undertaking a duty to act for the benefit of others as to matters within the scope of their fiduciary relationship. In the context of business associations, fiduciaries (i.e. corporate directors, who are also sometimes confusingly referred to as "trustees") are held to a good faith standard. James Fishman, *Improving Charitable Accountability*, 62 MD. L. REV. 218, 232 (2003); see *infra* notes 155-169 and accompanying text. In the charitable trust context, fiduciaries ("trustees") hold property subject "to equitable duties to deal with the property for a charitable purpose," RESTATEMENT (SECOND) OF TRUSTS § 348 (1959), and are governed by strict responsibilities to avoid all conflicts of interest, to preserve assets, and to act with prudence and due care. Fishman, *supra* at 228-31. This standard is more exacting than the standard applied in the context of business associations. *Id.* at 231.

128. The issue of agency costs has been the centerpiece of the debate for those attempting to develop a viable theory of the modern corporation. As Berle and Means observed seventy years ago, "The separation of ownership from control produces a condition where the interests of the owner and of the ultimate manager may, and often do, diverge, and where many of the checks which formerly operated to limit the use of power disappear." ADOLF A. BERLE, JR. & GARDINER C. MEANS, *THE MODERN CORPORATION AND PRIVATE PROPERTY* 6 (1933); see Evelyn Brody, *Agents Without Principal*, *supra* note 18, at 473-78 (1996); Geoffrey A. Manne, *Agency Costs and the Oversight of Charitable Organizations*, 1999 WIS. L. REV. 227, 252.

with exactness in either the entity's charter or the law.¹²⁹ The fundamental objective of fiduciary duties in the corporate context, then, is to bind managers to serve their principals' interests and thereby overcome the high agency costs inherent in the corporate form.¹³⁰

Close examination of the structure and economics of the nonprofit firm exposes paradoxes in extending corporate principles to charitable corporations. First and most fundamentally, the nonprofit faces greater obstacles in overcoming agency costs than its for-profit counterparts because it lacks residual claimants. The ability of the capital market to monitor and police the actions of managers in the for-profit sector is generally acknowledged.¹³¹ Interested shareholders can also serve those functions through the mechanisms of corporate democracy, including election of directors, proxy contests and other means of shareholder "activism." However, because the principal of the nonprofit corporation is not readily identifiable, there is no claimant with sufficient incentives to monitor agents' abuses. Further, even if some altruists were willing to act as monitors, existing legal regimes provide few direct remedies for abuses.¹³²

129. See Brody, *Agents Without Principals*, *supra* note 18, at 486; Manne, *supra* note 128, at 234.

130. Other theoretical analyses cast the agency problems in corporate governance in a different light. De-emphasizing the role of fiduciary duties, contractarian scholars argue for a combination of market incentives, enforceable contracts, and other external constraints on opportunism within firms. See, e.g., Henry N. Butler & Larry E. Ribstein, *Opting Out of Fiduciary Duties: A Response to the Anti-Contractarians*, 65 WASH. L. REV. 1, 2-8 (1990); John H. Langbein, *The Contractarian Basis of the Law of Trusts*, 105 YALE L.J. 625 (1995). By this account, fiduciary duties should be seen as "fundamentally contractual" with law enabling parties to adjust duties to suit their particular relationships and achieve economically efficient outcomes. This frame, however, assumes conditions sufficient to enable workable bargaining and mutual exchanges. Applying strict contractarian analyses to nonprofit organizations faces intractable problems given the absence of meaningful bargaining between patrons and agents and the lack of market for mechanisms to monitor their behavior. See, e.g., Victor Brudney, *Corporate Governance, Agency Costs, and the Rhetoric of Contract*, 85 COLUM. L. REV. 1403, 1403-04 (1985). See generally Deborah A. DeMott, *Self-Dealing Transactions in Nonprofit Corporations*, 59 BROOK. L. REV. 131 (1993).

131. While serious questions exist about the sufficiency of capital markets to accomplish these objectives, see *infra* notes 171-172, the extensive literature on corporate governance is in substantial agreement that the market for corporate control has some chastening effect on managers and directors.

132. See *Summers v. Cherokee Children & Family Servs.*, 112 S.W.3d 486, 506-07 (Tenn. Ct. App. 2002) (noting that nonprofit statute allows members to bring a "derivative-like" action, but where no members exist, it is left to the attorney general to respond to breaches of fiduciary duties and where necessary, to seek dissolution).

Lacking effective monitors to demand accountability, one might expect legal doctrine to provide substitute mechanisms to trigger regulatory review in well-defined circumstances. As we shall see, such is not the case.

A caveat is necessary at this point. Public and sponsored hospitals¹³³ provide an interesting wrinkle in this “absence of residual claimant” problem.¹³⁴ In this context, local governments and sponsors frequently

133. We use the term “sponsored” to refer to entities controlled by a religious organization, such as an order of Catholic sisters. Professor Singer predicts that Catholic-sponsored hospitals and attorneys general in particular are on a collision course, as:

[A]ttorneys general and local communities [are] beginning to rigorously question the use of charitable assets. At the same time, Church law clearly vests control of the health care institution and, to a large extent, disposition of its assets in the sponsoring religious congregation. Challenges to sponsor strategies are beginning. There is little doubt that the continued need of sponsors to respond to ministry pressures will, more frequently, result in litigation to contest sponsor authority and direction.

See Singer, *supra* note 2, at 164-65.

134. Of course, some nonprofits (including charitable entities) have structures that mimic “ownership” to some extent. Both public benefit and mutual benefit corporations may have members with rights to elect directors. Under the Revised Model Nonprofit Corporation Act (RMNCA), members of corporations are entitled to vote for directors, while public benefit corporations may have members. REVISED MODEL NONPROFIT CORP. ACT §§ 6.02, 6.03 (1987). Despite having control and governance powers inherent in possessing voting rights, members are not analogous to shareholders in all other respects. Most obviously, they lack any claim to profits: Distributions to members are forbidden except that mutuals may distribute to members on dissolution. Reserved powers are rights of control vested in members that normally are held by the corporation’s Board of Directors. These reserved powers may include power over major operational decisions, sales or conversions, and approval of budgets. See generally Dana Brakman Reiser, *Decision-makers Without Duties: Defining the Duties of Parent Corporations Acting As Sole Corporate Members in Nonprofit Healthcare Systems*, 53 RUTGERS L. REV. 979 (2001). In addition, members may also may have “reserved powers” to make operational decisions thus bypassing the traditional powers of boards and management. Reserved powers are rights of control vested in members that normally are held by the corporation’s Board of Directors. These reserved powers may include power over major operational decisions, sales or conversions, and approval of budgets. See Brakman Reiser, *supra*, at 991. Some nonprofit statutes have recognized these distinctions and applied slightly stricter fiduciary standards to boards of public benefit corporations because of the general absence of members to monitor governance of those organizations. REVISED NONPROFIT MODEL CORP. ACT § 8.30 (1987); Elizabeth A. Moody, *The Who, What, and How of the Revised Model Nonprofit Corporation Act*, 16 N. KY. L. REV. 251, 274 (1988) (noting that RMNCA drafters believed it “essential to find devices to hold directors [of nonprofits without members] accountable”). As a general matter, however, members are best understood as relating to the nonprofit organization by virtue of their participation and

behave as “owners” that provide a consistency of vision and accountability, thereby possibly being even more efficient than shareholders in their oversight of the corporation’s managers. While their existence may ameliorate the “residual claimant” problem in one sense, sponsored hospitals present another analytical challenge. While they may indeed represent well patients’ interests, particularly when they are an on-going enterprise, they also have significant interests of their own, which are easily and powerfully exercised. In short, corporate theories do not account for the “member” corporation, whose members have their independent missions, loyalties, and financial pressures that might be resolved by redeployment of the assets of “subsidiary” corporations.

A second factor undermining the efficacy of fiduciary law in nonprofit corporations is that their goals are multi-faceted and often not well-defined. While managers of business corporations must strictly observe the over-arching objective of profit maximization, their nonprofit counterparts face a more complex array of goals. Although generating net income is surely an important objective (especially in commercial not-for-profit organizations), it is also necessary to simultaneously accommodate the other, competing objectives of the organization articulated in the mission. Thus, nonprofit managers and directors must reconcile business objectives and mission. Complicating the task further is the fact that the mission objectives are often stated in general terms that lack the precise, quantifiable frame posed by the profit maximization standard.¹³⁵ While vague standards may appear to ensure flexibility and maximize director discretion,¹³⁶ the other side of the coin is that they may invite freewheeling

limited governance role in the corporation as distinguished from having a financial investment in the entity. *Id.* at 270 (“[M]embers generally relate to the organization by participation rather than by the financial interest generated by an investment.”); *id.* at 273 (noting that the membership relationship in nonprofits is much more personal than shareholders’ relationship to for-profit corporations).

135. See Goldschmid, *supra* note 57, at 641 (“The obligation of the nonprofit directors and officers with respect to the corporation’s mission creates a more difficult and complex decision-making process for them than for their for-profit peers.”); see also Manne, *supra* note 128, at 235-36 (“[T]he analytical power of the theory of the firm does not readily transfer to the realm of nonprofits. . . . [S]trong conclusions in the for-profit context regarding incentives and capacities to minimize agency problems are weaker in the nonprofit context.”).

136. See, e.g., Goldschmid, *supra* note 57, at 641 (noting it would be in accordance with the duty of care in business to the responsibilities for directors of the nonprofit hospitals to accept the lower bid from one of several suitors because the winning bidder would provide a higher level of public benefit to the community).

regulatory interventions that can bring carefully planned business strategies to a halt.

The efficacy of fiduciary principles is further hampered by the scarcity of precedents. Only a handful of cases address the duties of care¹³⁷ and loyalty;¹³⁸ mention of the duty of obedience is even rarer.¹³⁹ This is in part due to state law limiting standing to challenge breaches of the fiduciary duties to attorneys general, members, and directors.¹⁴⁰ However, state charity enforcers, particularly attorneys general, are notoriously circumscribed by a lack of investigative resources and the dearth of information about managerial abuses or contemplated business decisions owing to the minimal disclosure requirements applicable to nonprofits.¹⁴¹ Also limiting precedent is the attraction of settlement to both states and boards: State regulators and attorneys general focus on “fixing the

137. A search on Westlaw, <http://www.westlaw.com>, using the following terms: “‘duty of care,’ w/5 director! trustee & nonprofit charitable not-for-profit,” found only ten reported decisions involving the duty of care in nonprofit corporations, three of which were cases decided on other grounds, and therefore did not explicate the duty. Search on Westlaw, *All State Cases Database* (Mar. 11, 2004).

138. A search on Westlaw, <http://www.westlaw.com>, using the terms: “‘duty of loyalty,’ w/5 director! trustee & nonprofit charitable not-for-profit,” produced five nonprofit cases which discussed the duty of loyalty. Search on Westlaw, *All State Cases Database* (Mar. 11, 2004); *see also* 2 FURROW ET AL., *HEALTH LAW* §§ 5-15 to 5-16 (2000).

139. A search on Westlaw, <http://www.westlaw.com>, using the search terms: “‘duty of obedience’ w/ 5 director! trustee & nonprofit charitable not-for-profit,” produced one nonprofit case which discussed the duty of obedience. Only one case has cited the duty of obedience since 1984. Search on Westlaw, *All State Cases Database* (Nov. 20, 2004); *see also* 2 FURROW ET AL., *supra* note 138, § 5-17 (listing cases and describing the duty of obedience).

140. *See* DANIEL L. KURTZ, *BOARD LIABILITY: GUIDE FOR NONPROFIT DIRECTORS* 92 (1988) (most states deny standing to persons other than members, directors and attorneys general); *see also* 2 FURROW ET AL., *supra* note 138, § 5-18 (standing occasionally but rarely recognized for donors and others with “special interest”); *Developments in the Law-Nonprofit Corporations*, 105 HARV. L. REV. 1581, 1594-98 (1992); Manne, *supra* note 128, at 241 (“Standing limitations for nonprofit entities are grounded largely in the outdated notion of the state as *parens patriae*, and thus . . . have relegated enforcement to the exclusive province of the state.”). While some statutes and court decisions have granted standing to members and directors of not-for-profits, this adds little protection because it tends to make the goat the keeper of the cabbage patch. Rarely is standing recognized even for donors and others with a “special interest,” much less for members of the community the nonprofit serves. *See, e.g.*, *Jackson v. Stuhlfire*, 547 N.E.2d 1146 (Mass. App. 1990) (allowing members of nonprofit to bring suit); *John v. John*, 450 N.W.2d 795 (Wis. App. 1989), *cert. denied* 498 U.S. 814 (1990) (allowing directors to sue co-director).

141. Fishman, *supra* note 127, at 259-65.

problem,” not necessarily getting to root causes.¹⁴² Boards, comprised of volunteers, are notoriously risk-averse and eager to avoid sullyng their own or their institution’s standing in the community.¹⁴³ A consensus view is that applying for-profit corporate fiduciary standards to charitable corporations has proved inadequate to deter wrongdoing or to encourage responsible stewardship.¹⁴⁴ As Harvey Goldschmid stated, “[T]he central paradox of nonprofit corporate governance . . . is the fact that the nation’s nonprofit institutions are the recipients of so much public and private largess—in terms of gifts, grants, tax benefits, volunteer efforts, and other subsidies—and yet are subjected to so few accountability constraints.”¹⁴⁵ Questioning the efficacy of fiduciary law generally,¹⁴⁶ many academic commentators have proposed stricter standards for nonprofits.¹⁴⁷

By the same token, wholesale importation of for-profit corporate law gives short shrift to the nuanced role of directors of commercial

142. See *id.* at 268-69.

143. See Manne, *supra* note 128, at 245; see also Goldschmid, *supra* note 57, at 643 (citing forbearance by state regulators and understaffing as limiting enforcement of fiduciary duties).

144. JAMES J. FISHMAN & STEPHEN SCHWARZ, *NONPROFIT ORGANIZATIONS: CASES AND MATERIALS* 86 (1995); see also Demott, *supra* note 130, at 146-47 (“[I]t is foolish to import for-profit norms respecting self-dealing generally into the nonprofit context. Governance mechanisms are so much weaker in the nonprofit sector that loose controls on self-dealing create unacceptably high risks of misconduct.”); Goldschmid, *supra* note 57, at 643 (describing fiduciary standards as “aspirational” and proposing stronger enforcement); Manne, *supra* note 128, at 239 (“Much has been written about the application of fiduciary duties to directors of nonprofits, and all of it call for some reform in this area [T]he current regime is commonly viewed as inadequate.”); see also Hansmann, *Reforming Nonprofit Corporation Law*, *supra* note 4, at 568 (describing standard of conduct regarding conflicts of interest for nonprofit directors as “too weak”).

145. Goldschmid, *supra* note 57, at 632; see also Brody, *Agents Without Principals*, *supra* note 18, at 457-71; Manne, *supra* note 128, at 227-30.

146. Singer, *supra* note 68, at 237 (citing “subtle nuances and reasonable characterizations that can be attached to signing bonuses and other forms of executive compensation” that make it difficult to prove breaches of duty of loyalty).

147. DeMott, *supra* note 130, at 135-36 (noting the charitable trust model as a potential alternative to the corporate model adopted in the RMNCA); Thomas H. Boyd, Note, *A Call to Reform the Duties of Directors Under State Not-For-Profit Corporation Statutes*, 72 IOWA L. REV. 725, 744 (1987) (proposing that trustee standards should apply to public benefit nonprofits, while corporate standards should apply to mutual benefit nonprofits); see Hansmann, *Reforming Nonprofit Corporation Law*, *supra* note 4, at 570 (arguing that a strict prohibition on director self-dealing in nonprofit corporations would have “an enormously salutary effect”).

nonprofits—one that demands a balance of mission and margin. Critical to the success of any legal regime is preserving the managerial discretion necessary for the efficient operation of the nonprofit as a business enterprise.¹⁴⁸ External review imposes costs, such as increased risk aversion, transaction costs, and uncertainty in business decisions. The most obvious risk is that overly intrusive oversight may reduce efficiency, as impaired managerial discretion may constrain risk-taking and innovation.

Less widely appreciated is the danger that such reviews may pose to the corporation's charitable mission. As the Delaware Supreme Court recently acknowledged, strict application of corporate standards may be anomalous in the nonprofit setting: "Although principles of corporate law generally govern the activities of . . . a [charitable] corporation, its fiduciaries have a special duty to advance its charitable goals and protect its assets."¹⁴⁹ The threat of extensive second-guessing by regulators may tend to cause directors of charitable enterprises to substitute for their own judgments those of the regulators. When governmental actors exercise a heavy hand, they risk blurring public and private decision-making.

Finally, extensive regulatory oversight may undermine the norms that guide managers' behavior. As recent scholarship examining the role of trust and other extra-legal forces suggests, norms and other forms of social ordering that arise outside of the legal system strongly impact behavior of business managers.¹⁵⁰ There is reason to believe norms play a vital function in nonprofits: Board members and managers take their cues from their institution's mission and history and are driven by social forces such as prestige and embarrassment rather than threat of legal sanction.¹⁵¹ Yet the

148. Stephen Bainbridge summarizes the problem as follows:

Neither discretion nor accountability can be ignored, because both promote values essential to the survival of business organizations. Unfortunately, they are ultimately antithetical: one cannot have more of one without also having less of the other. Managers cannot be made more accountable without undermining their discretionary authority. Establishing the proper mix of discretion and accountability thus emerges as the central corporate governance question . . .

STEVEN M. BAINBRIDGE, *CORPORATION LAW AND ECONOMICS* 38 (2002).

149. *Oberly v. Kirby*, 592 A.2d 445, 472-3 (Del. 1991).

150. Margaret M. Blair & Lynn A. Stout, *Trust, Trustworthiness, and the Behavioral Foundations of Corporate Law*, 149 U. PA. L. REV. 1735, 1739 (2001) (trustworthy behavior helps explain the "puzzling persistence of cooperative patterns of behavior in firms in circumstances in which legal and market sanctions are ineffective or unavailable").

151. Describing the paradox of the fact that nonprofit managers tend to "adhere to good practices, and demonstrate fidelity to the organization's mission and the eleemosynary ideal" despite facing only abstract legal standards and scant enforcement, Professor

impact of legal commands is uncertain. Law may work to support social norms by its expressive effects¹⁵² or weaken them by perversely undermining their social significance.¹⁵³

B. Applying the Fiduciary Duties to Commercial Nonprofits in the Health Care Sector

Despite the inadequacies associated with applying corporate law to the nonprofit context, it has become the template for all state nonprofit statutes.¹⁵⁴ Almost every state applies the for-profit standard, rather than the more exacting trust standard, to nonprofit corporations. Somewhat startling is the fact that the special considerations raised by the non-distribution constraint and the mission of the nonprofit corporation are given only nodding recognition in statutes and case law dealing with fiduciary duties. As discussed below, a third duty, sometimes called a duty of obedience, pays some heed to directors' responsibilities to protect and promote their corporation's charitable mission. However, to date the case law governing nonprofits has failed to satisfactorily integrate the dictates of charitable responsibilities with the duties of care and loyalty imported from the for-profit corporate model. We discuss briefly the standards of the three fiduciary duties and then analyze some of the conundrums they pose for directors of nonprofit health care charities.

1. The Duty of Care

The duty of care is traditionally characterized by a three-part test

Fishman writes, "Why is the level of fidelity so high? Why do most fiduciaries do what is right? The answer may be that most charitable fiduciaries have internalized the norms of appropriate behavior. Accountability is a normative issue that reflects the role of the nonprofit sector in law and society." Fishman, *supra* note 127, at 242.

152. See Cass R. Sunstein, *On the Expressive Function of Law*, 144 U. PA. L. REV. 2021, 2024-25 (1996).

153. See Larry E. Ribstein, *Law v. Trust*, 81 B.U. L. REV. 553 (2001).

154. See REVISED MODEL NONPROFIT CORP. ACT § 8.30(a) (1987) (nonprofit duty of care and good faith); *id.* § 8.31 (nonprofit duty of loyalty). In drafting the Model Nonprofit Corporation Act, the ABA essentially used the for-profit model act as a template, electing not to employ a different approach, as recommended by some, that recognized the dramatic differences between the two corporate forms. As a result, the scarce common law that has evolved in the interim has, until recently, not developed a distinct "nonprofit" body of law.

inquiring into whether directors acted “in good faith,” with that level of care that an ordinarily prudent person would exercise in like circumstances and in a manner they reasonably believe to be in the best interest of the corporation. The seeming negligence-focused formulation, however, is mitigated by the application of the business judgment rule, which establishes a rebuttable presumption that directors who employ appropriate processes in the course of their decision-making have satisfied the duty of care. The business judgment rule essentially changes the negligence standard suggested by the technical articulation of the duty of care to one of gross negligence or recklessness by focusing on the decision-making process. Directors who make decisions that are informed, in good faith, and clear of conflicts of interest will avoid judicial scrutiny altogether.¹⁵⁵ The important caveat that the decision be the product of an *informed* judgment limits somewhat the rule’s potentially all-encompassing sweep and bars its application in situations of nonfeasance.¹⁵⁶ In addition, recent decisions by some courts suggest that boards that consciously disregard risks fail to satisfy the “good faith” requirement and will not enjoy the rule’s protection.¹⁵⁷

State courts have applied the business judgment rule to nonprofit directors, utilizing standards derived from the corporate context.¹⁵⁸

155. See generally AM. LAW INST., PRINCIPLES OF CORPORATE GOVERNANCE: ANALYSIS AND RECOMMENDATIONS § IV (Proposed Final Draft, 1992); DENNIS J. BLOCK ET AL., THE BUSINESS JUDGMENT RULE: FIDUCIARY DUTIES OF CORPORATE DIRECTORS (4th ed. 1993).

156. See *Cede & Co. v. Technicolor, Inc.*, 634 A.2d 345 (Del. 1993); *Smith v. Van Gorkum*, 488 A.2d 858 (Del. 1985); BAINBRIDGE, *supra* note 148, at 242-86 (contrasting precedent treating the business judgment rule as a substitute standard of review versus a rule of abstention).

157. See *In re Abbot Labs. Derivative S'holders Litig.*, 325 F.3d 795 (7th Cir. 2003); *In re Walt Disney Co. Derivative Litig.*, 825 A.2d 275 (Del. Ch. 2003). The Chief Justice of the Delaware Supreme Court states that the evolving standard requires “honesty of purpose and eschews a disingenuous mindset of appearing or claiming to act for the corporate good, but not caring for the well-being of the constituents of the fiduciary.” E. Norman Veasey, *Corporate Governance and Ethics in the Post-Enron WorldCom Environment*, 38 WAKE FOREST L. REV. 839, 851 (2003).

158. The Minnesota Supreme Court recently expressed concern that “[d]irectors of nonprofits may take fewer risks than would be optimal if they were overly concerned about liability for well meaning decisions.” *Janssen v. Best & Flanagan*, 662 N.W.2d 876, 883 (Minn. 2003); see also *Beard v. Achenbach Mem'l Hosp. Ass'n*, 170 F.2d 859 (10th Cir. 1948); *Scheuer Family Found., Inc. v. 61 Assocs.*, 582 N.Y.S.2d 662 (N.Y. App. Div. 1992); 2 FURROW ET AL., *supra* note 138, §5-15; FISHMAN & SCHWARZ, *supra* note 144, at 185 (describing the rule as “more appropriately known in the nonprofit context as the best judgment rule” and as providing “if a director has made a decision by informing herself in

Although common law rarely addresses explicitly the propriety of applying the business judgment rule to nonprofit corporations, those courts that have faced the question have accepted the rule.¹⁵⁹ In reality, however, despite the lofty standard of diligence provided by statutory and common law formulations, the duty of care very rarely results in courts imposing sanctions upon directors.¹⁶⁰ The business judgment rule protects almost all judgments by directors as long as they are plausibly “informed.” Some scholars have sought to explain this “schizophrenic” state of affairs by stressing the central role of trust in shaping behavior and suggesting that the law may reinforce trustworthy behavior by influencing the internal preferences of actors in contrast to affecting the external incentives they encounter.¹⁶¹

2. *The Duty of Loyalty*

In the business corporation context, the duty of loyalty flows from the directors’ duty to maximize shareholder wealth. This philosophical underpinning poses obvious difficulty for application to the nonprofit corporation, which does not have shareholders, whose legal form rests on a commitment to a charitable enterprise, and whose mission therefore is not

good faith without a disabling conflict of interest, there will be neither judicial inquiry nor liability even if the action was unfortunate for the organization or its membership.”); MICHAEL W. PEREGRINE & JAMES R. SCHWARTZ, *THE APPLICATION OF NONPROFIT CORPORATION LAW TO HEALTH CARE ORGANIZATIONS* 44-45 (2002).

159. *Janssen v. Best & Flanagan*, 662 N.W.2d 876, 883 & n.2 (Minn. 2003) (noting it found no case rejecting the business judgment rule in the nonprofit context, and that the Supreme Courts of Alabama, Hawaii, and South Dakota, as well as intermediate appellate courts of Colorado, New York, Ohio, South Carolina, Tennessee, and Wisconsin have applied the business judgment rule to nonprofit boards); *see also Beard*, 170 F.2d 859; *Woo Chul Lee v. Interinsurance Exch.*, 50 Cal. App. 4th 694 (Cal. Dist. Ct. App. 1996); *Oberly v. Kirby*, 592 A.2d 445, 462 (Del. 1991) (“A court cannot second-guess the wisdom of facially valid decisions made by charitable fiduciaries, any more than it can question the business judgment of the directors of a for-profit corporation.”); *Scheuer Family Found.*, 582 N.Y.S.2d at 662.

160. In cases involving for-profit corporations, under the business judgment rule the standard of care is almost uniformly applied only to review the process by which decisions are made, not the result. In only a handful of cases have courts found directors liable under this standard, and few, if any, find liability for even egregious mistakes in judgment. *See Charles Hansen, The ALI Corporate Governance Project: Of the Duty of Care and the Business Judgment Rule, A Commentary*, 41 BUS. LAW. 1237 (1986). *See generally* Melvin Aron Eisenberg, *The Director’s Duty of Care in Negotiated Dispositions*, 51 U. MIAMI L. REV. 579 (1997).

161. Blair & Stout, *supra* note 150, at 1737-38.

primarily wealth maximization. The question then becomes whether the corporate notion of the duty of loyalty can be usefully reformulated to ensure appropriate director behavior and preservation of the charitable mission in the nonprofit context.¹⁶²

The duty of loyalty also governs the individual board member's relations with the corporation of which she is a director. Interestingly, neither courts nor legislatures have interpreted the duty of loyalty in the for-profit context as prohibiting outright self-dealing and other conflicted interest transactions.¹⁶³ In general, the law prohibits only those self-dealing transactions that are not approved or ratified by the board of directors or shareholders under specified standards. In the business corporation context, approval may be gained by the vote of a disinterested majority of the board of directors or by a majority of disinterested shareholders provided the terms of the transaction are fully disclosed prior to the vote.¹⁶⁴ In the nonprofit context, most states appear to have applied the business corporation standard in addressing the duty of loyalty,¹⁶⁵ although some states have imposed somewhat more stringent standards for self-dealing transactions.¹⁶⁶

3. *The Duty of Obedience*

A third duty, applicable only to the directors of nonprofit corporations, is the duty of obedience. Although articulated as a distinct

162. The case law using duty of loyalty in this way is virtually non-existent. *But see* Summers v. Cherokee Children & Family Servs., Inc., 112 S.W.3d 486, 504-31 (2002) (describing the duty of loyalty as one intended to ensure effective performance of mission by directors).

163. Neither wants to deny the nonprofit entity the potential of a board member facilitating beneficial contracts or business relationships for the nonprofit. KURTZ, *supra* note 140, at 60-61, 63.

164. *See, e.g.*, DEL. CODE ANN. tit. 8, § 144 (2001); REVISED MODEL BUS. CORP. ACT § 8.31 (1983); Cede & Co. v. Technicolor, Inc., 634 A.2d 345, 365 (Del. 1993) (characterizing Delaware statute as "a legislative mandate that . . . an approving vote of a majority of informed and disinterested directors shall remove any taint of director or directors' self-interest in a transaction").

165. *See* REVISED MODEL NONPROFIT CORP. ACT § 8.31 (1987); *see also, e.g.*, 15 PA. CONS. STAT. ANN. § 5715 (West 1995); S.C. CODE ANN. § 33-31-830(a)(3) (Law. Co-op. Supp. 2003).

166. *See, e.g.*, CAL. CORP. CODE § 5233 (West 2004) (requiring that transactions be fair and reasonable at the time entered into and approved by a majority of the board, which must consist entirely of disinterested members; it must also be shown that the board determined, after reasonable inquiry, that a more advantageous deal could not be obtained).

fiduciary duty in only a handful of cases,¹⁶⁷ the concept appears to have been broadly popularized by the work of Daniel Kurtz,¹⁶⁸ presumably to overcome the perceived deficiencies of applying the duty of loyalty to the nonprofit corporate board. Broadly construed, the duty of obedience expresses the obligation of nonprofit directors to observe and advance the mission of the charitable corporation by adhering to its purposes, usually as set forth in the entity's articles of incorporation or bylaws. However, in the few instances in which it is specifically mentioned by courts, it has been invoked to indicate directors' responsibility to assure that their corporations obey the law and not stray from the dictates of the purposes expressed in their articles and bylaws.¹⁶⁹ In various cases in which courts have dealt with nonprofit hospitals seeking to change their business to provide health care services other than acute care, the concept limits such changes unless permitted by the corporation's articles.¹⁷⁰

C. Practical Problems with Applying Fiduciary Duties to Nonprofit Boards and Managers

Even in the for-profit context, the efficacy of common law and statutory duties in ensuring that directors meet their fiduciary duties is the subject of considerable debate. A raft of studies examining the failures of oversight in the wake of Enron, WorldCom, and other corporate scandals

167. See, e.g., *Manhattan Eye, Ear & Throat Hosp. v. Spitzer*, 715 N.Y.S.2d 575 (Sup. Ct. 1999).

168. See KURTZ, *supra* note 140, at 84-85.

169. See FISHMAN & SCHWARZ, *supra* note 144, at 230 ("The duty of obedience resembles the trustees' duty to administer a trust in a manner faithful to wishes of creator Thus, the director has a duty to follow the purposes and powers as expressed in the [corporation's] governing legal documents." (citation omitted)). The duty of obedience is regarded by some commentators as a particularized obligation under the duty of loyalty or care. See e.g., Goldschmid, *supra* note 57, at 650.

170. For example, in *Queen of Angels Hosp. v. Younger*, 136 Cal. Rptr. 36 (Ct. App. 1977), a religious order sought to close a hospital in order to provide health services to the indigent through outpatient neighborhood clinics. Even though the articles of incorporation indicated several purposes, the court interpreted them to require continuing operation of a hospital. *Id.* at 40-41. In *Attorney Gen. v. Hahnemann Hosp.*, 494 N.E.2d 1011, 1021 (Mass. 1986), trustees of a hospital sought to sell its assets in order to become a grant-making institution for health care institutions. The Court allowed the trustees to amend the articles to do so but noted without such provision they would have violated fiduciary duties. *Id.* at 1018-19.

points to the insufficiency of director oversight in the for-profit sector.¹⁷¹ The emerging consensus that fiduciary duties are no substitute for other means of assuring honesty and diligence by corporate managers would seem to apply a fortiori to nonprofit entities.¹⁷² Adding to the problems

171. See, e.g., First Interim Report of Dick Thornburgh, Bankruptcy Court Examiner, *In re Worldcom Inc.*, No. 02-15533(AJG) (Bankr. S.D.N.Y. Nov. 4 2002), <http://news.findlaw.com/hdocs/docs/worldcom/thornburgh1strpt.pdf> (citing . . . “numerous failures, inadequacies and breakdowns” among the “Board of Directors, the Audit Committee, the Company’s system of internal controls and the independent auditors.”); AM. BAR ASS’N TASK FORCE ON CORPORATE RESPONSIBILITY, REPORT OF THE AMERICAN BAR ASSOCIATION TASK FORCE ON CORPORATE RESPONSIBILITY 25, 29 (2003) (finding outside directors “overly dependent upon and overly passive with respect to senior executive officers” and recommending that “outside directors abandon the passive role many have been content to play and replace it with a new culture stressing constructive skepticism and an active, independent oversight role”); WILLIAM C. POWERS, JR. ET AL., SPECIAL INVESTIGATIVE COMMITTEE OF THE BOARD OF DIRECTORS OF ENRON CORP., REPORT OF INVESTIGATION 148 (2002), <http://i.cnn.net/cnn/2002/LAW/02/02/enron.report/powers.report.pdf> (describing oversight by Enron’s Board and Management as “cursory”; stating that Board “did not fully appreciate the significance of some of the significant information that came before it”; and characterizing controls put in place governing self dealing as inadequate); see also William H. Donaldson, Chairman, U.S. Securities and Exchange Commission, Remarks at the 2003 Washington Economic Policy Conference before the National Association for Business Economics (Mar. 24, 2003), <http://www.sec.gov/news/speech/spch032403whd.htm> (“[I]nattention to good corporate governance practices over the past decade or more is at the heart of what has gone so terribly wrong in corporate America in the past few years [A]t too many companies, the chief executive position has steadily increased in power and influence. In some cases, the CEO had become more of a monarch than a manager. Many boards have become gradually more deferential to the opinions, judgments and decisions of the CEO and senior management team. This deference has been an obstacle to directors’ ability to satisfy the responsibility that the owners—the shareholders—have delegated and entrusted to them.”).

172. Besides suggesting that the potential for abuse was larger than previously suspected, the well-documented shortcoming in the for-profit sector is all the more startling because of the enormous phalanx of analysts and experts that monitor the securities markets and institutional investors ostensibly monitoring directors’ behavior. The lack of comparable watchdogs in the nonprofit sector suggests that directorial abuse might be even harder to detect. Further, extensive corporate scholarship identifies a number of factors, all applicable to nonprofit boards, which impair effective director oversight. For example Professors Bebchuk, Fried and Walker persuasively explain excesses in executive compensation by demonstrating the subtle conflicts that arise out of managers’ influence over the appointment of directors, the effects of board decision-making dynamics, and the impact of directors’ lack of independently supplied information. See Lucian Arye Bebchuk et al., *Managerial Power and Rent Extraction in the Design of Executive Compensation*, 69 U. CHI.

associated with relying on fiduciary principles are multiple administrative limitations facing state charitable enforcers.¹⁷³ Attorneys general lack resources, access to information, and expertise to effectively monitor conduct of the extensive and economically significant commercial nonprofit sector.¹⁷⁴

The numerous examples of abject breaches of oversight responsibilities by directors of major commercial health care charities raise serious doubt as to how effectively the fiduciary duties serve their prophylactic function of averting abuse and encouraging director vigilance. For example, as described in Subsection II.A, in a number of high profile conversions of nonprofit health plans in the early 1990s, insiders personally profited from lucrative arrangements and sales that took place for vastly undervalued amounts, resulting in losses of billions of dollars of charitable assets.¹⁷⁵ The fact that few, if any, of the directors involved in these cases were held to account under fiduciary theories confirms the view that the duties are “relatively weak weapon[s] in the

L. REV. 751 (2002); William W. Bratton, *Enron and the Dark Side of Shareholder Value*, 76 TUL. L. REV. 1275 (2002).

173. Although the Delaware courts have recently signaled an intention to apply the requirement of good faith more aggressively, that standard nevertheless requires a showing that directors “*consciously and intentionally disregarded their responsibilities*, adopting a ‘we don’t care about the risks’ attitude” *In re Walt Disney Co. Derivative Litig.*, 825 A.2d 275, 289 (Del. Ch. 2003); *see also In re Abbot Labs. Derivative Litig.*, 325 F.3d 795, 807-11 (7th Cir. 2003) (finding absence of good faith where directors were aware of extensive safety problems leading to large civil fine and took no efforts to remedy them).

174. James Fishman has catalogued a number of deficiencies including the fact that few states even have charities sections within the attorneys general office, staffing is minimal, and responsibilities are often divided with other agencies in a way that impairs effective oversight. Fishman, *supra* note 127, at 262-63. In addition, attorneys general lack staff to efficiently review information provided in mandatory reporting such as Form 990 and are unable to effectively share information with IRS or other state enforcers. *Id.* at 263-65.

175. Examples are legions of vastly underpriced sales of assets of nonprofits, often engineered by insiders who ultimately profited by stock ownership in or lucrative employment agreements with the purchaser. *See, e.g., Colombo, supra* note 96, at 785 (estimating actual value of assets of California’s Health Net HMO to be approximately 500% higher than originally estimated and describing funding of charitable foundations on conversion of PacifiCare Health Systems at less than 1% of actual value of the enterprise). For a detailed account of the numerous instances of under-valuation in such conversions and the successful efforts of the Consumers Union to have hundreds of millions of dollars turned over to independent foundations, *see Eleanor Hamburger et al., The Pot of Gold: Monitoring Health Care Conversions Can Yield Billions of Dollars for Health Care*, 29 CLEARINGHOUSE REV. 473 (1995).

arsenals of attorneys general”¹⁷⁶ for dealing with such problems. Although subsequent intervention by consumer groups caused attorneys general to review and ultimately challenge some (but far from all) of these transactions, few if any directors were personally prosecuted. Responding to the problem posed by conversions, by 1998 over twenty-five states and the District of Columbia had adopted legislation.¹⁷⁷ However, most of the nonprofit conversion statutes do not change the substantive standard for review of fiduciary breaches.¹⁷⁸

As an example of the problems associated with relying on fiduciary law to police nonprofit governance, consider the complete breakdown in governance that was central to the demise of AHERF, discussed in Subsection II.A. Several careful studies of AHERF place prime responsibility on its boards for effectively ceding governance to the CEO and accepting a model of corporate control designed to prevent effective oversight. Multiple and overly large boards thwarted effective discussion or analysis of corporate policy and the CEO’s domination of the board (through selection process and personal ties) discouraged any meaningful board input.¹⁷⁹ When one holds this framework up against the lenient standard widely applied under the duty of care, however, it is entirely possible that the board members might have avoided personal liability.¹⁸⁰

176. Singer, *supra* note 68, at 237.

177. Christopher W. Frost, *Financing Public Health Through Nonprofit Conversion Foundations*, 90 Ky. L.J. 935, 953 (2001-2002).

178. See generally MODEL ACT FOR NONPROFIT HEALTHCARE CONVERSIONS (1998), *reprinted* in GREANEY & SCHWARTZ, *supra* note 63; FISHMAN & SCHWARZ, *supra* note 144, at 185; Kevin F. Donohue, *Crossroads in Hospital Conversions—A Survey of Nonprofit Hospital Conversions*, 8 ANNALS HEALTH L. 39 (1999).

179. Burns, et al., *supra* note 72, at 21-22; see also Michael W. Peregrine & James R. Schwartz, *Revisiting the Duty of Care of the Nonprofit Director*, 36 J. HEALTH L. 183, 201 (2003).

180. A recent account by one of AHERF’s insiders that is highly critical of top management explains that the Boards were supplied with extensive information, perhaps so much so that they were unable to digest it and properly supervise management.

It might be reasonable to suppose that the [AHERF] trustees were unable to see and perhaps prevent the oncoming train wreck because they were underinformed. In fact, nothing could be further from the truth. The trustees of the constituent corporations of HERF and of AHERF itself were regularly given reams of information, including detailed financial statements. Although it would have taken a reader of financials with extraordinary insight to discern from one of the constituent corporation’s statements how all of AHERF was doing, there was enough crossover on the various boards that there was a core of trustees who had most if not all of the relevant information available to them. The more likely scenario, in fact is that the trustees had too much information; they were given so much to absorb that they could not winnow out what was important.

Even with such a remarkable record of inattentiveness, the business judgment rule may have afforded protection, as it requires only that directors be reasonably informed. Moreover, to pass the process-oriented information hurdle, directors can rely on ostensibly trustworthy surrogates to supply expertise and evaluation. Assuming the AHERF boards were reasonably attentive to information placed before them and relied on the advice of executives and responsible intermediaries, there is every likelihood that their conduct would enjoy the protection of the business judgment rule.¹⁸¹

The ineffectiveness of the fiduciary duties in policing board behavior has spurred charitable regulators to invoke charitable trust law to supply a more stringent standard of conduct. For example, the Attorney General of Minnesota's business compliance reviews of the Allina Health System¹⁸² and HealthPartners¹⁸³ examined in extraordinary detail the day-to-day business decisions of those companies.¹⁸⁴ Attorney General Hatch claimed that the

ANDREW E. THURMAN, *INSIDE AHERF: LESSONS LEARNED*, <http://www.thurmanhealthlaw.com/INSIDE%20AHERF.doc> (last visited Mar. 16, 2004).

181. Ultimately, the AHERF president "pleaded no contest to a single misdemeanor count of misusing charitable funds by virtue of having diverted endowment funds of a hospital to finance the organization's operating costs." FREMONT-SMITH & KOSARAS, *supra* note 56, at 9-10. The AHERF CFO pled to a single misdemeanor and paid a small fine. Three senior AHERF financial executives, including the CFO entered into civil consent decrees with the United States Securities and Exchange Commission and, without admitting wrongdoing, paid fines. THURMAN, *supra* note 180, at 1.

182. See Press Release, Minnesota Attorney General's Office (Sept. 24, 2001), http://www.ag.state.mn.us/consumer/PR/pr_allina_mou_92401.htm; see also Vince Galloro, *Watch It! Attorneys General Become More Active as Healthcare Finances Grab Public Eye*, MOD. HEALTHCARE, Aug. 13, 2001, at 16 (describing fourteen-month investigation of Allina and revelations that its HMO subsidiary spent \$56 million on consultants over three year period and "coached executives through team-building exercises, such as playing ring toss, and showed movies to teach . . . officials about group dynamics"); *supra* Subsection II.B.1.

183. See *In re HealthPartners, Inc.*, No. MC 03-001587 (Minn. Dist. Ct. Prob. Div. June 10, 2003) (stipulation and order for the appointment of a Special Administrator).

184. The Attorney General determined management's expenses, travel and executive compensation to be "lavish," to the point of deeming inappropriate a room service charge for breakfast while attending a conference where a continental breakfast was available. The Attorney General's findings of inappropriate luxuries may be found at Minn. Att'y Gen, Summary of Executive Compensation Expenses, http://www.ag.state.mn.us/consumer/PDF/HealthPartners_ExecComp_pdf (last visited Nov. 17, 2004); Minn. Att'y Gen, Summary of HealthPartners Consulting Expenses, http://www.ag.state.mn.us/consumer/PDF/HealthPartners_Consulting_Expenses.pdf (last visited Nov. 17, 2004); and Minn. Att'y Gen, Chapter I: Travel and Entertainment, <http://www.ag.state>.

boards had failed in their oversight of senior management and had neglected their responsibilities to exercise independent judgment. Citing a variety of “governance failures” by the HealthPartners’s Board, the Attorney General petitioned for the appointment of two “special administrators” to act as trustees of the HealthPartners charitable trust.”¹⁸⁵

The Attorney General’s legal theory rested on an amalgam of charitable trust and corporate law. Its legal brief asserted that Minnesota law subjects nonprofit board members to charitable trust standards by virtue of the fact that the corporation holds charitable assets. It charged that poorly documented or excessive expenditures “waste[d]” corporate assets and ineffective oversight breaches directors’ fiduciary duties.¹⁸⁶ Rather confusingly, the Attorney General cited the duties of care and obedience from nonprofit *corporate* law for these propositions along with conclusory statements that the stricter charitable trust standard should apply.¹⁸⁷ It is highly doubtful that a court would find a breach of fiduciary duty under the nonprofit corporate standard in these circumstances (the court never reached the question of whether charitable trust law could be imported to supply a stricter standard).¹⁸⁸ The corporate waste doctrine is exceedingly difficult to satisfy¹⁸⁹ and, as we have seen, duty of care claims

mn.us/consumer/PDF/HealthPartners_Travel_Entertainment.pdf (last visited Nov. 17, 2004). See also Brody, *Whose Public?*, *supra* note 13, at 1005.

185. Brody, *supra* note 13, at 1005.

186. Michael Peregrine & James Schwartz, *Key Nonprofit Law Developments in 2003*, 13 BNA HEALTH L. REP. 128, 130 (2004).

187. PEREGRINE & SCHWARTZ, *supra* note 158 (citing *Manhattan Eye, Ear & Throat Hosp. v. Spitzer*, 715 N.Y.S.2d 575, 593 (Sup. Ct. 1999) for the duty to ensure “that the mission of the charitable corporation is carried out”). It also cites commentary summarizing directors’ general fiduciary duties under nonprofit corporate law, e.g., KURTZ, *supra* note 140 (“[D]irectors should be diligent and attentive.”), but makes no reference to the business judgment rule. *Id.*

188. Commentators have sharply questioned whether theories of corporate waste or breach of fiduciary duty can be brought in instances of director nonfeasance such as HealthPartners petition. See PEREGRINE & SCHWARTZ, *supra* note 158, at 26-27.

189. Under Delaware law, “waste entails an exchange of corporate assets for consideration so disproportionately small as to lie beyond the range at which any reasonable person might be willing to trade.” *Lewis v. Vogelstein*, 699 A.2d 327, 336 (Del. Ch. 1997) (citing *Grobow v. Perot*, 539 A.2d 180, 189 (Del. 1988)); see also *Saxe v. Brady*, 184 A.2d 602, 610 (Del. Ch. 1962). This extraordinarily high standard of proof has led some courts to doubt it can ever be met absent proof of self dealing. *Steiner v. Meyerson*, No. Civ. A. No. 13139, 1995 WL 441999, at *5 (Del. Ch. July 19, 1995) (“But rarest of all—and indeed, like Nessie, possibly nonexistent—would be the case of disinterested business people making non-fraudulent deals (non-negligently) that meet the legal standard of

are easy defended by invoking the business judgment rule.

The doctrinal and policy flaws of borrowing the charitable trust standard are discussed in Part IV. However, two important collateral aspects of attorney general activism in the face of doctrinal uncertainty in this area should be noted. As discussed in Part II, one highly controversial aspect of *Allina* was the relief obtained by the Minnesota Attorney General—a spin off of the HMO subsidiary. As troubling, however, was the Attorney General's petition for authority to select eight of the special administrators who were to serve as the new entity's board. He sought this same power to appoint directors, first informally and later with court approval, in the HealthPartners case (ultimately the probate division of the district court ordered that one be appointed special administrator with responsibilities to report instances of board failure to act in good faith).¹⁹⁰ The state's attempt to substitute its own decision-makers for the directors of the nonprofit corporation does considerable violence to the independence of the nonprofit sector. The problems associated with this intervention go beyond the merits of the claimed failures of the current board. The threat of direct intervention by politically-selected regulators blurs the line between public and private.¹⁹¹ If not reserved for instances of outright corruption, the power to replace decision-makers may be too potent a weapon to entrust to courts, especially when attorneys general may accomplish the result by the mere threat of seeking judicial relief.

D. The Misuse or Neglect of Mission in Analyzing Directors' Fiduciary Duties

Finally, we consider the curiously neglected role of institutional mission in informing directors' duties. Conversions, closures, asset sales, and other organic changes involving nonprofit corporations require directors' most assiduous adherence to their fiduciary duties. Fiduciary questions arise in many contexts, including whether the conversion or change of purpose is consistent with the purpose of the nonprofit organization; whether the purchaser is appropriate in view of the entity's charitable purposes; whether directors approving the decision resolved

waste!").

190. *In re HealthPartners, Inc.*, No. MC 03-001587 (Minn. Dist. Ct. Prob. Div. June 10, 2003) (stipulation and order for the appointment of a Special Administrator).

191. Brody, *Whose Public?*, *supra* note 13, at 942. Further, some have inferred political impropriety in elected officials' appointment of some special administrators or board members. See Strom, *supra* note 92 (raising concerns about politicians' appointment of friends, colleagues and political contributors, and quoting experts who suggest that such appointments are more appropriately made by courts).

conflicts of interest; and how the directors decided to use the assets acquired by the nonprofit corporation. Though these questions would seem to force regulatory agencies and courts to confront directly the role of mission in nonprofit corporate decision-making, the standard to be applied remains muddled. Arising in different statutory contexts, and often applying different substantive standards, the case law presents a less than uniform picture. Three approaches can be discerned in the case law: the pure corporate law standard; the mixed corporate/nonprofit mission standard; and the corporate/regulatory policy standard. None, however, offers a coherent formula for accommodating mission values into the fiduciary duties of directors.

1. *Pure Corporate Standard*

In a number of cases, courts confronting organic changes have purported to apply the corporate fiduciary standard in reviewing directors' decisions. For example, in *Health Midwest*, the court declared unequivocally "corporate law applies to all aspects of this transaction."¹⁹² In so doing, it declined to apply the Kansas *cy pres* statute to the transaction, finding that the statute did not apply to changes in corporate purposes.¹⁹³ Applying the corporate standard in a straightforward fashion, it went on to hold that the business judgment rule required deference to the board's decision to convert, its choice of a buyer, and its evaluation of an appropriate sales price. Likewise, it summarily rejected a challenge to executive compensation arrangements for executives involved in the transaction.¹⁹⁴ However, despite its invocation of a pure corporate standard, the *Health Midwest* court could not resist invoking mission-related obligations in reviewing one financially important (and parochial) aspect of the board's decision. The Kansas court struck down the board's decision to pay the sale

192. *Health Midwest v. Kline*, No. 02-CV-08043, 2003 WL 328845, at *2 (D. Kan. Feb. 6, 2003) (citing *United Methodist Church v. Bethany Med. Ctr.*, 969 P.2d 859 (1998)); *see also* REVISED MODEL NONPROFIT CORP. ACT § 8.30 (1987).

193. *Health Midwest*, 2003 WL 328845, at *19 ("The Kansas *cy pres* statute governs changes to the purposes of charitable trusts, devises and bequests. The *cy pres* statute does not apply to changes to the purposes of nonprofit corporations. The *cy pres* statute applies only to any restricted gifts and not the entity as a whole. No restricted gifts have been identified herein and therefore the *cy pres* statute does not apply." (citations omitted)).

194. The court observed that "[t]he appropriateness of the packages (even though they appear on their face to be excessive) has no bearing in regard to whether the Agreement should be approved. *Health Midwest's* decision to approve the compensation is an internal matter of the Missouri company and is subject to review by a Missouri court." *Id.* at *19.

proceeds into a Missouri foundation, noting that the board had elsewhere concluded that twenty percent of Health Midwest's assets had previously served Kansas residents. The court made no effort to explain why corporate law analysis including the business judgment rule did not mandate judicial abstention here, except to suggest rather obliquely that mission factors compelled the result.¹⁹⁵

2. *Mixed Corporate/Nonprofit Mission Standard*

Some courts have more explicitly weighed mission responsibilities in interpreting nonprofit directors' fiduciary duties. For example, in *MEETH* the court invoked the duty of obedience to buttress its conclusions that the hospital directors had neglected their obligation to fully consider all options for avoiding closure of the hospital and had not received adequate value in the sale of its assets.¹⁹⁶ Yet, the court gave little deference to the judgment of the directors and in fact never mentioned the business judgment rule in reviewing the merits of the decision to "monetize the assets" of MEETH for use in what the directors regarded as a more needed and financially viable charitable use. Nor did it explain how the Board was to go about weighing mission and business responsibilities. Similarly, in *Queen of Angels*, the court was willing to override the business judgment of the hospital board where it interpreted the nonprofit corporation's mission as commanding unwavering allegiance to the continued operation

195. The court stated:

The attorney general . . . has persuaded the Court that the decision to merge into a Missouri Foundation is a "perversion of corporate purpose" and that the Kansas boards have neglected their duties to the communities in their service areas and have breached the trust placed in them. The announced foundation plan does not confirm that Health Midwest's Kansas subsidiaries' historic charitable purposes will remain intact following the transaction.

Id. at *26.

196. The court noted:

It is axiomatic that the Board of Directors is charged with the duty to ensure that the mission of charitable corporation is carried out. This . . . "duty of obedience" . . . requires the director of a not-for-profit corporation to be "be faithful to the purposes and goals of the organization," since "[u]nlike business corporations, whose ultimate objective is to make money, nonprofit corporations are defined by their specific objectives: perpetuation of particular activities are central to the raison d'être of the organization."

Manhattan Eye, Ear & Throat Hosp. v. Spitzer, 715 N.Y.S.2d 575, 593 (Sup. Ct. 1999) (alteration in original) (citation omitted).

of a hospital.¹⁹⁷

3. Corporate/Regulatory Policy Standard

Perhaps the most confusing analysis of mission is found in the regulatory context. In its evaluation of the conversion and sale of CareFirst to WellPoint Health Networks, the Maryland Insurance Commissioner applied a multi-pronged regulatory standard to determine whether the transaction satisfied the statute's broad public interest criteria.¹⁹⁸ The statute gives the Commissioner broad discretion to determine what constitutes due diligence, setting forth eight criteria that may be brought to bear. As discussed above,¹⁹⁹ the Commissioner's decision at times invoked for-profit fiduciary standards and at others rejected them. Indeed, in one passage, the opinion considered a particularly rigid application of the for-profit standard, inquiring whether the directors of CareFirst had an obligation to accept the highest bid and thus were bound to ignore mission-based considerations in selecting a buyer. Under for-profit corporate law in many states, the *Revlon* doctrine obligates fiduciaries to act as a broker and accept the highest bid, once the decision to sell is definitive.²⁰⁰ While application of this rule to nonprofits would be controversial, strict application of corporate fiduciary standards might suggest that in a change of control transaction, a nonprofit board is bound to opt for the best financial offer even though another bidder may offer nonfinancial terms more in keeping with the mission of the nonprofit corporation.²⁰¹ Although special counsel vigorously supported applying the

197. *Queen of Angels Hosp. v. Younger*, 136 Cal. Rptr. 36, 41 (1977).

198. The *CareFirst* decision is discussed *supra* notes 117-126 and accompanying text; see CAREFIRST CONVERSION INFORMATION, *supra* note 117, at 5-7.

199. See *supra* notes 117-26 and accompanying text.

200. *Revlon Inc. v. MacAndrews & Forbes Holdings Inc.*, 506 A.2d 173, 182 (Del. 1985) (holding that once the board of a target company of a takeover bid "no longer faced threats to corporate policy and effectiveness, or to the stockholders' interests," their role "changed from defenders of the corporate bastion to auctioneers charged with getting the best price for the stockholders at a sale of the company.").

201. In change-of-control transactions, the nonprofit board may seek to achieve nonfinancial objectives. For example it may wish to obtain "capital improvement commitments, access to acute care commitments, preservation of workforce, and preservation of employee benefits." Peregrine & Schwartz, *supra* note 179, at 199. For an argument in favor of applying *Revlon* to nonprofits, see Colin T. Moran, *Why Revlon Applies to Nonprofit Corporations*, 53 BUS. LAW. 373 (1998).

Revlon Rule, the Maryland Insurance Commissioner concluded it did not need to decide the issue as it found the director's conduct wanting for failure to exercise "due diligence."

In other passages, the opinion departed sharply from for-profit fiduciary principles. For example, acknowledging that courts employ the business judgment rule in cases involving directors' breach of the duty of care, the Commissioner announced that the presumption did not apply in a regulatory context.²⁰² The opinion deemed the business judgment rule inapposite in an administrative proceeding governed by a broad regulatory mandate. Thus, the Commissioner concluded he had broad latitude to conduct his own *de novo* review of whether a transaction is in the public's interest.

IV. THE ELUSIVE SLIDE FROM A FIDUCIARY TO CHARITABLE TRUST STANDARD

Part III establishes that the strict importation of for-profit corporate law principles and applying mixed for-profit and nonprofit mission or regulatory policy standards is inefficacious in the nonprofit health care enterprise context. This Part argues that the invocation of charitable trust principles, either directly or implicitly, is fundamentally unsound. It contends, first, that doctrinal developments militate strongly against applying charitable trust standards except where an express trust exists. Although some states have chosen to buck the trend and retain a broad charitable trust standard for their nonprofits, courts and attorneys general should take care to recognize that those are *sui generis* cases owing to their statutory law. Further, there is no reason to believe that these states' approaches advance sound public health care policy.

Next we argue that conceptually, charitable trust law, which assumes an identifiable settlor, beneficiaries, and trust purpose, is ill-suited to the nonprofit corporation. We also find that in stretching the law governing charitable trusts beyond recognition, attorneys general have undertaken a wholly impractical and *ad hoc* course. There are reasons to believe that rigid application of charitable trust principles will undermine sound health policy aimed at maintaining a health care delivery system sufficient to meet the nation's needs. By the same token, these efforts make it impossible for nonprofit boards to have any clear sense of what power they have to direct the corporate mission in a way that is market-responsive, or to deploy assets consistent with a long-term strategic plan.

202. See *supra* note 124 and accompanying text.

Finally, we conclude that by blending charitable trust and corporate fiduciary law in their oversight of nonprofit board decision-making, attorneys general and other charity regulators have opportunistically capitalized on doctrinal confusion in this area. While acknowledging that corporate law requires some development to regulate the nonprofit sector well, we conclude that it is unquestionably the better doctrinal starting point. Specifically, nonprofit corporate doctrine should explicitly recognize the centrality of mission to the charitable enterprise, and presume that directors are best situated, at least in the first instance, to advance the corporation's mission. Recognition of directors' superior expertise and dedication to mission preservation would hopefully ameliorate the trends described in this Article that counter policies uniquely important in the health care sector and that may have a particular deleterious impact on long-term access to appropriate health care in many communities. That is, by inappropriately interfering with directors' responsibility to balance mission and margin, the vibrant and creative impact of the health care sector may be seriously impeded. Further, it may hamper the efficient reorientation of segments of the sector, such as redeployment of charitable assets and conversion to for-profit form. Finally, by broadly invoking various policy concerns that implicate federal tax law, state licensure and access statutes, and health care fraud law in their state law analyses, attorneys general usurp power, distort policy, and subject entities to inconsistent application of these laws.

A. *The Impact of the Adoption of Modern Nonprofit Statutes*

Approximately twenty-nine states have adopted all or part of the Revised Model Nonprofit Corporation Act (RMNCA) or its predecessor.²⁰³ In retrospect, it seems surprising that the RMNCA does not more helpfully address the issues associated with the most distinctive aspect of the charitable corporation, its nonprofit mission. Like most nonprofit statutes, it requires that a nonprofit corporation have a public benefit, religious, or mutual benefit purpose.²⁰⁴ At the same time, most states adopting modern nonprofit statutes are relatively clear that corporate law standards generally apply in these matters. Problems arise, however, because the RMNCA and most state nonprofit acts do not address the extent to which public benefit,

203. The original model act was adopted in 1942, but did not address directors' duties; the revised model act was adopted in 1987. See James Edward Harris, *The Nonprofit Corporation Act of 1993: Considering the Election To Apply the New Law to Old Corporations*, 16 U. ARK. LITTLE ROCK L. REV. 1, 3 n.11 (1994).

204. REVISED MODEL NONPROFIT CORP. ACT § 2.02(a)(2) (1987).

mutual benefit, and religious purposes may alter the frame of analysis applied by directors in exercising their fiduciary duties or by courts in assessing their conduct.²⁰⁵ As we have seen, this gap has been only episodically addressed by courts and has invited attorney general activism in the form of transporting charitable trust law to fill the void. It should be noted that a few states, such as Illinois, New Hampshire, and Virginia, have gone in an entirely different direction, enacting statutes that explicitly impose a charitable trust upon the property of nonprofit corporations. While this approach unquestionably gives courts and attorneys general clear and significant authority over mission decisions by nonprofit boards, the law of other states should not be read to vest such discretion. We survey and analyze below the state statutory approaches to the issue.

1. *Model Nonprofit Corporation Act States*

Most states apply corporate law principles to charitable corporations, either by judge-made law or the adoption of all or part of the Model Nonprofit Corporation Act.²⁰⁶ This “modern trend” of significantly

205. Mission is little addressed by either the Model Act or the common law. It has long been assumed that a board may alter its mission by amending its articles of incorporation. The process is rather uncomplicated; the Revised Model Act provides that a “corporation may amend its articles of incorporation at any time to add or change a provision that is required or permitted in the articles or to delete a provision not required in the articles.” *Id.* § 10.01. Nowhere is it suggested that such amendments may not affect the corporate purposes.

206. The following states’ nonprofit corporate statutes are based upon the Model Act adopted in 1964: ALA. CODE §§ 10-3A-1 to -225 (1999); ALASKA STAT. § 10.20.005 (Michie 2002); ARIZ. REV. STAT. ANN. §§ 10-3301 to -3304 (West 2004); CAL. CORP. CODE §§ 5510–27 (West 1990); COLO. REV. STAT. §§ 7-123-101 to -137-204 (1999); D.C. CODE ANN. §§ 29-301.01 to -321.01. (2001); GA. CODE ANN. §§ 14-3-101 to -1703 (Harrison 2003); 805 ILL. COMP. STAT. 105/101.01- 105/101.80 (2004); IOWA CODE ANN. §§ 504A.1- .101 (West 1999); KY. REV. STAT. ANN. §§ 273.070- .991 (Michie 2003); ME. REV. STAT. ANN. tit. 13B, §§ 101-110 (West 1981); MINN. STAT. §§ 317A.001- .909 (2004); MO. REV. STAT. §§ 355.001- .881 (2001); MONT. CODE ANN. §§ 35-2-113 to -1402 (2003); NEB. REV. STAT. §§ 21-1901 to -19,177 (1997); NEV. REV. STAT. 82.006- .546 (1999); N.M. STAT. ANN. §§ 53-8-1 to -99 (Michie 2001); N.C. GEN. STAT. §§ 55A-1-01 to -17-05 (2003); N.D. CENT. CODE §§ 10-33-01 to -147 (2003); OR. REV. STAT. §§ 65.001- .990 (2003); 15 PA. CONS. STAT. ANN. §§ 5101-6145 (WEST 1995); S.D. CODIFIED LAWS §§ 47-22-1 to -78 (Michie 2000); TEX. REV. CIV. STAT. ANN. art. X § 1396-1.01 to 1396-11.01 (Vernon 2003); UTAH CODE ANN. §§ 16-6a-101 to 16-6a-304 (2001); VT. STAT. ANN. tit. 11B, §§ 1.01- 17.05 (1997); WASH. REV. CODE §§ 24.03.005- .925 (1994); W. VA. CODE ANN. §§ 31E-1-101 to -15-1520 (Michie 2003); WIS. STAT. §§ 181.0103- 1703 (2002). The following states’ statutes are based upon the 1987 Revised Model Nonprofit Corporation

displacing trust law with corporate law was famously articulated in *Stern v. Lucy Webb Hayes National Training School*:²⁰⁷

The charitable corporation is a relatively new legal entity which does not fit neatly into the established common law categories of corporation and trust. . . . [T]he modern trend is to apply corporate rather than trust principles in determining the liability of the directors of charitable corporations, because their functions are virtually indistinguishable from those of their “pure” corporate counterparts.²⁰⁸

As we have seen, a number of more recent opinions like *Health Midwest* have followed *Stern* and applied the corporate standard rejecting categorical attempts to import charitable trust law to guide evaluations of directors’ decisions:

The Kansas cy pres statute governs changes to the purposes of charitable trusts, devises and bequests. The cy pres statute does not apply to changes to the purposes of nonprofit corporations. The cy pres statute applies only to any restricted gifts and not the entity as a whole (citation omitted). No restricted gifts have been identified herein and therefore the cy pres statute does not apply.²⁰⁹

Consistent with the common law trend, the Revised Model Nonprofit Corporation Act of 1987 adopted virtually the same fiduciary duty standard applicable to business corporations²¹⁰ and specifically rejected the stricter

Act: MISS. CODE ANN. §§ 79-11-101 to -529 (2001); and TENN. CODE ANN. §§ 48-51-101 to -68-105 (2002). See also Peregrine & Schwartz, *supra* note 179, at 192.

207. 381 F. Supp. 1003, 1013 (D.C. Cir. 1974).

208. *Id.* at 1013. As far back as 1967, a New Jersey court observed that the hospital was a charitable corporation, governed by the law applicable to charitable corporations which is rooted partially “in the law of trusts, to some extent in the law of corporations; to some extent it may partake of both or indeed be *sui generis*.” *Paterson v. Paterson Gen. Hosp.*, 235 A.2d 487, 489 (N.J. Super. Ct. Ch. Div. 1967). Thus, although the court did not say that the board had unlimited discretion to deviate from its charter, it clearly and ultimately viewed the case as subject to the law governing nonprofit corporations, as opposed to trusts. *Id.* at 489. Delaware followed suit, repeatedly affirming that charitable corporations are subject to corporate rather than trust law. *Oberly v. Kirby*, 592 A.2d 445, 467 (Del. 1991). The court further noted that philanthropists understand the difference between a trust and nonprofit corporation when they make their gifts, and when they use the corporate form, they “invoke the far more flexible and adaptable principles of corporate law.” *Id.*

209. *Health Midwest v. Kline*, No. 02-CV-08043, 2003 WL 328845, at *19 (D. Kan. Feb. 6, 2003).

210. Section 8.30 of the Revised Model Nonprofit Corporation Act adopts the standard

trust standard.²¹¹ Thus, for purposes of assessing *liability* of corporate directors, the Revised Act and most state nonprofit corporation laws apply the corporate standard discussed earlier.²¹²

It must be acknowledged that the RMNCA is not without ambiguity. While it is clear that the Revised Act was designed to shift the standard applicable to the nonprofit *director* from the trust to business standard,²¹³ the Act suggests the possibility that the *corporation*, as distinct from the director, may continue to be subject to state common law that applies trust rules to the property held by the nonprofit corporation.²¹⁴ Several state attorneys general have exploited this uncertainty to apply a different standard to the *regulation* of the assets of nonprofit corporation, as opposed to imposition of director liability.

That corporate law governs directors' fiduciary duties, but trust law would govern their power to manage charitable assets, makes little sense

of conduct almost identical to that of the Revised Model Business Act:

[D]irector shall discharge his or her duties as a director, including his or her duties as a member of a committee: (1) in good faith; (2) with the care of an ordinarily prudent person in a like position would exercise in similar circumstances; and (3) in a manner the director reasonably believes to be in the best interest of the corporation.

REVISED MODEL NONPROFIT CORP. ACT § 8.30 (1987); *see* Moody, *supra* note 134, at 275 (noting that section 8.30 "clearly settles the dispute as to whether directors of nonprofit corporations should be held to the standard of the director of a business corporation or the standard of a trustee"). *See generally* PEREGRINE & SCHWARTZ, *supra* note 158 (general standard for directors of for-profit and nonprofit corporations same in almost all states).

211. Section 8.30 of the Revised Act sets for the general standards of conduct for nonprofit board directors: A director shall not be deemed to be a trustee with respect to the corporation or with respect to any property held or administered by the corporation, including without limit, property that may be subject to restrictions imposed by the donor or transferor of such property. REVISED MODEL NONPROFIT CORP. ACT § 8.30(e) (1987).

212. *See generally* Peregrine & Schwartz, *supra* note 179, at 185 (explaining that the Revised Model Act tracks directors' duties articulated in the Model Business Corporation Act).

213. *See* 1 MARILYN E. PHELAN, NONPROFIT ENTERPRISES: CORPORATIONS, TRUSTS AND ASSOCIATIONS § 4:02 (2000).

214. REVISED MODEL NONPROFIT CORP. ACT, § 8.30 cmt. 1 (1987). Several states that have substantially adopted the Revised Act have not adopted 8.30(e), thereby leaving open the question of how the relationship between statutory and common law applies to the nonprofit director. Some commentators advance this interpretation as well. *See* Frost, *supra* note 177, at 946; Singer, *supra* note 68, at 237; *cf.* 1 PHELAN, *supra* note 213, § 4:02 ("The charitable corporation is a relatively new legal entity that does not fit neatly into the established common-law categories of corporation or trust.").

doctrinally or analytically. The drafters of the Model Act clearly intended to recognize nonprofits as corporations, and to regulate them as such. While the corporate law model has its deficiencies, it is unquestionably superior, both analytically and practically, to a charitable trust approach to governing nonprofits. States can more easily tailor corporate law to the unique aspects of the nonprofit sector either statutorily or, for example, by differently articulating the business judgment rule. Because only a handful of states had common law one way or the other addressing the relationship of trust law to the assets of nonprofit corporations before the enactment of nonprofit corporate statutes, clarifying the law should not be difficult. Most state courts facing this issue today are doing so for the first time. The corporate standard of governance facilitates the articulation of clear parameters within which nonprofit boards may alter the corporate mission, which power is essential to the functioning of commercial not-for-profits. This result is consistent with the comments to the Revised Model Act, which merely leave open the possibility that trust law would still apply to charitable assets.

2. *Nonprofit Corporate Law and Quasi-Cy Pres*

New York has sought a middle ground between the corporate standard and charitable trust law. Yet, New York law makes clear that even states that have attempted to address the unique characteristics of the nonprofit form have not avoided activism by the attorney general or confusing guidance from the courts on the scope of board autonomy to direct mission.

New York clearly subscribes to corporate law principles in the governance of the charitable corporation. Unlike most states, however, it also addressed the ownership and mission questions unique to the charitable corporation²¹⁵ by rejecting the concept that the assets of a

215. According to *MEETH*:

Not-for-profit corporations operate under legal regimes designed for traditional for-profit corporations. However, fundamental structural differences between not-for-profit corporations and for-profit corporations render this approach incapable of providing effective internal mechanisms to guard against directors' improvident use of charitable assets. For example, in the for-profit context, shareholder power ensures that Boards make provident decisions, while in the not-for-profit context, this internal check does not exist. To put it another way, a nonprofit corporation has no "owners" or private parties with a pecuniary stake to monitor and scrutinize actions by the directors.

Manhattan Eye, Ear & Throat Hosp. v. Spitzer, 715 N.Y.S.2d 575, 592 (Sup. Ct. 1999). Both the attorney general and a court must be involved in the disposition of substantially all of the nonprofit's assets, "to ensure that the interests of the ultimate beneficiaries of the

nonprofit that accrue from a gift are subject to a trust;²¹⁶ requiring notice to the attorney general, and court approval, whenever an amendment to the articles of incorporation affects the corporate purposes or powers;²¹⁷ recognizing the duty of obedience; and treating the disposition of assets upon dissolution under a process that the courts refer to as *quasi-cy pres*.²¹⁸ As conceptualized by the *MEETH* court, "A charitable Board is essentially a caretaker of the not-for-profit corporation and its assets. As caretaker, the Board 'ha[s] the fiduciary obligation to act on behalf of the corporation . . . and advance its interests.'" ²¹⁹

Despite its attempts to affirmatively deal with the unique characteristics of the nonprofit form, New York law fails to articulate a clear definition of mission or the extent to which the board may alter the nature of the nonprofit's business while still remaining faithful to that mission. Consequently, courts' conflicting signals about the scope of the attorney general's power over charities has created uncertainty for nonprofit boards. For example, the *MEETH* board asserted that its strategic plan was not a new or different mission, and consequently sought to implement its planned transition to out-patient services without amending its articles of incorporation; this approach dispensed with any requirement of obtaining court approval of a change in purpose. The court disagreed with *MEETH*'s view on the scope of its mission, of course,

corporation, the public, are adequately represented and protected from improvident transactions." *Id.* Further, the *MEETH* court observed that the legislature imposed a higher standard of care upon the director of the nonprofit. *Id.* at 593.

216. N.Y. NOT-FOR-PROFIT CORP. LAW § 513(a) (McKinney 1997). Subsection b adds: "Except as may be otherwise permitted under article eight of the estates, powers and trusts law or section 522 (Release of restrictions on use or investment), the governing board shall apply all assets thus received to the purposes specified in the gift instrument." *Id.* § 513(b). The legislative history elucidates: "[t]he board is under a duty to apply such funds in accordance with the directions of the donor, but within the framework of the corporation law rather than the trust law." *Alco Gravure v. Knapp Found.*, 490 N.Y.S.2d 116, 121 n.7 (1985) (quoting Memorandum of the Joint Legislative Committee to Study Revision of Corporation Laws (Jan. 13, 1969)).

217. N.Y. NOT-FOR-PROFIT CORP. LAW § 804 (a) (ii) (McKinney 1997).

218. *See Gravure v. Knapp Found.*, 64 N.Y.2d 458, 474-76 (1985) (summarizing the legislative history of the 1969 statute). The *quasi-cy pres* concept is embodied in N.Y. NOT-FOR-PROFIT CORP. LAW § 1005(a) (3) (A) (2002) (addressing the distribution of the assets of a nonprofit undergoing dissolution). *See also In re Thurston*, 746 N.Y.S.2d 343, 346 (Sur. Ct. 2002) (explaining how *quasi-cy pres* concept works in the nonprofit corporate context).

219. *Manhattan Eye, Ear & Throat Hosp.*, 715 N.Y.S.2d at 593 (citing *Pebble Cove Homeowners' Assoc., Inc. v. Shoreatlantic Dev. Co.*, 595 N.Y.S.2d 92 (1993)) (alteration in original).

but more troubling is that the court's analysis fails to guide other charitable corporations making significant changes that arguably fall within the original mission.

The *MEETH* court also invoked the duty of obedience,²²⁰ declaring that “the duty of obedience, perforce, must inform the question of whether a proposed transaction to sell all or substantially all of a charity’s assets promotes the purposes of the charitable corporation when analyzed under section 511.”²²¹ The court treated the proposed *MEETH* transaction as analogous to a conversion, “inasmuch as in both there is a charitable organization which alleges that it is incapable of continuing its primary mission of operating a hospital, seeks approval of the sale of all its assets, and plans to apply the sale proceeds towards a *newly revised mission*.”²²² In applying the duty of obedience, the court characterized its role as ensuring that nonprofit boards are “faithful to the purposes and goals of the organization”—nonprofits are not ultimately about making money, but about the “perpetuation of [the] particular activities [that] are central to the *raison d’être* of the organization.”²²³ The court also held that the duty of obedience mandates that the board depart from its core mission only as a “last resort.”²²⁴ While this court’s interpretation of the duty of obedience seems more liberal than that embodied by charitable trust law, it certainly was not so in application to the facts of the *MEETH* case, and resulted in a much different outcome than *Littauer*,²²⁵ which did not invoke the duty of obedience.

Quasi-cy pres is also intended to ensure fidelity to mission, by requiring boards to dispose of charitable assets upon dissolution to entities that will subscribe to the dissolving corporation’s original purpose. As interpreted by New York’s highest court, *quasi-cy pres* is less restrictive than the charitable trust *cy pres* concept. It:

accords greater authority to the corporation’s board of directors and the courts than governs the distribution of the assets held by a trustee under a will or other instrument making a disposition for charitable purposes . . . or than was the *cy pres* standard at

220. According to the *MEETH* court, the duty of obedience had only been previously raised in breach of duty situations, and never in the context of the sale of assets. 715 N.Y.S.2d at 593.

221. *Id.*

222. *Id.* at 594 (emphasis added).

223. *Id.* at 593.

224. *Id.* at 595.

225. *Nathan Littauer Hosp. Ass’n v. Spitzer*, 287 A.D.2d 202 (N.Y. App. Div. 2001).

common law (“as nearly as possible”).²²⁶

Interestingly, however, MEETH was not dissolving. Rather, the board sought to monetize the hospital facility to enable it to establish clinics. Thus, the court appears to have taken some liberties in its invocation of the cy pres doctrine. This is, of course, consistent with the trend of other states employing charitable trust principles to strengthen their ability to second-guess nonprofit boards.

By contrast, and further confusing the matter of what constitutes a mission change, the *Littauer* court held that a change in corporate membership, which the attorney general characterized as a disposition of assets,²²⁷ was not a change in the underlying purpose, nor the overall business purpose, of the hospitals.²²⁸ The court observed: “Plainly, the statute is designed to require prior court approval only in instances where the proposed amendment truly seeks to change the nature, object or powers of a particular corporation.”²²⁹ The court also rejected *amici* arguments that a requirement of compliance with the Catholic Ethical and Religious Directives in addition to the articles of incorporation constituted a curtailment of corporate powers requiring judicial approval. The court distinguished between corporate powers and purposes, and the services the

226. *In re Multiple Sclerosis Serv. Org.*, 68 N.Y.2d 32, 35 (1986). The Court of Appeals further stated:

Under the quasi cy pres standard of the Not-For-Profit Corporation Law, a Supreme Court Justice in determining whether to approve the plan of distribution proposed by the corporation’s board, and if not to what other charitable organizations distribution should be made, should consider (1) the source of the funds to be distributed, whether received through public subscription or under the trust provision of a will or other instrument; (2) the purposes and powers of the corporation as enumerated in its certificate of incorporation; (3) the activities in fact carried out and services actually provided by the corporation; (4) the relationship of the activities and purposes of the proposed distributee(s) to those of the dissolving corporation, and (5) the bases for the distribution recommended by the board.

Id.

227. The Attorney General argued that the two hospitals’ filing of restated articles of incorporation, which reserved certain governance and management powers for the new corporate parent, required notice to the attorney general and court approval under N.Y. NOT-FOR-PROFIT CORP. LAW § 510 (McKinney 1997), which governs disposition of assets, and N.Y. NOT-FOR-PROFIT CORP. LAW § 804 (McKinney 1997), which addresses changes to corporate purposes and powers. *Nathan Littauer Hosp. Ass’n v. Spitzer*, 287 A.D.2d 202, 204 (N.Y. App. Div. 2001).

228. 287 A.D.2d at 204; *see also supra* notes 77-81 and accompanying text.

229. *Id.* at 205.

entity actually provides, stating: “the decision to delineate in a restated certificate of incorporation a specific or potential restriction upon the services to be provided by the corporation is not the functional equivalent of altering the corporation’s underlying purpose or curtailing its power to achieve its overall objectives.”²³⁰ In sum, then, the New York statute’s attempt to regulate boards’ oversight of the nonprofit mission has, in the courts’ hands, generated confusion without promoting attention to the role of mission. Since charitable corporations pursuing a dynamic strategic plan are likely to avoid court intervention, of greater relevance to the daily operation of the charitable corporation is the wide gulf between the attorney general’s and nonprofit sector’s concept of the scope of an entity’s mission, and what actions comprise a change to mission requiring judicial approval.

3. *Statutory Charitable Trust States*

As noted above, Illinois and New Hampshire have long been clear in their treatment of the nonprofit corporation—nonprofit assets are subject to charitable trust by virtue of statute.²³¹ Virginia has just recently joined this statutory charitable trust group.²³² This Subsection will focus its discussion on New Hampshire, where the attorney general has asserted his statutory charitable trust power over the health care industry quite aggressively.

New Hampshire law specifically delineates “health care charitable trusts,” to include health care providers and payors.²³³ As a result, the New

230. *Id.* at 207.

231. RESTATEMENT (THIRD) OF TRUSTS § 10 cmt. b, at 198 (Tentative Draft No. 1, 1996). The examples and illustrations included in the draft, however, are dissimilar to the scenarios presented here. An Illinois appellate court, applying the state’s Charitable Trust Act in *Riverton Area Fire Prot. District v. Riverton Volunteer Fire Dep’t*, 566 N.E.2d 1015 (Ill. App. Ct. 1991), held that a not-for-profit corporation held its assets as trustee of a charitable trust; no trust documents were required to evidence the creation of the trust, rather, the court observed, “charitable trusts are remedial and created by statute.” *Id.* at 1019.

232. 2002 Va. Acts. ch. 792, § 2.2-507.1 (codified as amended at VA. CODE ANN. § 2.2-507.1 (Michie 2004)). The next section of the act gives the courts the same subject matter jurisdiction over the assets of the charitable corporation as they have over the assets of a charitable trust. *Id.* § 17.1-513.01 (codified as amended at VA. CODE ANN. § 17.1-513.01 (Michie 2004)).

233. N.H. REV. STAT. ANN. § 7:19-b(I)(d) (2004). Charitable organization is defined as any entity established for the public health, other charitable purpose, or solicits for any charitable purpose. N.H. REV. STAT. ANN. § II(b)(2) (Lexis Supp. 2001).

Hampshire Attorney General's opinion letter in *Optima Health*, in which he demanded the "unmerger" of two hospitals was, in the abstract, legally sound.²³⁴ From a public policy perspective, however, we argue against states statutorily imposing a trust on nonprofit assets. As discussed throughout this Article, characterizing nonprofit holdings as trust assets devalues those assets, making it significantly more difficult for nonprofits to partner and obtain access to affordable capital. The New Hampshire Attorney General's devolution of the Optima deal would certainly make any potential affiliate think twice before partnering with an entity incorporated in a charitable trust state. Further, it is wholly unclear whether and how the Attorney General's opinion accounted for the health policy questions implicated by the hospitals' merger.

The New Hampshire Attorney General's response to the Optima merger was dramatic and has become a significant example of the potential of an attorney general to require *cy pres* proceedings to ensure, as conceived by the attorney general, that the charitable corporation abides by the articulated purposes of the corporation.²³⁵ The *Optima* opinion has also become "seminal" for other states because it not only relies upon the New Hampshire Charitable Trust Act, but also comprehensively brings together charitable trust common law from across the country.²³⁶ For precisely this reason, the opinion has contributed significantly to the current doctrinal confusion regarding the application of trust law to nonprofit assets. *Optima* relies upon California common law, Illinois statutory charitable trust cases, and express charitable trust cases, without explaining the doctrinal distinctions between the law of states that are "statutory charitable trust states" and those that are not, or the inapplicability of express trust cases to most nonprofit health care

234. N.H. DEP'T OF JUSTICE, *supra* note 10. The multi-hospital merger, in its inception, was the product of a 1994 deal between Elliot Hospital and the Catholic Medical Center (CMC). *Id.*

235. Specifically, the Attorney General observed that "[a]lthough a charitable corporation may not be governed as a trust in every respect, courts have held that the assets of a charitable corporation are impressed with a charitable trust that restricts the use of the assets to the defined purposes of the corporation." N.H. DEP'T OF JUSTICE, *supra* note 10.

236. Footnote 10 of the Attorney General's opinion, *id.*, is a string cite of a collection of charitable trust cases virtually identical to footnote 7 in ROBERT A. BOISTURE & DOUGLAS N. VARLEY, STATE ATTORNEYS GENERAL'S LEGAL AUTHORITY TO POLICE THE SALES OF NONPROFIT HOSPITALS AND HMOs (1995), <http://www.volunteertrustees.org/legal.html> (describing attorneys general's authority to use charitable trust doctrine against hospitals). These cases are much more nuanced than either report concedes.

providers.²³⁷

Finally, *Optima* is a very hard case from a public policy perspective. There is no question that the Attorney General was responding to the community's unhappiness with the merger, which resulted from application of the Catholic Ethical and Religious Directives to the new entity,²³⁸ the elimination of acute care services at one of the campuses, and finally, *Optima's* 1997 decision to affiliate with out-of-state Covenant Health Systems, itself a Catholic multi-state hospital system. The community also felt misled by the hospital leaders about what the implications of the transaction would be. This is not at all atypical, and captures why, as a doctrinal matter, questions of nonprofit ownership, mission accountability, and satisfaction of the beneficiaries remain unresolved. Every community wants to retain its hospital, ideally, with the most up-to-date technology and a full panoply of services. These aspirations are frequently inconsistent with what the providers in the community can financially sustain, and what, from a public policy perspective, represents a responsible allocation of resources. So, the question becomes who dictates how the assets of the nonprofit provider are best used—the board, the community, or some arm of the state (the attorney general, the department of health, or a court). The *Optima* opinion does not engage the full scope of these issues, and is therefore poor precedent for their resolution.

Ultimately, *Optima Health* was dissolved at a cost of ten million dollars, and the two hospitals returned to their original independence.²³⁹ Whether the outcome was worth the price is probably impossible to ever determine.

237. N.H. DEP'T OF JUSTICE, *supra* note 10; cf. BOISTURE & VARLEY, *supra* note 236.

238. That few people seemed to understand the implications of the Ethical and Religious Directives bolstered the Attorney General's argument that the new entity's mission was unclear and inconsistent with both of the predecessor organizations. Notably, the Attorney General expressed significant concern that, in his view, *Optima* was disregarding CMC's traditionally commitment to religious health care and was potentially violating the Ethical and Religious Directives in its delivery of health services at the newly established acute care facility. N.H. DEP'T OF JUSTICE, *supra* note 10.

239. See Julia L. Eberhart, *Merger Failure: A Five-Year Journey Examined*, HEALTHCARE FIN. MGMT., Apr. 2001, at 37, 39.

B. Attorneys General's Attempts To Integrate Charitable Trust Doctrine with Nonprofit Corporate Law

1. Attorneys General's Activism

As cases discussed in this Article reveal, attorneys general who find current law inadequate to accomplishing their goals of increased oversight of nonprofit boards are attempting to strengthen their power with a reinvigorated charitable trust law blended with corporate analyses whenever possible. In most cases, the attorneys general assert that the legislatures' enactment of statutes clearly applying corporate principles to nonprofits did not wholly displace pre-existing common law applying trust principles to not-for-profits. In some instances, attorneys general have had to import the charitable trust law upon which they are relying from other states. We explore in this Section the attorneys general's use, or misuse, of common law to accomplish these ends. California common law is an extremely important source for the proposition that the assets of a charitable corporation comprise a charitable trust, subject to the oversight of the state attorney general, and limited to the purposes articulated in the articles of incorporation.²⁴⁰ In 1964, the California Supreme Court, in *Holt v. Coll. of Osteopathic Physicians & Surgeons* directed that "charitable contributions must be used only for the purposes for which they were received in trust."²⁴¹ A decade later, a California appellate court precluded Queen of Angels Hospital from closing its hospital and converting its assets to the operation of health clinics.²⁴² The primary purpose for which Queen of Angels was organized, argued the attorney general, was the operation of a hospital, and that is what it must do, as long as it remains in existence.

It is highly uncertain whether these cases remain good law in California. In 1980, California adopted a nonprofit corporate act, which became the model for the ABA's Revised Model Nonprofit Corporation Act.²⁴³ In so doing, the California legislature applied corporate fiduciary duties to nonprofit directors, and, like New York, incorporated some charitable trust concepts with respect to nonprofit corporate assets. Whether the California statute occupies the entire field of nonprofit governance, or whether some vestiges of the pre-1980 common law remain

240. See, e.g., *Queen of Angels Hosp. v. Younger*, 136 Cal. Rptr. 36, 39 (Ct. App. 1977).

241. 394 P.2d 932, 935 (Cal. 1964).

242. *Queen of Angels Hosp.*, 136 Cal. Rptr. at 39.

243. REVISED MODEL NONPROFIT CORP. ACT (1987).

viable, is an unanswered question in California.²⁴⁴ This question is not unique to California, of course. The relevance of trust-based common law after a state's enactment of a nonprofit corporate statute must be considered in every state.

Taking advantage of the doctrinal ambiguity, attorneys general have argued four different cases for subjecting the assets of a nonprofit hospital corporation to a charitable trust: that assets acquired from general donations are subject to a trust; that assets intermingled with acquisitions resulting from general donations cannot be separated, thereby necessitating that all of the charities' assets be subject to a trust; that a consequence of tax exemption is the imposition of a trust on the nonprofit's assets; and that restricted gifts are subject to a trust. Only the last of which, restricted gifts, finds support in the doctrine of traditional trust law. Restricted gifts comprise what is generally understood to be charitable trust property, irrespective of whether the donor uses the designation "charitable trust":²⁴⁵ The donor gives money or property for a very specific articulated use by the corporate recipient. The property is subject to a trust,²⁴⁶ with the corporation as trustee.²⁴⁷ Thus, it is uncontroversial that if a nonprofit hospital corporation is sold, converts, or dissolves, it must treat separately any trust property it received during its existence, ensuring that in its capacity as trustee, it is faithful to the settlor's intent.

The disputes between states and hospitals arise from attorneys general's use of charitable trust law more expansively, by asserting that *all* of the assets of the nonprofit corporation are subject to a trust. This

244. Our thanks to James Schwartz for helping us sort through the morass that California law appears to be to a New Yorker.

245. Property held by a charitable corporation is subject to a charitable trust most typically when the donor attaches conditions to a gift. "A disposition to [a hospital or university] for a specific purpose, such as to support medical research, perhaps on a particular disease, or to establish a scholarship fund in a certain field of study, creates a charitable trust of which the institution is the trustee for purposes of the terminology and rules of this Restatement." RESTATEMENT (THIRD) OF TRUSTS § 28 cmt. a (2003).

246. In New York, pursuant to the not-for-profit corporate statute, the corporation would not become a trustee or be subject to charitable trust law, but corporate law. Nonetheless, if the corporation receives a gift with conditions, or that uses trust language, it is bound by the intentions of the donor, unless it undergoes a *quasi-cy pres* proceeding. See *Alco Gravure, Inc. v. Knapp Found.*, 479 N.E.2d 752, 757 n.7 (N.Y. 1985).

247. The members of the board of directors are not trustees, in the strict sense, however, because they do not hold title to the property of the corporation. RESTATEMENT (THIRD) OF TRUSTS § 5 cmt. g (2003).

assertion takes three different forms, none of which charitable trust law supports. The first form of the argument posits that whatever is acquired by general donations to the hospital becomes property subject to a trust because donors expected and understood that their gifts would be used for and by the recipient hospital.²⁴⁸ This argument is wrong; outright donations to a charity, particularly those solicited in connection with a campaign or fund-raising event, simply do not satisfy the prerequisites for the establishment of a trust.²⁴⁹ A slight variant of this first argument is that the assets owned by a charitable corporation with restrictions on use articulated in its articles of incorporation are subject to a constructive charitable trust, protecting them from a non-compliant use.²⁵⁰ Though not

248. See, e.g., *Banner Health Sys. v. Long*, 663 N.W.2d 242, 247 (S.D. 2003). In *Banner Health*, although the court rejected the Attorney General's argument that an implied charitable trust resulted from donations made to support the corporation's general purposes, it did recognize the possibility of a constructive trust if "Banner was unjustly enriched by the sale of the assets and removal of the proceeds from the local communities at the expense of those communities . . ." *Id.* at 248. The court then suggested that if indeed the facts support the finding of a constructive charitable trust, the directors could be in breach of their fiduciary duties for having used the trust property in a manner adverse to the interests of the beneficiaries. *Id.* at 249; see also *supra* notes 99-107 and accompanying text.

249. See, e.g., *Nat'l Found. v. First Nat'l Bank*, 288 F.2d 831, 834, 836 (4th Cir. 1961) (finding that donations made to a local chapter of National Foundation in response to a general appeal did not constitute a charitable trust to the local chapter, but rather were an unrestricted gift to National Foundation); *Persan v. Life Concepts, Inc.*, 738 So. 2d 1008 (Fla. Dist. Ct. App. 1999) (making a gift to a charity for a specific purpose does not create a charitable trust; creation of trust must be express, with intent established beyond a reasonable doubt); *United Methodist Church v. Bethany Med. Ctr.*, 969 P.2d 859 (Kan. 1998) (not a case where originating donor created a trust but rather a situation where five Methodists incorporated for the purpose of collecting donations for a hospital but no single donor, including the church, acted as a trust settlor); see also 76 AM. JUR. 2D § 141 (2004). This outcome is consistent with the Restatement of Trusts: "An outright devise or donation to a nonproprietary hospital or university or other charitable institution, expressly or impliedly to be used for its general purposes, is charitable but does not create a trust as that term is used in this Restatement." RESTATEMENT (THIRD) OF TRUSTS § 28 (2003).

The distinction between a trust and an unrestricted gift is controlled by the intention of the donor to impose enforceable duties upon the recipient. See 15 AM. JUR. 2D § 120 (2004); see also *Eychaner v. Gross*, 747 N.E.2d 969 (Ill. App. Ct. 2001) (resolving dispute as to whether university evidenced intent to place in trust with theater council either theater building or intangible interests in maintaining the theatre), *rev'd*, 779 N.E.2d 1115 (Ill. 2002).

250. See, e.g., *Banner*, 663 N.W.2d at 250. This result occurs from a convoluted combination of trust and statutory analysis, and depends upon a finding that non-members'

doctrinally grounded, the rationale advanced for this position is not unappealing: that “[a]ny other rule of law would allow a charitable nonprofit corporation to eviscerate the charitable purpose for which it was formed without recourse for those who donated funds for that purpose.”²⁵¹ A response to this argument is that a donor committed to the perpetual mission of her designated charity might have protected her intent by creating a trust; that she did not subjects her to the risk of a charitable board taking the entity in a new direction.

The second argument in favor of imposing a charitable trust on the entire assets of a nonprofit corporation assumes that, because it is impossible to separate out assets acquired from general fund-raising (which are wrongly designated trust assets by this analysis) from non-donated assets, all assets must be treated as subject to a trust.²⁵² The adage that “two wrongs don’t make a right” comes to mind. Third, attorneys general posit that nonprofit hospital assets that are under-written by the government through tax exemptions and payments by government health plans should be subject to a trust.²⁵³ This sweeping approach is free-floating social policy masquerading as trust law.

The attorneys general in North Dakota, South Dakota, and New Mexico all attempted to use charitable trust principles to block Banner’s removal of the proceeds from Banner’s liquidation of its assets in their respective states. Recall that Banner is a multi-state health care system that sold its holdings in North Dakota, South Dakota and New Mexico so that it

rights are affected by the elimination of the restrictions. *Id.* at 248-49. Such would unlikely succeed in a state whose nonprofit corporate statute does not resemble North Dakota’s.

251. *Id.* at 250.

252. This “implied trust” argument is also explained as a “base capital” concept—that the originally donated assets facilitated the generation of other assets or value, such that the entire body must be subject to trust. *See Coffey et al., supra* note 102, at 4. A Massachusetts case represents a situation where the hospital was originally founded as a result of a trust and whose assets were later indistinguishable from subsequent gifts. *Att’y Gen. v. Hahnemann Hosp.*, 494 N.E.2d 1011, 1021 (Mass. 1986) (finding that where assets of a charitable trust dedicated to the operation of homeopathic hospitals are so intertwined with other hospital funds, the board would violate fiduciary duties if it dedicated funds from the trust, or funds donated prior to the change in corporate purpose by donees who understood the purpose to be governed by the trust, to a new purpose).

253. *See Horwitz, supra* note 5, at 1347; *see also Coffey et al., supra* note 102, at 5 (observing that the South Dakota Supreme Court did not address the contention that the taxes not paid by the hospitals enable them to enhance their value). The North Dakota trial court rejected the argument that by accepting tax benefits, a nonprofit corporation converted to a charitable trust. *Id.*

could concentrate its operations in and around Colorado and Arizona. The attorneys general sought to limit Banner's ability to liquidate its holdings and move the proceeds by establishing the existence of a constructive or implied trust. They relied on two now familiar arguments: first, that the donations, and possibly the entirety of the hospital's assets, particularly from local citizens, were intended for the community hospital, and not the larger corporation, and therefore created a trust for the benefit of the community; second, that the tax benefits obtained through the hospitals' exemption created a trust of which taxpayers are the beneficiaries—otherwise, unjust enrichment would result from the hospitals' retention of the value of the benefits accruing from tax forgiveness.²⁵⁴

In the only case that actually produced a court opinion, the attorney general of South Dakota convinced the South Dakota Supreme Court to integrate charitable trust law with the state's nonprofit corporation act, producing a legal precedent which is doctrinally flawed and impossible to apply. While the South Dakota Supreme Court agreed that nonprofits are subject to the state's nonprofit corporate statute, it also held that the corporate statute did not supersede the law of charitable trusts.²⁵⁵ Thus, the court concluded that it may be necessary to impose a constructive charitable trust on the hospital assets to protect donors.²⁵⁶ Finally, the court suggested that if the attorney general could establish that Banner was in a fiduciary relationship with the various communities it served, pursuant either to trust law or the general common law governing fiduciary relationships, Banner's decision to sell its facilities may have breached its duties as a fiduciary.²⁵⁷

The South Dakota Supreme Court correctly rejected any possibility that charitable corporate assets are subject to an express trust²⁵⁸—the

254. *See id.*

255. *Banner Health*, 663 N.W.2d at 247. The court specifically sought to preserve the relevance of the following statutory language preserving a court's ability to employ the implied trust device when equity so requires:

The enumeration in §§ 55-1-7 to 55-1-10, inclusive, of cases wherein an implied trust arises does not exclude or prevent the arising of an implied trust in other cases nor prevent a court of equity from establishing and declaring an implied, resulting, or constructive trust in other cases and instances pursuant to the custom and practice of such courts.

Id. at 246-47 (quoting S.D. CODIFIED LAWS § 55-1-11 (Michie 2004)).

256. *Id.* at 249.

257. *Id.*

258. RESTATEMENT (THIRD) OF TRUSTS § 27 (2003) provides that a trust may be created for

specific elements of an express trust are absent.²⁵⁹ Without further explication, the Supreme Court recognized the possibility of an implied trust “based on theories of unjust enrichment, breach of fiduciary duties, and improper amendment of the charitable corporation’s articles of incorporation.”²⁶⁰ This outcome is unsupported by precedent. The imposition of a charitable trust as a result of tax-exemption²⁶¹ and fund-raising finds no support in charitable trust doctrine. The literature states that charitable trusts result only from express and not implied trusts.²⁶² Further, the law has been clear that unrestricted charitable donations do not create a trust;²⁶³ donations to hospitals, particularly those solicited in connection with a campaign or fund-raising event, do not satisfy the

private or charitable purposes, or a combination thereof. RESTATEMENT (SECOND) OF TRUSTS § 372 (1959) provides that “[a] trust for the promotion of health is charitable.” See also RESTATEMENT (THIRD) OF TRUSTS § 28(d) (2003). For an historic overview of the development of the legal recognition and treatment of charitable trusts in the United States, see Nina J. Crimm, *A Case Study of a Private Foundation’s Governance and Self-Interested Fiduciaries Calls for Further Regulation*, 50 EMORY L.J. 1093, 1098-1133 (2001).

259. Property held by a charitable corporation is subject to a charitable trust most typically when the donor attaches conditions to a gift, whether or not she explicitly designates that it be held as a charitable trust. RESTATEMENT (THIRD) OF TRUSTS § 13 (2003) (“The manifestation of intention requires an external expression of intention as distinguished from undisclosed intention. There may, however, be a sufficient manifestation of intention to create a trust without communication of that intention to the beneficiary or to the trustee or any third person.”). By virtue of the restrictions, the donee corporation becomes the trustee obliged to ensure that the property is devoted to the specified purposes. The members of the board of directors are not trustees, in the strict sense, however, because they do not hold title to the property of the corporation. *Id.* § 5 cmt. g. “A disposition to [a hospital or university] for a specific purpose, however, such as to support medical research, perhaps on a particular disease, or to establish a scholarship fund in a certain field of study, creates a charitable trust of which the institution is the trustee for purposes of the terminology and rules of this Restatement.” *Id.* § 28.

260. *Banner Health*, 663 N.W.2d at 248. South Dakota’s nonprofit corporate statute is unremarkable; it allows amendments to the articles “in any and as many respects as may be desired,” S.D. CODIFIED LAWS § 47-22-14 (Michie 2004), so long as “[n]o amendment to the articles of incorporation shall affect any existing cause of action in favor of or against such corporation, or any pending action to which such corporation shall be a party, or the existing rights of persons other than members,” *id.* § 47-22-22 (emphasis added in *Banner Health*, 663 N.W.2d at 249).

261. See, e.g., *Hughes v. Good Samaritan Hosp.*, 158 S.W.2d 159 (Ky. 1942); *Levin v. Sinai Hosp.*, 46 A.2d 298 (Md. 1946).

262. See, e.g., Coffey et al., *supra* note 102, at 4.

263. See *supra* note 249 and accompanying text.

prerequisites for a trust.²⁶⁴ This analysis is consistent with the Restatement of Trusts: “[a]n outright devise or donation to a nonproprietary hospital or university or other charitable institution, expressly or impliedly to be used for its general purposes, is charitable but does not create a trust as that term is used in this Restatement.”²⁶⁵

The South Dakota Supreme Court accepted an extremely complicated analysis that provides literally no guidance to the nonprofit sector as to the circumstances that may give rise to a charitable trust. For multi-state nonprofit systems, even the specter that an attorney general might seize its assets can cripple the organization by devaluing those assets and suggesting protracted litigation to potential suitors. The reality of attorneys general’s attempts to capture charitable assets at the very least extends the time it takes to close any deal, and dramatically increases transaction costs, including attorneys’ fees, which, of course, are paid from the charitable proceeds the attorney general is seeking to preserve. These problems increase exponentially when several attorneys general enter the fray, as happened with *Banner*.

Finally, attorney general involvement with multi-state charitable corporations raises the question of who is looking out for the *national* public interest in the allocation of charitable resources.²⁶⁶ Large health care systems have the financial ability to sustain the rural or urban hospital that struggles to break even each year, has limited access to affordable financing, and lacks the resources to invest in the capital necessary to stay current with the technology required to survive in the current health care market.

264. See, e.g., *Nat’l Found. v. First Nat’l Bank*, 288 F.2d 831, 834, 836 (4th Cir. 1961) (finding donations made to the local chapter of the National Foundation in response to a general appeal is not a charitable trust to local chapter but an unrestricted gift to National Foundation); *United Methodist Church v. Bethany Med. Ctr. Inc.*, 969 P.2d 859 (Kan. 1998) (not a case where originating donor created a trust but rather a situation where five Methodists incorporated for the purpose of collecting donations for a hospital but no single donor, including the Kansas East Conference, acted as a trust settlor). See generally 15 AM. JUR. 2D § 141 (2004).

265. RESTATEMENT (THIRD) OF TRUSTS § 28 (2003).

266. See Brody, *Whose Public?*, *supra* note 13, at 968. Evelyn Brody gives substantial thoughtful attention to *Banner* in her article. She notes that “In terms of the national public interest, however, relocation could be a positive-sum game: The governing board of a charity might determine that the overall social benefit can be increased by moving its activities from a state with a low utility to a state with a higher one.” *Id.*

2. Charitable Trust Law Is Conceptually Ill-Suited to the Nonprofit Corporation

Subjecting a commercial enterprise and its board to charitable trust principles is strained in application and constrained in outcome. Traditional trust standards prioritize preservation of trust assets and strict adherence to the settlor's intent. The duty of loyalty requires strict obedience to the specifications of the trust instrument,²⁶⁷ and administration of the trust solely in the interests of the beneficiaries.²⁶⁸ While several doctrines somewhat blunt the edge of charitable trust requirements,²⁶⁹ the trust standards are nonetheless exacting and unforgiving in their insistence that trustees devote their energies selflessly and diligently toward accomplishing the settlor's objectives.

Further, trust law as the organizational mechanism for nonprofit corporations has little to commend it.²⁷⁰ First, it is analytically ill-suited to

267. RESTATEMENT (THIRD) OF TRUSTS §§ 227 cmt. b, 228(b) (1991).

268. Although attorneys general so far have not sought to apply the trustee fiduciary standards to the directors of charitable corporations, two of the most prominent hospital counsel in this area, Michael Peregrine and James Schwartz, suggest that hospitals should protect against attorneys general imposing constructive trusts upon charitable assets for fear that the imposition of the trust fiduciary standards cannot be far behind. Peregrine & Schwartz, *supra* note 179, at 193. If their prediction proves accurate, corporate rules that subject directors to what essentially amounts to a gross negligence standard would be replaced with a charitable trust regime of simple negligence. *Id.* at 192. Further, a trustee may not engage in transactions with the trust for their personal benefit. Evelyn Brody, *The Limits of Charity Fiduciary Law*, 57 MD. L. REV. 1400, 1419-20 (1998); *see also* Boyd, *supra* note 147, at 734-35; RESTATEMENT (THIRD) OF TRUSTS § 227 cmt. c (1991). Trustees must fully disclose any conflicted transaction, which nonetheless must be fair and reasonable, and in the interests of the beneficiaries. *See* I PHELAN, *supra* note 213, § 4:03. Corporate rules are not nearly so strict.

269. Courts employ the doctrines of *cy pres* to relieve the distress to a trust whose purpose no longer exists, or for which the means dictated by the settlor to accomplish the purposes have become impossible. In so doing, the courts typically evaluate the degree of departure from the original intent before approving a substitute purpose. *See* Greil Mem'l Hosp. v. First Ala. Bank, 387 So. 2d 778, 781 (Ala. 1980) (finding a testamentary gift to charitable corporation made for sole purpose of "curing and preventing tuberculosis" was a charitable trust which assets could only be used for that purpose, despite change in treatment of TB; abandonment of purpose caused legacy to lapse); *see also* Taylor v. Baldwin 247 S.W.2d 741, 750 (Mo. 1952) (holding that courts will intercede where there is a substantial departure from the charity's dominant purpose). Courts sometimes distinguish the trust's purpose, to which the trustees must adhere, from the means about which the trustees may use their discretion, as long as it is not otherwise addressed in the charter. *Id.* at 756.

270. *See* Fishman, *supra* note 127, at 226-87 (explaining the distinctions between

the organizational form of the charitable corporation. A charitable trust is created by the grant of a settlor (the principal) to accomplish a specific and defined benefit for the public; the trustees (agents) are charged with fulfilling the settlor's wishes. The typical charitable corporation, however, has no settlor. To remedy this analytical deficiency, attorneys general are treating taxpayers and donors as the settlors; as a result, the donors/taxpayers become both the settlors and the beneficiaries of the trust.²⁷¹ Interestingly, no attorney general has suggested treating the corporation itself as the settlor; this alternative is obviously unappealing to a regulator, because it would leave the corporate board accountable to itself.²⁷²

Focusing on the identity of the settlor and the beneficiary understates the analytic difficulties, however. The notion that trustees must adhere to the settlor's original intent is justified by the "theory that the right to testation is a fundamental aspect of private property."²⁷³ This rationale simply does not apply to the means by which nonprofit corporations have accumulated their assets. Obviously, where a donation to a hospital carries a testator's express restrictions as to its use, a trust is created and the testator's desires are respected. Typically, however, the assets of a health care corporation have been acquired or built from myriad sources, including the entity's profits, bond issues, tax subsidies, governmental aid

charitable trusts and nonprofit corporations and detailing the benefits of the nonprofit corporate form, for example with respect to governance and ability to respond to changed circumstances).

271. Further, it is unclear precisely who the intended beneficiaries were in the cases of some hospitals' founding. Many Jewish hospitals, for example, were founded as much to ensure residency placements for young Jewish doctors who were precluded from such opportunities in most of America's prestigious academic medical centers. PAUL STARR, *THE SOCIAL TRANSFORMATION OF AMERICAN MEDICINE* 174 (1982). So, in many instances, the physicians who comprise the medical staff were as much the intended beneficiaries as the community that would constitute the hospital's patient base.

272. Brody, *Whose Public?*, *supra* note 13, at 957.

273. Ilana H. Eisenstein, Comment, *Keeping Charity in Charitable Trust Law: The Barnes Foundation and the Case for Conservation of Public Interest in Administration of Charitable Trusts*, 151 U. PA. L. REV. 1748, 1758 (2003). Charitable trusts are exempt from the Rule Against Perpetuities. *Id.* at 1763.

Allowing the trust terms to run in perpetuity produces several public costs. First, economic costs of dead-hand control include limitations on alienability of property, limited marketability, and a decrease in productivity of trust assets and property. . . . Second, time and changing conditions create a risk of obsolescence and thereby detract from the charitable efficiency of the organization.

Id. at 1763-64 (footnotes omitted).

and unrestricted donations. Even with private property, public policy strives to limit dead-hand control. It is simply bad policy and law to attach gratuitous restrictions on the significant holdings of a commercial nonprofit enterprise, particularly one that operates in a dynamic industry such as health care. This is not to say, of course, that there are not public policy detriments of allowing too permissive use of charitable assets. Donors and taxpayers may be discouraged from supporting entities that do not promise some reasonable commitment to the community good for which the contribution was originally intended. As potentially debilitating, donors might confer only restricted gifts, to guarantee the use to which their support is put.

An over-arching theme is a desire to avoid the burden of adhering to legal constructs that preclude responding to changed circumstances and force the misuse and wasting of charitable assets. How health care is delivered has evolved from predominantly acute care in the 1960s to predominantly outpatient care today and will likely be comprised of pharmaceuticals and “continuum care” tomorrow. What health care is delivered depends upon the ever-changing demographics of the community, including the age, education, and socio-economic status of the population. Where health care is delivered must respond to population shifts. Those responsible for directing the uses of the privately-held assets that substantially comprise the U.S. health care system must have the flexibility and autonomy to make the timely decisions necessary to respond to these changes. On the other hand, their power should not go unchecked.

Thus, it is no surprise that charitable trust law presents a potentially appealing source from which to fill the legal void attorneys general face when concerned about a nonprofit board’s deviation from its mission. Nonetheless, trust law is ill-suited to address the myriad questions that arise in a corporate context: Does fealty to mission require merely that the nonprofit subscribe to a valid charitable purpose or must it assiduously and forever adhere (absent state consent) to the mission originally articulated in the corporation’s formation documents? If the answer is somewhere in between, so that nonprofit boards may variously deploy assets in response to significant market changes, the question becomes at what point in this middle ground state approval is required.

3. It Is Impracticable To Apply Charitable Trust Law to Nonprofit Corporations

Finally, we explore the potential impact of wholesale importation of charitable trust standards to govern oversight of the modern commercial

enterprise. Strictly applied, charitable trust law would invite detailed judicial review of all board decisions that implicate the nonprofit's mission and perhaps application of the *cy pres* doctrine to determine whether the prior business activity under review has become impossible, inexpedient, or impracticable,²⁷⁴ and whether the new business plan is as "near as possible" to the settlor's original intent.²⁷⁵ This approach would obviously pose enormous practical difficulties for the court. For example, ascertaining whether the settlor's original intent has become impossible, inexpedient, or impracticable to fulfill in the context of a multi-million dollar commercial enterprise attempting to respond to a rapidly changing health care market would be an enormously complicated, perhaps intractable, inquiry. Also troublesome is the artificiality of determining the "settlor's intent" (are the settlors current taxpayers and donors or those who supported the entity at the time of its establishment, or an aggregation of all taxpayers?) from articles of incorporation that can be decades if not a century old. It makes little sense to require the corporation to remain as "near as possible" to its original mission when to do so might result in economic demise, represent a misallocation of significant health care resources, or is simply not in the best interest of the community that is the current beneficiary of the nonprofit's activities.

First, as is illustrated by this discussion, the notion that a trust comprises a third party beneficiary contract between settlor and trustee²⁷⁶ is a legal construction ill-conceived for the charitable corporate context. Because the beneficiaries of the charitable corporation/trust cannot be identified, they must be represented in *parens patriae* by the attorney general. Unlike the private trust context, where the beneficiaries have a clear incentive to monitor the trustees, and to litigate if the trustees fail in their obligations,²⁷⁷ attorneys general have neither access to the information necessary to monitor the charitable corporation/trust, nor the resources necessary to determine or pursue the beneficiaries' interests.

Second, using trust law to oversee governance of nonprofit corporations is inefficient. While trust law in the *private* trust context is

274. See generally FISHMAN & SCHWARZ, *supra* note 144, at 100.

275. *In re Multiple Sclerosis Serv. Org.*, 68 N.Y.2d 32, 35 (1986).

276. Robert H. Skitoff, *An Agency Costs Theory of Trust Law*, 89 CORNELL L. REV. 621 (2004). Because the law of trusts incorporates both in rem benefits of property law and the "in personam flexibility of contract law," the alternative theory of trust law is grounded in property law—that the trust conveys a beneficial interest in the trust property to the beneficiaries. *Id.* at 629.

277. This argument obviously does not apply if the beneficiaries are as yet unborn, or are incompetent. *Id.* at 663, 668.

arguably the most efficient means of achieving the best interests of the trusts beneficiaries “within the settlor’s legally permissible objectives,”²⁷⁸ critics increasingly question whether trust law is efficient for charitable trusts.²⁷⁹ Irrespective of how that debate is resolved, trust law unquestionably should not extend to the nonprofit corporation.

Finally, much of trust law, specifically, the duty of care, attends to beneficiaries’ presumed lower “risk tolerance”; trustees are charged with the preservation of the trust assets. By comparison, corporate law’s business judgment gloss on the duty of care seeks to preserve boards’ risk-taking instincts, all the better to pursue opportunities that will maximize profits.²⁸⁰ Neither model is ideal for the nonprofit corporation, whose primary goal is community service, irrespective of its profit potential, and without necessary reference to asset valuation. On the other hand, nonprofit health care providers are acutely aware that they cannot accomplish their mission without financial wherewithal.

In sum, application of trust doctrine to nonprofit corporations is analytically strained—no identifiable “settlor,” beneficiaries, or “trust instrument” exists, so attorneys general and courts engage in a fictitious analysis that is confusing at best. At worst, applying the inflexible standards of trust law can be devastating to the economic survival of a significant health care enterprise and might cause dissipation of the corporation’s assets, which conflicts precisely with the ostensible goal of charitable trust.

4. *Directors’ Duties in Transitions of Nonprofit Corporations*

The case law evaluating directorial decisions regarding organic changes gives mixed and conflicting guidance with respect to the proper role of mission in that process. Most states appear to accept in principle that corporate fiduciary standards should apply to nonprofit directors. Yet developing a framework for allowing consideration of charitable purposes in appropriate cases remains elusive. As the discussion of applying the *Revlon* principle to nonprofit conversions illustrates,²⁸¹ in some

278. Skitoff explains that the trustees’ duty of impartiality as among different classes of beneficiaries whose interests may conflict is the “salient distinguishing characteristic of trust law as organizational law.” *Id.* at 652. This concept is likely inapplicable to the charitable corporation unless, in the hospital context, patients and doctors are conceived of as competing classes of beneficiaries.

279. *See, e.g.,* Eisenstein, *supra* note 273. Eisenstein suggests that in some circumstances, the public is best served by allowing the trust to fail. *Id.* at 1781-83.

280. *See* Skitoff, *supra* note 276, at 656-57.

281. *See supra* note 200.

circumstances strict application of the corporate standard may fail to take into account mission-related issues that should be appropriately considered by directors in evaluating changes. At the same time, where statutes or common law invite consideration of mission-related factors, there is a real risk that regulators, courts and attorneys general will substitute their judgments for the discretion of directors.

5. *Impact*

For managers of nonprofit health care enterprises, legal uncertainty breeds inefficiency and impairs pursuit of charitable goals. Most obviously, threats of second-guessing by charity regulators impede managers' ability to deploy assets and plan strategically in a dynamic health services market. The interventions by the Minnesota and New York Attorneys General with Allina, MEETH, and Littauer created uncertainty that pervades the business decisions of the entire nonprofit health care sector in those states. One can scarcely doubt that management, acutely aware that attorneys general may question routine business expenditures, now asks itself how everyday decisions might appear if they were widely publicized.²⁸² Moreover, interventions that question long-established business structures raise significant policy questions. For example, the demand that Allina spin off its HMO implicates the permissible relationships among the component parts of an integrated delivery system and ultimately whether an integrated delivery system is even possible. Further, the Attorney General's position in *Allina* questions whether the corporate purposes of a system member may be subsumed by the system's over-arching mission.

We have also seen that the role of mission in informing directors' decision-making is quite ambiguous. When assessing whether boards have satisfied their fiduciary responsibilities, courts and charity regulators sometimes invoke mission responsibilities, and sometimes ignore them. For example, the *MEETH* and *Littauer* decisions send mixed messages about boards' autonomy to interpret and direct their mission in New York.

282. As described in a recent *New York Times* article: "Charities and foundations have been bracing for stronger regulatory intervention in their affairs, and many are already taking steps to beef up their governance. . . ." Stephanie Strom, *Questions About Some Charities' Activities Lead to a Push for Tighter Regulation*, N.Y. TIMES, Mar. 21, 2004, at 23; see also Brody, *Whose Public?*, *supra* note 13, at 943 (describing numerous examples of activism by attorneys general and concluding, that "should charities too quickly accede to state demands over matters of discretionary governance, the sector as a whole can see a degradation in charities' willingness to take risks, and in volunteer board members' willingness to serve").

Likewise, the legal posture assumed by the three attorneys general who challenged Banner's re-deployment of its assets across state lines threatens the very existence of multi-state systems—these systems risk losing their assets if they attempt to move them out of the local communities in which they are currently invested. Further, a public policy requiring that the assets of a charitable corporation constitute a trust belonging to the public makes the entity less attractive as a potential strategic partner, which may negatively affect the value of those assets.

At a more quotidian level, attorneys general's overzealous prosecutions may deter service on boards by just the kind of experienced professionals that both state and federal regulators hope to see actively engaged in corporate oversight. Further, nonprofit boards may be made excessively risk-averse by the specter of overreaching regulatory oversight. Studies suggest that they are prone to overestimate risk and be less inclined to pursue innovative business strategies. Lacking financial incentives, volunteer nonprofit directors appear to be driven by a combination of social norms and their personal loyalty to the mission of the institution they serve.

In this environment, it is important to remember the norm-shaping impact of law. As scholarship has stressed, an important objective of the law is to shift social norms and social meaning.²⁸³ As we have argued, this role is particularly significant in nonprofit fiduciary law because of the absence of financially interested monitors and the ambiguity surrounding the objectives guiding corporate agents. With respect to health care nonprofits, we conclude that the legal milieu in which they operate seems inimical to fostering good stewardship. A legal regime that is slow to insist on director vigilance but intrudes on decisions of central importance regarding mission likely reinforces directorial abdication.

Finally, we consider the law of nonprofits from an institutional perspective. Attorneys general play a complicated role in the current environment. They fill a variety of roles with respect to the nonprofit sector: prosecutor, consumer advocate, public representative as *parens patriae*, supervisor of charitable trusts, regulator, and politically accountable officer of the state.²⁸⁴ Abhorring a vacuum, many have assumed a multi-faceted role in the oversight of the governance of

283. See, e.g., Cass Sunstein, *supra* note 152, at 2043; Eric A. Posner, *Symbols, Signals, and Social Norms in Politics and the Law*, 27 J. LEGAL STUD. 765 (1998).

284. See FISHMAN & SCHWARZ, *supra* note 144, at 254-56 (enumerating common law and statutory powers of the attorney general regarding charitable corporations and trusts); Brody, *Whose Public?*, *supra* note 13, at 938-39.

nonprofit organizations that extends far beyond enforcing fiduciary and trust principles. As described in Part II, this has led some down a path to micromanaging business affairs, seeking direct appointive power over board positions, and parochial control over the redeployment of charitable assets. From a health care policy perspective, it has also resulted in de facto centralization of several important regulatory functions. Attorneys general have used their leverage over nonprofits in asset sales, conversions, and mergers to direct the geographic and service dimensions of the charitable sector. As they candidly admit, attorneys general offices see themselves as assuring an appropriate allocation of society's scarce charitable resources, and they freely use legal tools (and capitalize on the doctrinal ambiguities in the law) to do so. One must seriously question, however, whether a mandate exists for reposing so much discretion in that office and, even if it did, whether attorneys general command the resources to responsibly assume it.

We also speculate that attorney general activism may have untoward spillover effects on other governmental actors. Given the centralization of authority in the attorneys general, it is perhaps not surprising that states have not felt impelled to forthrightly consider the difficult issues posed by the changing landscape of charitable health care. Most states have weakened or abandoned certificate of need review; hospital closings are not closely supervised; and the preservation of the health care safety net is largely unattended. Thus, few departments of health actively supervise the geographical locations of charitable health facilities or the range of services they offer. These are public policy tasks essential to maintaining a viable health care system that are properly vested in state departments of health, which should not abdicate their responsibilities to attorneys general offices that are wholly unequipped to fill this function.

V. CONCLUSION: A PROPOSAL FOR MISSION PRIMACY

Our review of the application of fiduciary and charitable trust law to nonprofit health care corporations has identified a number of significant problems. First, there is widespread confusion about the boundaries between those bodies of law. Difficulties are compounded by the impact of those laws on the health sector—notably the uneasy fit of importing wholesale for-profit corporate principles to govern entities having decidedly different attributes and goals, and the inappropriateness of using rigid trust concepts to guide management of dynamic commercial enterprises. This confusion has led to opportunism among certain attorneys general who have sought judicial relief, which inappropriately transfers power over business and mission decisions to them. It has also

spawned uncertainty in business planning that may weaken the nonprofit sector's ability to serve its societal purposes. Finally, ambiguity about the role of boards and attorneys general may have diverted health policymakers and regulators from dealing squarely with the central task of fairly and efficiently allocating charitable assets.

As discussed throughout this Article, commercial nonprofits in health delivery and payment must anticipate and respond to demographic shifts, reimbursement reform, and technical innovation. Attorneys general should not usurp departments of health and insurance, which are the agencies properly responsible for ensuring that the business climate in which providers and payors operate can supply high quality and affordable health care. To give a concrete example, attorneys general's insistence that nonprofit hospitals forever adhere to their original purpose of serving the local community as a free-standing acute care facility can have detrimental long-term consequences for that community's access to appropriate health care. It may force them to forgo the efficiencies, stability, and capital accruing from affiliation with a financially strong national system; or it may deny the community a sensible re-deployment of its charitable assets, e.g. from acute care hospital services to disease prevention or outpatient clinics. A final concern, focusing on institutional competence, is that the attorney general's office lacks the expertise, resources, and legal mandate to micro-manage business affairs of commercial enterprises or to macro-manage the allocation of health services within the community.

This Part offers several core principles that should guide future judicial, legislative, and regulatory adjustments affecting nonprofit health care organizations. Admittedly, few of the problems we have identified are subject to easy correction by isolated changes, e.g., judicial interpretation of doctrine, attorney general forbearance, or modest legislative adjustments. What we offer below, however, can provide a useful first step: guidance as to the central issues that should be addressed in redefining nonprofit accountability so as to ensure that governmental oversight is both coordinated and appropriate.

A. Principles for Reorienting Nonprofit Organization Law and Policy

The complex tangle identified in this article of confused doctrine, lack of managerial accountability, and overreaching by attorneys general poses challenges for all branches of government dealing with nonprofit governance. Because there is so much variation in state law in this area, a precise road map for implementing change is impossible. However, we identify below three core principles to guide legislatures, courts, and regulators as they move toward developing governance standards for

nonprofit enterprises in health care.

Our analysis takes as a starting point that the evidence of modest benefits flowing from the nonprofit sector supports continued reliance on legal mechanisms that enable and require those institutions to achieve their charitable missions.²⁸⁵ At the same time, the record of community benefits is not so compelling as to support use of legal tools to preserve nonprofit entities at any cost.²⁸⁶

1. Ownership and Accountability

The fundamental question of who, broadly speaking, “owns” the nonprofit corporation merits close attention. Many questions, such as defining and evaluating community benefit, ascertaining directors’ obligations under changing conditions, and enumerating the rights and obligations of controlling members, cannot be addressed without a clearer understanding of to whom (or what) nonprofit fiduciaries should be accountable. As discussed above, corporate scholars continue to debate whether for-profit governance should adopt a model of shareholder primacy, director primacy, or some other objective function reflecting societal goals that underlie the corporate form. In the nonprofit sector, the debate has scarcely moved in the last twenty years.²⁸⁷ The absence of discussion is remarkable because the issue is, if anything, more pointed for nonprofits than for commercial profit corporations. That is, lacking shareholders, the candidates for primacy are a more diffuse and amorphous group: the class of beneficiaries to be served by the charity; the directors who manage those objectives; members, where present; donors and taxpayers; or the representative of the public beneficiary class such as the Attorney General. Moreover, the absence of the disciplining effect of a capital market or vigilant, interested shareholders to vindicate abuses in court exacerbates the agency costs inherent in the nonprofit form. At the same time, the similarities between the commercial nonprofit sector and the for-profit sector are also striking. Commercial nonprofits do not rely

285. See *supra* Part I.

286. See generally M. Gregg Bloche, *Should the Law Prefer Nonprofits?*, in *THE PRIVATIZATION OF HEALTH CARE REFORM* 186 (M. Gregg Bloche ed., 2003).

287. Evelyn Brody’s impressive body of scholarship has addressed these issues. As she summarizes the dilemma, “[N]onprofit ‘accountability’ is a difficult question. Accountable to whom? For what? While nonprofits as suppliers of goods and services must respond to their customers, and as employers must respond to their professional staffs and employees, the same types of resource dependency affect for-profit firms.” Brody, *Agents Without Principals*, *supra* note 18, at 534-35 (footnote omitted).

heavily on donations and, from a financial standpoint, are driven by a need to satisfy customers in the marketplace.²⁸⁸

The failure of courts and commentators to resolve questions of ownership and mission accountability is in part explained by the plurality of competing interests. Starting with the perspective that tax expenditures and legal support create a strong public entitlement, some regard the public at large (or its representative) as the appropriate locus of accountability. Others identify as the appropriate party in interest the beneficiaries of the nonprofit's charitable mission. Still others advocate including the "patrons"—donors and customers who provide the financial wherewithal to fulfilling the nonprofit's charitable mission. Finally, there is the perspective of the "sponsor" or "member" of the nonprofit corporation, whose control and support keep the enterprise running. Choosing among these competing parties in interest is ultimately a normative and political question that underlies any workable definition of "accountability." As Evelyn Brody framed the issue: "Who are the 'principals' to whom society *wants* the charity to answer . . . ?"²⁸⁹ Like many before us, we will dodge that question. Instead we offer a framework for allowing courts and legislatures to address the issue by allocating presumptive decisionmaking authority to those entrusted with serving the nonprofit's purposes, but insisting that they follow clearly articulated mission statements.

As a general guiding principle, we suggest that "mission primacy" should be recognized as a central objective of the nonprofit enterprise with the corollary that directors enjoy presumptive deference in defining and, within limits, amending that mission. This focus would incorporate mission-centered values into interpretations of the traditional fiduciary duties of care and loyalty. At the same time, like the model of "director primacy" advanced for proprietary corporations,²⁹⁰ it would preserve managerial discretion to balance the various constituents of the nonprofit firm including donors, consumers, and the community. Consequently, this standard would accommodate the relational imperatives of the modern business environment in health care by deferring to managerial expertise, avoiding interference with discretionary judgments, and encouraging

288. *Id.* at 535 ("Effectively, then, nonprofits are generally as untethered to their donors as large for-profit firms are to their shareholders."); *id.* at 536 ("In many ways, the formal legal and economic differences between nonprofit organizations and proprietary firms are more of degree than of kind.").

289. *Id.* at 512.

290. See BAINBRIDGE, *supra* note 148, at 192-240.

appropriate risk-taking. Finally, mission primacy accounts for the particular circumstances of nonprofit governance because it preserves the central values of trust and volunteerism that are needed to reinforce legal duties.

Mission primacy, then, would extend the concept of the duty of obedience to underscore directors' core responsibilities as stewards of a nonprofit enterprise to advance its public purpose. It has been seen that by embracing the for-profit corporate model, nonprofit governance law has often blindly applied fiduciary norms that neglect mission values entirely. Our approach would hold directors to a fiduciary standard that requires them to weigh mission considerations in all decisions. At the same time, however, by requiring courts to grant deference to directors' judgments, the rule would reduce risks of unwarranted judicial interference and preserve the norm-shaping role of fiduciary law. Thus, mission primacy would allow legitimate mission-centered factors to override corporate fiduciary standards in some cases while imposing a more exacting standard of care or loyalty where mission issues predominate. Several examples will serve to illustrate the way in which mission primacy would affect application of fiduciary duties.

In cases involving organic change, such as the hospital closure at issue in *MEETH*, mission primacy would mandate consideration of and deference to the board's evaluation of mission, and its determination of the most appropriate means to accomplish that mission.²⁹¹ Where a board's actions are questioned under duty of care or loyalty standards, mission factors may help give content to the inchoate considerations that contribute to the board's deliberation. Likewise, mission primacy may compel deference in the economically important cases involving multi-state charitable corporations consolidating their holdings, or whose mission calls for the reallocation of revenues from profit-producing facilities to facilities in financially distressed communities, irrespective of whether such aid crosses state lines.

Mission primacy would not affect the attorney general's extant authority to ensure compliance with the duty of care by appropriate due diligence, particularly when a board decides to dispose of the charity's assets. However, it would prevent courts from blindly applying corporate

291. In this regard mission primacy would likely have required a less categorical evaluation of purpose in *MEETH*. See *Manhattan Eye, Ear & Throat Hosp. v. Spitzer*, 715 N.Y.S.2d 575, 595 (Sup. Ct. 1999) ("[I]t is sophistry to contend that this means that MEETH is not seeking a new and fundamentally different purpose . . ."). This approach is more in keeping with the court's approach in *Littauer*, discussed *supra* notes 12, 77, 79 and accompanying text.

principles in a manner that overlooks mission entirely. For example, charitable corporations selling their assets frequently find themselves courted by prospective buyers with diverse missions, whose offers vary dramatically. As discussed above, some would apply the *Revlon* doctrine to nonprofits, thus imposing an obligation on directors to sell for the highest price and ignore mission-based considerations, once a decision to sell the corporation has been made.²⁹² Mission primacy would avoid this trap, allowing a nonprofit board to weigh mission preservation against price, and to select a buyer whose offer best accommodates both of these concerns.

2. Clarify the Charitable Trust/Corporate Law Boundary

This Article counsels strongly against states' reliance on charitable trust law to regulate nonprofit assets, except, of course, where an express trust exists. We have argued that applying charitable trust law to corporate assets is doctrinally unsound and produces outcomes that potentially waste, rather than preserve, scarce charitable assets. The alternative approach, adopted by most of the courts to have directly addressed the issue, is to look to corporate law as the foundation for the law governing all aspects of charitable corporations. While this reflects our preferred doctrinal path, recognition of the mission primacy principle proposed above is necessary to assure that both boards and charity regulators observe core nonprofit purposes.

Clarifying that corporate law governs disputes involving nonprofit business decisions would remove an important obstacle to efficient business planning by multi-state entities in most cases. Thus, corporate analysis with a focus on mission would likely have resulted in Banner being able to re-deploy its assets to Colorado and Arizona with relative ease. Except for the circumstances where express trusts existed, Banner's holdings should not otherwise have been impressed with a trust—traditional trust doctrine does not support the imposition of a trust on the basis of generalized donations or tax subsidies. In some circumstances, legislative action would be required to assure corporate principles prevail; for example, in the Banner litigation, the South Dakota Supreme Court recently decided that the enactment of the nonprofit corporate statute did not preempt the pre-existing charitable trust statutory or common law.²⁹³ Clarification of the rather murky doctrine of implied charitable trust in

292. See *supra* note 200 and accompanying text.

293. *Banner Health Sys. v. Long*, 663 N.W.2d 242, 247 (S.D. 2003).

those states that recognize the concept would also serve to remove uncertainty in this area.

Recognizing that the corporate standard has not been a model of successful prophylactic law in the for-profit sector, it might be appropriate to adjust fiduciary standards applicable to nonprofits in some circumstances. For example, most state nonprofit corporate statutes bar loans to directors, a development that long preceded parallel developments in the for-profit sector under Sarbanes-Oxley law. Moreover, an evolving recognition that the business judgment rule's impact should be tempered by requiring good faith and informed decision-making should be encouraged.²⁹⁴ Further, administrative improvements may well be needed to assure that fiduciary derelictions are detected and remedied. At the same time, enhanced enforcement mechanisms must be carefully designed so as not to undermine the social norms that play a critical role in assuring fiduciary performance.²⁹⁵

While regulatory interventions in the health care sector would have to be sensitive to the multiple regulatory entities sharing oversight of the sector, it is unquestionably the case that states need to invest the resources in some charity agency that will provide better regulatory guidance to the nonprofit sector, and will review the increasingly available information about nonprofit entities to detect potential problems.

3. Clarify and Delineate State Agencies' Supervisory Responsibilities

Viewed from the perspective of health care policy, the most important—and most vexing—public policy question emanating from our analysis of charitable nonprofit law is how the law can best achieve the appropriate distribution of health care resources. These concerns undoubtedly drive attorneys general to undertake many of the actions for which we take them to task in this Article. Simply stated, the problem we identify in this regard is one of institutional competence and transparency. To the extent that there are market failures, there are alternative and more focused means of regulation and allocation including licensure, certificate of need regulation, and subsidies from state departments of

294. See cases listed *supra* note 159 and accompanying text.

295. James Fishman has usefully advanced the idea of employing a charity commission that would operate under the aegis of the attorney general to review complaints about charitable corporations. Fishman, *supra* note 127, at 266-72 (reviewing the scope of proposals and changes made by nonprofit experts). The attorney general's office would only be required to become involved where a viable complaint could not be resolved at the commission level. *Id.* at 272-75.

health; exercise of the state's tax-exemption powers; and contracting by state entities funding health services. Without expressing a preference for any particular regulatory regime, we believe that policy-making through these agencies is preferable because it is more likely to identify and evaluate deficits in safety net services. By contrast, allocating broad and unrestricted discretion to attorneys general to make allocative decisions behind a veil of corporate or charitable trust litigation appears to be a recipe for ad hoc and inefficient decision-making.²⁹⁶ Whatever oversight agenda of the charitable sector a state attorney general decides to pursue, we view it as essential that the office clearly articulate its public policy concerns, expectations, and standards of review. The need for such guidance is particularly acute if attorneys general intend to assert standing on mission issues, for which there is virtually no precedent to guide nonprofit boards.

Our final point concerns problems that flow from attorneys general seeking to replace directors, or to appoint a "super-board" with veto or special administrative powers. Such appointments provoke charges of political cronyism, which threatens the private and necessarily non-political nature of nonprofit tax-exempt charities. More important, political appointments to charitable boards create inherent conflicts for the appointees—whether they are bound to act as they independently believe appropriate to fulfill their fiduciary duties, or whether they are required to pursue the preferences of the regulator who appointed them. The inevitable blurring of the line between public and private accountability occasioned by these interventions threatens to undermine director discretion; quite startlingly, rather than improving accountability to mission, it shifts director fealty to the interests of political officials.

296. We acknowledge the myriad problems surrounding efficient deployment of charitable resources. Certificate of Need processes, which were originally intended to reduce health care expenditures and eliminate inequitable distribution of resources by regulating significant capital investments on new facilities or equipment, currently exists in fewer than half the states. See M. Gregg Bloche, *The Invention of Health Law*, 91 CAL. L. REV. 247, 298 (2003); Lauretta Higgins Wolfson, *State Regulation of Health Facility Planning: The Economic Theory and Political Realities of Certificate of Need*, 4 DEPAUL J. HEALTH CARE L. 261, 261-62 (2001).

The Battle over Self-Insured Health Plans, or “One Good Loophole Deserves Another”

Russell Korobkin, J.D.*

Enacted in 1974, the federal Employee Retirement Income Security Act (ERISA)¹ has been a major roadblock to advocates of increased regulation of health insurance benefits in the era of “managed care.” Originally drafted as a pension law, ERISA, as enacted, applies to all fringe benefits provided by private employers to their employees. The statute shields benefit plans, including health insurance, from state regulation in two ways. First, ERISA’s “preemption” clause prohibits state laws that “relate to” employee benefit plans.² Second, although ERISA’s “savings clause” exempts state laws that “regulate insurance” from the statute’s preemptive force,³ this exception is in turn limited by the “deemer clause,” which prevents state insurance regulations from reaching employer health care benefits plans (EHBPs) that are self-insured,⁴ as opposed to those that purchase insurance coverage from a third party. Put another way, ERISA obstructs state regulation on two levels: The statute partially shields all EHBPs from state regulation, and self-insured EHBPs enjoy an enhanced level of protection.

A large chorus of critics has lodged two different types of complaints about ERISA. On one hand, critics contend that managed care arrangements threaten consumer health and that the expansion of these insurance systems requires the government to police health insurers more closely. ERISA preemption impedes possible state regulatory efforts.⁵ On

* Professor of Law, UCLA and Faculty Associate, UCLA Center for Health Policy Research. Comments on earlier drafts from Tom Baker, John Day, Mark Hall, Peter Jacobson, John Jacobi, Bill Sage, Rich Saver, and workshop participants at the UCLA Center for Health Policy Research are gratefully acknowledged. Invaluable research assistance was provided by Heather Richardson.

1. Employee Retirement Income Security Act of 1974, Pub. L. No. 93-406, 88 Stat. 829 (codified as amended at 29 U.S.C. §§ 1001-1461 (2000)).

2. 29 U.S.C. § 1144 (2000).

3. *Id.* § 1144(b)(2)(A).

4. *Id.* § 1144(b)(2)(B).

5. See, e.g., Robert Covington, *Amending ERISA’s Preemption Scheme*, 8 KAN. J.L. & PUB.

the other hand, to the extent that ERISA's savings clause enables state regulation of managed care to avoid preemption, critics complain that ERISA creates an inequitable two-tiered regulatory system, in which employees in "insured" plans receive protections of state law denied to employees in "self-insured" plans.⁶

In the past three terms, two important United States Supreme Court decisions, *Rush Prudential HMO v. Moran*⁷ and *Kentucky Ass'n of Health Plans v. Miller*,⁸ expanded the scope of ERISA's savings clause, giving states greater latitude to regulate managed care. A third decision, *Aetna Health Inc. v. Davila*,⁹ added to the significance of *Rush Prudential*. At the same time, the Court did not change its interpretation of the deemer clause: In fact, there have been no Supreme Court rulings dealing with the deemer clause since 1990.¹⁰ The result is that, as the scope of ERISA preemption has contracted, the gap in regulatory protections enjoyed by employees in insured and self-insured plans has expanded. An employer's decision about whether to purchase third-party insurance or to self-insure its employees' health care expenses has taken on increasing significance, creating two competing incentives. Employers who wish to avoid the costs associated with state regulation have a greater incentive to establish self-insured EHBPs, and the supporters of regulation have more incentive than ever to fight self-insurance.

A change in federal law would moot this issue. Congress could amend ERISA to provide equal legal treatment for employees in insured and self-insured EHBPs, or the courts could reinterpret the savings and deemer

POL'Y 1 (1999); Mark A. Edwards, *Plan Protections for ERISA Self-Insured Employee Welfare Benefit Plan Participants: New Possibilities for State Action in the Event of Plan Failure*, 1997 WIS. L. REV. 351; Margaret G. Farrell, *ERISA and Managed Care: The Law Abhors a Vacuum*, 29 J. HEALTH L. 268 (1996); James Saya, *Removing a Roadblock to Reforming Health Care*: New York State Conference of Blue Cross & Blue Shield Plans v. Travelers Insurance Company, 3 CONN. INS. L.J. 127 (1997); Jana K. Strain & Eleanor D. Kinney, *The Road Paved with Good Intentions: Problems and Potential for Employer-Sponsored Health Insurance Under ERISA*, 31 LOY. U. CHI. L.J. 29 (1999).

6. See, e.g., Farrell, *supra* note 5; Troy Paredes, *Stop-Loss Insurance, State Regulation, and ERISA: Defining the Scope of Federal Preemption*, 34 HARV. J. ON LEGIS. 233 (1997); William H. Pitsenberger, "An Apparently Irrational Distinction": A Suggestion for Using Equal Protection Arguments To Overcome Conflicts in ERISA Preemption, 32 J. HEALTH L. 307 (1999); Strain & Kinney, *supra* note 5.

7. 536 U.S. 355 (2002).

8. 123 S. Ct. 1471 (2003).

9. 124 S. Ct. 2488 (2004).

10. That case was *FMC Corp. v. Holliday*, 498 U.S. 52 (1990).

clauses in a way that would eliminate or minimize the distinction. Neither of these scenarios is likely in the near future, however. Although a federal “Patients’ Bill of Rights” that includes amendments to ERISA might one day be enacted, none of the leading legislative proposals would eliminate special protections for self-insured health benefits plans. A change in the relevant judicial interpretations of ERISA is even less likely because the Congress that enacted ERISA clearly intended for the statute to protect self-insured benefit plans from state regulation.

With the statutory difference in treatment between insured and self-insured plans unlikely to disappear any time soon, the reach of state regulation of health insurance rests on how many businesses choose to self-insure their EHBPs. At present, the number is surprisingly high, owing in part to the popularity of a loophole in ERISA that enables employers without sufficient resources to bear the risk of their employees’ health care costs to purchase “stop-loss” insurance—a product that reimburses the employer for costs above a specified threshold amount—and still qualify for ERISA’s protection from state regulatory requirements.

ERISA also contains a second, less well-understood loophole, however, that states can exploit to minimize the number of EHBPs able to invoke ERISA as a shield against state regulation of health insurance: ERISA’s text and structure permit states to regulate the terms and conditions of stop-loss insurance. Some states have already taken advantage of this loophole to a limited extent, although none has exploited it as fully as they might. Correctly interpreted, ERISA provides state regulators with the tools to effectively staunch employers’ ability to manipulate the statute’s preemption provision for the sole purpose of avoiding state insurance mandates.

This Article explores the battle between employers who seek to maximize and state regulators who seek to minimize the scope of ERISA preemption. Part I describes the relevant statutory structure of ERISA and the implications of that structure, with emphasis on the three recent Supreme Court decisions that increased the legal importance of the distinction between insured and self-insured health plans. Part II describes how the availability of stop-loss insurance allows employers to exploit a loophole in ERISA’s deemer clause in order to avoid exposure to state regulation. Part II also defends the role federal courts have played in permitting the exploitation of this loophole on the ground that it is consistent with ERISA’s text and its underlying congressional intent. Part III describes how a loophole in ERISA’s savings clause allows state regulators to close the deemer clause loophole. It argues that, as is true for the deemer clause loophole, the text and underlying intent of ERISA

counsels that the courts should not intervene to block the exploitation of this loophole. When it enacted ERISA, Congress established a muddled set of rules. Properly understood, ERISA's ground rules should allow employers and regulators to battle to a stalemate.

I. THE INSURANCE-SELF-INSURANCE GAP

A. ERISA's Structure

In the early 1970s, as a response to a number of failures of employer-sponsored pension funds, Congress proposed to replace a patchwork of state pension plan regulations with a federal regulatory structure.¹¹ In the process of drafting ERISA, however, Congress expanded the proposal's scope to preempt state laws that relate to any "employee benefits plan," including employer-provided health insurance.¹² But while ERISA, as enacted, provides detailed substantive regulations of pension plans, it includes virtually no substantive regulation of EHBPs,¹³ leaving such plans largely unregulated, save for a few recently-enacted federal health benefits regulations, such as minimum hospital length-of-stay rules for childbirth,¹⁴ mental health care coverage requirements,¹⁵ and limits on preexisting condition exclusions.¹⁶

There is one significant exception, however, to ERISA's preemption of state laws that relate to health care plans. The statute's savings clause protects from preemption state laws that "regulate[] insurance."¹⁷ Although there is no legislative history explaining the addition of the

11. See, e.g., Patricia Butler, *ERISA Preemption Manual for State Health Policy Makers*, 2000 NAT'L ACAD. FOR STATE HEALTH POL'Y 5.

12. Although most ERISA benefits plans are employer-sponsored, other entities, such as labor unions, can also sponsor such plans. See 29 U.S.C. § 1144 (2000).

13. See *Shaw v. Delta Air Lines, Inc.*, 463 U.S. 85, 91 (1983). ERISA and its regulations do provide for a number of procedural regulations of EHBPs. For example, administrators of EHBPs are fiduciaries and have a range of obligations as such, see 29 U.S.C. § 1002 (2000); EHBPs must provide summary plan descriptions to participants, see 29 C.F.R. 2520.102-3 (2004); and ERISA provides plan participants with a federal cause of action to recover promised benefits that the plan fails to provide, 29 U.S.C. § 1132(a) (2000).

14. Newborns' and Mothers' Health Protection Act of 1996, 42 U.S.C. § 300gg-4 (2000).

15. Mental Health Parity Act of 1996, 29 U.S.C. § 1185a (2000), 42 U.S.C. § 300gg-5 (2000).

16. 26 U.S.C. § 9801 (2000); 29 U.S.C. § 1181 (2000); 42 U.S.C. § 300gg (2000).

17. 29 U.S.C. § 1144(b)(2)(A) (2000).

savings clause,¹⁸ its presence clearly suggests that Congress did not intend for the preemption principle to go so far as to subvert traditional, core areas of state regulatory authority.

In ERISA's text, however, the savings clause is followed by the deemer clause,¹⁹ with the latter limiting the scope of the former. The deemer clause provides that employee benefit plans "shall [not] be deemed to be an insurance company or other insurer . . . or to be engaged in the business of insurance . . . for purposes of any law of any State purporting to regulate insurance companies [or] insurance contracts"²⁰ Thus, ERISA preserves the traditional right of states to regulate the insurance industry, but those regulations may not extend to cover EHBPs, even though EHBPs often serve an insurance function and might otherwise find themselves subject to state laws governing insurance.²¹

The deemer clause's limitation on the scope of the savings clause makes sense only in the context of one of ERISA's underlying goals: providing a uniform legal structure for employers that operate in multiple states.²² Neither ERISA nor any other federal law requires employers to provide any fringe benefits. ERISA's supporters thought that by protecting large, multi-state employers from the burden of dealing with multiple sets of regulatory requirements, employers would be more likely to provide fringe benefits.²³ The end result of Congress's attempt to balance the competing goals of deferring to traditional state functions and promoting legal uniformity is that states may regulate insurance companies, even if such regulations indirectly "relate to" EHBPs because such plans purchase insurance, but states may not directly regulate the plans themselves.

Supporters of increased regulation criticize as inequitable or outright illogical the fact that EHBPs enjoy greater freedom from state control than

18. *Id.*; *Metropolitan Life Ins. Co. v. Massachusetts*, 471 U.S. 724, 745 (1985).

19. 29 U.S.C. § 1144(b)(2)(B) (2000).

20. *Id.*

21. *Cf.* HOUSE COMM. ON EDUC. & LABOR, ERISA OVERSIGHT REPORT OF THE PENSION TASK FORCE OF THE SUBCOMM. ON LABOR STANDARDS 10 (Comm. Print 1977) (concluding that the deemer clause "create[s] what may amount to a legal fiction in a given circumstance" in which a plan engages in insurance activities).

22. *See Ingersoll-Rand Co. v. McClendon*, 498 U.S. 133, 142 (1990) (describing ERISA's purpose of ensuring that benefit plans are not subject to divergent regulatory schemes in different states); *Pilot Life Ins. Co. v. Dedeaux*, 481 U.S. 41, 46-47 (1987) (same).

23. Donald T. Bogan, *ERISA: The Savings Clause, 502 Implied Preemption, and State Law Remedies*, 42 SANTA CLARA L. REV. 105, 118 n.51 (2001); Farrell, *supra* note 5; Patricia M. Ochmann, *Managed Care Organization Manage To Escape Liability: Why Issues of Quantity v. Quality Lead to ERISA's Inequitable Preemption of Claims*, 34 AKRON L. REV. 571, 580 (2001).

do insurance companies when the two types of organizations serve the same purpose of guaranteeing the provision of needed medical care.²⁴ This disparate treatment is not irrational, however, because state laws regulating insurance companies impose a less severe administrative burden on multi-state employers than would state laws directly regulating EHBPs themselves. For a multi-state employer that wishes to self-insure its employees' health care benefits, inconsistent state regulatory requirements, if permissible, would impose upon it the cost of developing a separate insurance plan for its employees in each state.²⁵ If that multi-state employer purchases insurance for its employees, it might have to purchase different insurance policies for employees in each state, but the employer need not concern itself with the task of complying with different state regulations—such responsibility would fall on the insurance companies.

Admittedly, this distinction can appear minor, especially given that self-insured employers can (and often do) hire insurance companies to design and administrate their self-insured EHBPs.²⁶ In other words, conflicting state regulations of EHBPs would not cause CEOs of large national companies to spend their late-night hours struggling to master the regulatory intricacies of all fifty states. But qualitative distinctions between relative burdens created by regulation must be made in any structure of federal preemption. Otherwise, all state regulation would be preempted, because all state laws can be said to have some attenuated effect on preempted subject matter. For example, without such qualitative distinctions, ERISA presumably would preempt state food handling laws because such regulations affect the available options and costs to employee

24. See, e.g., Margaret G. Farrell, *ERISA Preemption and Regulation of Managed Health Care: The Case for Managed Federalism*, 23 AM. J.L. & MED. 251, 264 (1997); Douglas J. Witten, *Regulation of "Downstream" and Direct Risk Contracting by Health Care Providers: The Quest for Consumer Protection and a Level Playing Field*, 23 AM. J.L. & MED. 449, 466 (1997); cf. *FMC Corp. v. Holliday*, 498 U.S. 52, 65-66 (1990) (Stevens, J., dissenting) (stating that there is no rational reason to permit employee benefit plans to contract for certain rights vis-a-vis employees when state law prohibits similarly situated insurance companies from contracting for the same rights).

25. Of course, many multi-state employers chose to provide different health care plans to their employees in different states notwithstanding the administrative costs of doing so. See, e.g., David Reich-Hale, *Big Employers Self-Funding HMO Costs*, NAT'L UNDERWRITER: LIFE & HEALTH / FIN. SERVICES EDITION, Oct. 11, 1999, at S-21 (describing one large employer that self-insures its employee's medical care in one state but purchases third-party insurance in others).

26. Farrell, *supra* note 5.

benefit plans that wish to provide lunch as a fringe benefit.²⁷

Despite the fact that ERISA's differential treatment of EHBPs and third-party insurance companies that sell health insurance to EHBPs is logically defensible, this differential treatment leads to a troubling inequity for employees. If an EHPB purchases third-party insurance, it is classified as an "insured" plan, and state regulations govern any set of benefits that it purchases. If a plan self-insures, however, these same state regulations do not apply. Consequently, employees in an insured plan benefit from state regulatory protections, whereas similarly-situated employees in a self-insured plan do not.

This apparent inequity is perhaps made more objectionable by the fact that few employees know whether their EHPB is insured or self-insured. "Self-insured" rarely means "self-administrated," as most self-insured plans hire a third-party administrator (TPA) for their EHPB (and, as noted above, TPA services are often provided by insurance companies).²⁸ This means that most employees in self-insured EHBPs submit claim forms to and have their covered medical expenses paid by an entity other than their employer, oblivious to the distinction that the TPA is paying claims with the employer's money rather than with its own.²⁹

The extent of the consequence to employees of whether their benefits plan is insured or self-insured became clear in 1985 (if not before) when the United States Supreme Court decided *Metropolitan Life Ins. Co. v. Massachusetts*.³⁰ *Metropolitan Life* concerned a Massachusetts law requiring group health insurance policies to provide a minimum level of benefits for mental health care. When two insurance companies sold policies to employee benefits plans without such a benefit, the Massachusetts Attorney General brought suit. The United States Supreme Court upheld a Supreme Judicial Court of Massachusetts judgment for the State³¹ on the ground that the Massachusetts mandate was an insurance regulation protected from preemption by the savings clause. In so doing, the Court rejected the insurance companies' argument that the savings clause should

27. See Russell Korobkin, *The Failed Jurisprudence of Managed Care and How To Fix It*, 51 UCLA L. REV. 457, 505 (2003).

28. BARRY F. FURROW ET AL., HEALTH LAW 423 (2d ed. 2000).

29. See, e.g., *Ins. Bd. v. Muir*, 819 F.2d 408, 409 (3d Cir. 1987) (noting that employees received Blue Cross and Blue Shield claim forms and received reimbursement from Blue Cross and Blue Shield but the Blues were providing administrative services for a self-insured plan).

30. 471 U.S. 724 (1985).

31. *Id.* at 734-35.

be read narrowly to protect only “traditional” insurance laws, such as those regulating insurance company reserves, and not “innovative” benefits mandates.³²

The *Metropolitan Life* court noted that, as a result of its ruling, employees in insured plans and employees in self-insured plans would be treated differently under state laws, because the deemer clause would prohibit Massachusetts from applying the mandate to self-insured EHBP.³³ To the extent that it found this distinction problematic, however, the Court laid the blame on Congress’s doorstep for structuring ERISA in the way that it did.³⁴ This dicta was reaffirmed as holding five years later in *FMC Corp. v. Holliday*,³⁵ the only deemer clause case the Supreme Court has ever decided. *FMC Corp.* concerned a self-insured EHBP with a subrogation clause, requiring the plan member to reimburse the plan for any medical care costs that the plan paid if the member recovered those costs in a liability action against a third party.³⁶ A plan member who recovered such expenses from a third-party refused to reimburse the plan on the ground that a state law prohibited subrogation.³⁷ The Court held that ERISA preempted the state law because it was an insurance regulation, and as such “[did] not reach self-funded employee benefits plans because the plans may not be deemed to be insurance companies, other insurers, or engaged in the business of insurance for purposes of such state laws.”³⁸

State mandated benefits laws, like the law at issue in *Metropolitan Life*,³⁹ were enacted as a reaction to the rise of managed care from the 1970s to the 1990s. The speed with which managed care arrangements replaced traditional indemnity insurance as the dominant form of health insurance led to a nation-wide backlash against the perceived aggressiveness of insurer attempts to contain costs by limited benefits and services.⁴⁰ State legislators introduced bills by the hundreds requiring insurers to cover a wide-range of benefits and otherwise mandating the terms of insurance contracts, and nearly every state passed a variety of specific mandates, if not

32. *Id.* at 739-47.

33. *Id.* at 747.

34. *Id.* at 747 & n.25 (stating that the court “merely give[s] life to a distinction created by Congress in the ‘deemer clause’” and citing legislative history).

35. 498 U.S. 52 (1990).

36. *Id.* at 54.

37. *Id.* at 55.

38. *Id.* at 61.

39. *Metropolitan Life*, 471 U.S. at 728.

40. Ochmann, *supra* note 23.

an entire "Patient's Bill of Rights."⁴¹ One researcher estimates that the number of state mandates rose from virtually none in 1970 to 850 in 1991, with the largest rate of increase coming before 1988.⁴²

The explosion of mandated benefits laws protected from preemption by the savings clause means the stakes associated with an EHBP's choice between purchasing third-party insurance or self-insuring its members' medical care costs are high. By self-insuring, an employer can avoid paying the cost of dozens of state insurance mandates, from in vitro fertilization to chiropractic treatment, as well as related state insurance policy regulations, such as the law at issue in *FMC Corp.* prohibiting subrogation by insurance providers. While no one mandate is likely to significantly increase the cost of health insurance, the aggregate affect of mandates can be quite significant.⁴³ These savings, available only to self-insured plans, flow straight to the EHBP's—and thus the employer's—bottom line.⁴⁴ Thus, the deemer clause not only affords different treatment to employers ex post based on the employer's decision to insure or self-insure, it also affects employers' ex ante incentives when making that choice. As a result, self-insurance is attractive not only to the small percentage of employers that operate in multiple jurisdictions⁴⁵ and might wish to minimize the administrative costs of insuring employees subject to inconsistent state rules, but also to any

41. See Russell Korobkin, *The Efficiency of Managed Care "Patient Protection" Laws: Incomplete Contracts, Rounded Rationality, and Market Failure*, 85 CORNELL L. REV. 1, 17-18 (1999).

42. Gail A. Jensen, *State Mandates on Private Insurance*, CATO REGULATION: THE REVIEW OF BUSINESS & GOVERNMENT, Aug. 1, 1992, <http://www.cato.org/pubs/regulation/reg15n4g.html>.

43. See, e.g., Renate M. Nellich, *Executive Partnerships in Reinsurance*, NAT'L UNDERWRITER: LIFE & HEALTH / FIN. SERVICES EDITION, Apr. 20, 1998, at 10 (reporting that benefits expenses among U.S. businesses grew from thirty percent to nearly forty-two percent of payroll between 1975 and 1998, with half the increase due to new mandated benefits).

44. In a perfectly functioning market in which employees had complete information and unlimited cognitive abilities, employers would have no incentive to provide less attractive fringe benefits to their employees than do competitors, because the employer would either have to spend the savings on other forms of employee compensation or risk losing its best employees. It is more plausible to assume, however, that while the availability of health care coverage affects many employees' choice of jobs, few employees consider the details of competing employers' health care plans when making such choices. See generally Korobkin, *supra* note 41.

45. This percentage of employers has been reported to be as low as five percent. See Gail A. Jensen et al., *State Insurance Regulation and Employers' Decisions to Self-Insure*, 62 J. RISK & INS. 185, 200 (1995) (describing the composition of their employer data set as including predominantly single-state employers).

employers that wish to avoid costly regulatory protections that states require insurance companies to provide.⁴⁶

B. The Supreme Court Expands the Reach of the Savings Clause

As the above discussion explains, the broader the interpretation given to ERISA's savings clause, the larger the gap between the legal protections afforded employees enrolled in self-insured and insured health benefits plans, and consequently the greater the incentive of employers to self-insure their medical benefits plans. In its last three terms, the Supreme Court decided two cases specifically involving the breadth of the savings clause in the context of health insurance and another that has indirect implications for the savings clause's importance. The Court resolved these disputes in ways that expand the savings clause's scope and importance. Thus, an indirect effect of the Court's rulings in *Rush Prudential HMO v. Moran*,⁴⁷ *Aetna Health Inc. v. Davila*,⁴⁸ and *Kentucky Ass'n of Health Plans v. Miller*⁴⁹ is to increase the incentive of employers to self-insure.

1. Rush Prudential HMO v. Moran

Rush Prudential HMO concerned one of the 1990s' most popular mandated benefits statutes, the status of which, under the savings clause, was disputed by the lower federal courts.

One of the most controversial features of managed health care is "utilization review," according to which a health insurer reviews treatments proposed by physicians to determine whether they are "medically necessary."⁵⁰ If the insurer's representative determines that a procedure does not satisfy the insurer's standard of medical necessity, the insurer refuses to authorize payment for it. Although the patient may pay for the treatment out of pocket, the costs of medical procedures that are expensive enough to justify utilization review are prohibitive for most patients, so a utilization review denial usually means that the patient will not receive his desired treatment.

46. See, e.g., Peter Schmidt, *Part I: The Basics of ERISA as It Relates to Health Plans*, in EBRI ISSUE BRIEF NO. 167 (SPECIAL REPORT SR-31), at 5 (1995) (reporting that the growing ranks of self-insured plans are "influenced" by employer desire to escape expanding state regulations).

47. 536 U.S. 355 (2002).

48. 124 S. Ct. 2488 (2004).

49. 123 S. Ct. 1471 (2003).

50. See, e.g., Korobkin, *supra* note 27, at 463.

One contributing factor to the public backlash against the health insurance industry in the 1990s was the perception that insurers were using utilization review as a method of minimizing costs by denying legitimate treatment requests.⁵¹ As a result, forty-one states enacted “external review” statutes,⁵² which require health insurers to permit patients to appeal adverse utilization review decisions to a neutral arbitrator and to pay for the treatment if that arbitrator determines that the treatment is medically necessary. Insurers challenged these regulations as preempted by ERISA, and a “circuit split” resulted. The Seventh Circuit held that the savings clause protected an Illinois external review statute.⁵³ Meanwhile, the Fifth Circuit ruled that ERISA preempted a substantively identical Texas statute because the remedies provided under the statute conflicted with ERISA’s remedy provisions.⁵⁴ The Fifth Circuit’s rule would deny the protection of state external review laws to any patient who receives his health insurance through an EHBP, thus treating members of insured and self-insured plans identically. The Seventh Circuit’s rule would grant the same treatment to external review statutes as to the mental health benefits mandate at issue in *Metropolitan Life*, consequently providing rights to employees in insured plans but not those in self-insured plans.

Over a sharp dissent by four justices,⁵⁵ the Supreme Court in 2002 upheld the Seventh Circuit’s position that ERISA’s savings clause protects state external review statutes and that such statutes are not otherwise preempted because they conflict with ERISA’s remedy provisions.⁵⁶ The decision was a major victory for supporters of managed care regulation. It also expanded the legal and practical significance of an employer’s decision to self-insure its EHBP rather than purchase third-party insurance.

51. For an analysis of why it might make business sense for health insurers to engage in such a strategy notwithstanding built-in market constraints on strategic underperformance of contractual obligations, see Korobkin, *supra* note 41, at 29-44.

52. Mark C. Nielsen, *Piercing the Preemptive Veil: Rush Prudential HMO v. Moran Opens the Door for Additional State Regulation of Managed Care Organizations*, 14 HEALTH L. 15, 15 (2002) (noting that external review statutes have been enacted by forty-one states and the District of Columbia).

53. *Moran v. Rush Prudential HMO, Inc.*, 230 F.3d 959 (7th Cir. 2000).

54. *Corporate Health Ins., Inc. v. Texas Dep’t of Ins.*, 215 F.3d 526 (5th Cir. 2000), *modified and reinstated by* 314 F.3d 784.

55. *Rush Prudential*, 536 U.S. at 388 (Thomas, J., dissenting). Chief Justice Rehnquist and Justices Scalia and Kennedy joined Justice Thomas’s dissent.

56. *Id.* at 385, 386.

2. *Aetna Health Inc. v. Davila*

The Supreme Court's decision in *Rush Prudential HMO* permits states to provide patients enrolled in insured (but not self-insured) EHBPs with procedural protections from erroneous utilization review denials, which has the indirect effect of providing EHBPs with a financial incentive to self-insure. The Court's June 2004 decision in *Aetna Health Inc. v. Davila*⁵⁷ further increased the incentive to self-insure by increasing the relative importance to patients of the external review statutes permitted under *Rush Prudential HMO*.

In theory, the legal system can adopt either (or both) of two approaches to prevent health insurance providers from minimizing costs by using the utilization review process to avoid providing services that satisfy the underlying medical necessity standard. One approach to this moral hazard problem relies on ex ante government regulation of services as a prophylactic device. In the case of utilization review, external review statutes serve this function. Providers are prevented by the external review process from refusing to provide at the time of sickness the level of care promised at the time of enrollment. The alternative approach relies on the threat of private litigation and resulting sanctions to deter careless or strategic behavior. Knowing that they can be sued by the patient for resulting damages should they improperly deny coverage of a requested treatment, providers will have an incentive to take appropriate care to furnish the services to which patients are entitled. They will also have an incentive in close cases to err on the side of providing questionable treatment to avoid the risk of litigation.

ERISA's remedy provisions permit a member of an EHP to bring a lawsuit under ERISA to "recover benefits due to him under the terms of his plan," to "enforce his rights under the terms of the plan," or to "clarify his rights to future benefits under the terms of the plan."⁵⁸ However, the Supreme Court has interpreted this portion of the statute narrowly, ruling that aggrieved plan participants can bring suit under the statute for the value of benefits improperly withheld but not for compensatory or punitive damages.⁵⁹ The significance of this limitation for the utilization review process depends on whether a patient improperly denied medically necessary medical treatment may bring suit under specific state statutes or general state tort law that permits a broader range of remedies.

57. 124 S. Ct. 2488 (2004).

58. 29 U.S.C. § 1132(a)(1)(B) (2000).

59. *Mass. Mutual Life Ins. Co. v. Russell*, 473 U.S. 134, 144 (1985).

Prior to 2000, most lower courts to address this question had held that ERISA's remedial provisions preempted all related state claims and, thus, an insurer's legal risk of a utilization review denial was limited to the cost of the desired treatment.⁶⁰ Under this rule, the direct financial incentive to conduct the utilization review process carefully and generously is limited, which increases the attractiveness of ex ante prophylactic regulation embodied in external review statutes to critics of managed care.

The Supreme Court's *Pegram v. Herdrich*⁶¹ decision in 2000 cast doubt on the conventional wisdom that ERISA preempts state law claims against health insurers arising out of utilization review denials. At issue in the case was whether an HMO violates its fiduciary duties under ERISA by basing physician compensation in part on how successful physicians are at limiting resource usage for patient care.⁶² In answering this question in the negative, the Court explained that allowing a patient who is denied medically necessary care to maintain a cause of action for breach of fiduciary duty would essentially duplicate her existing right to challenge medical necessity determinations under state law.⁶³ The Pennsylvania Supreme Court⁶⁴ and three federal circuit courts,⁶⁵ along with commentators,⁶⁶ read the *Pegram* dicta as signaling that ERISA does not preempt state law causes of action arising from utilization review denials.

In last term's decision in *Davila*, the Supreme Court reviewed and reversed the Fifth Circuit's decision to this effect, validating the pre-*Pegram* conventional wisdom that, when a patient receives health care coverage through an EHP, his ability to sue over a utilization review denial can be brought only under ERISA.⁶⁷ The Court made no distinction in its opinion between the rights of employees in insured and self-insured EHPs, despite the fact that it appears that one of the two plaintiffs in the case was

60. The leading case was *Corcoran v. United Health Care, Inc.*, 965 F.2d 1321 (5th Cir. 1992). For a description of the state of the law prior to 2000, see Korobkin, *supra* note 27, at 494-97.

61. 530 U.S. 211 (2000).

62. *Id.* at 217.

63. *Id.* at 235.

64. *Papps v. Asbel*, 768 A.2d 1089 (Pa. 2001).

65. *Cicio v. Does*, 321 F.3d 83 (2d Cir. 2003); *Land v. Cigna Healthcare of Fla.*, 339 F.3d 1286 (11th Cir. 2003); *Roark v. Humana, Inc.*, 307 F.3d 298 (5th Cir. 2002). One circuit disagreed with this reading of *Pegram*, *DiFelice v. Aetna U.S. Healthcare*, 346 F.3d 442 (3d Cir. 2003), as did the dissenters in *Rush Prudential*, 536 U.S. at 388 (Thomas, J., dissenting).

66. See, e.g., Korobkin, *supra* note 27, at 520; Russell Korobkin, *HMOs Get Authority To Strike a Fair Balance*, L.A. TIMES, June 14, 2000, at B9.

67. 124 S. Ct. at 2493.

enrolled in an insured EHBP while the other was a member of a self-insured EHBP.⁶⁸

In a concurring opinion, Justice Ginsburg (joined by Justice Breyer) argued that the Court should consider revisiting its earlier decision that read ERISA's remedial provisions so narrowly or, alternatively, that Congress should rewrite the statute.⁶⁹ Unless and until this happens, however, it now seems clear that state statutes requiring external review provide the primary, if not the only, legal check on health insurance providers using utilization review to minimize the cost of providing health care, thus increasing the differential flexibility that self-insured EHBPs not subject to external review laws enjoy relative to insured EHBPs. In other words, *Davila* amplifies the difference in legal treatment of self-insured and insured EHBPs established in *Rush Prudential*—increasing the desirability of self-insurance to employers who want to maintain maximum flexibility to reduce the costs of providing health care coverage and decreasing the desirability of self-insurance to state regulators who wish to maximize the extent of legal protection for employees.

3. *Kentucky Ass'n of Health Plans v. Miller*

In 2003, the Supreme Court followed its decision in *Rush Prudential* with another decision specifically concerning the breadth of the savings clause. As it did in *Rush Prudential*, the Court in *Kentucky Ass'n* again

68. The Court's opinion came in the consolidated cases of *Aetna Health Inc. v. Davila* and *Cigna Healthcare of Tex., Inc. v. Calad*. *Davila*, 124 S. Ct. at 2492-93. According to Cigna Healthcare's brief, Ruby Calad's EHBP was self-insured by her husband's employer, which in turn "delegated certain administrative responsibilities for the plan to petitioner Cigna Healthcare of Texas, Inc." Brief for Petitioner Cigna Healthcare of Texas, Inc. at 2, *Davila* (No. 02-1845). According to the joint brief of respondents *Davila* and *Calad*, Juan *Davila* "was a member of Aetna's HMO, which is not itself an 'ERISA plan.'" Brief for Respondents at 6, *Davila* (No. 02-1845). This language implies that Aetna, not *Davila*'s employer, was the risk bearing entity. The Supreme Court's opinion does not state whether the employees were in insured or self-insured EHBPs—in fact, the difference between insured and self-insured plans is mentioned nowhere in the opinion. The Court does state that the employees' "respective plan sponsors had entered into agreements with [Aetna and Cigna] to administer the plans." *Davila*, 124 S. Ct. at 2493 (emphasis added). This language could be read to imply that both insurance companies *only* administered the plans (and thus neither was the actual risk-bearing entity), but the better reading is probably that the Court did not think that whether risk was borne by the employees' employer or the insurance company administrator was relevant to the question of the preemptive effect of ERISA's remedial provisions.

69. *Davila*, 124 S. Ct. at 2503-04 (Ginsburg, J., concurring).

avored state regulatory power over broad federal preemption under ERISA.

At issue in *Kentucky Ass'n* were state “any willing provider” (AWP) laws, which require health insurers doing business in the state to contract with all physicians (or, in some cases other medical care providers such as chiropractors or pharmacists) willing to provide care to an insurer’s customers in accordance with the insurer’s standard terms and conditions.⁷⁰ At least half of the states have enacted some version of an AWP statute.⁷¹

Most managed care organizations (MCOs) oppose AWP laws on the ground that they take away a potent tool for containing health care costs. By selectively contracting only with certain providers, MCOs can force price concessions from those providers, both because they can guarantee a large quantity of business to the selected providers, and because the providers must worry that the MCO will refuse to contract with them at all if they do not grant such concessions.⁷² Patients’ advocates, on the other hand, often support the laws on the grounds that they give patients greater treatment options and that they allow patients who move from one insurance plan to another the ability to maintain their pre-established doctor-patient relationships.⁷³

As was the case with external review statutes, the circuit courts agreed that AWP laws “relate to” ERISA plans and are subject to federal preemption but split on the question of whether they are protected by the savings clause.⁷⁴ The Fourth and Sixth Circuits determined that AWP statutes qualify as insurance regulations, and thus are saved;⁷⁵ the Fifth and Eighth Circuits held that the laws fall outside the protection of the savings clause because they regulate entities outside the insurance industry or because they do not affect the allocation of risk between insurers and their customers.⁷⁶

70. Vickie Y. Brown & Barbara R. Hartung, *Managed Care at the Crossroads: Can Managed Care Organizations Survive Government Regulation?*, 7 ANNALS HEALTH L. 25, 36 (1998); Butler, *supra* note 11, at 67; Justin Goodyear, *What Is an Employee Benefit Plan?: ERISA Preemption of “Any Willing Provider” Laws After Pegram*, 101 COLUM. L. REV. 1107, 1116 (2001).

71. Farrell, *supra* note 5, at 270.

72. See, e.g., Korobkin, *supra* note 27, at 510.

73. *Id.* at 509-10.

74. *Id.* at 511-12.

75. *Kentucky Ass’n of Health Plans v. Nichols*, 227 F.3d 352 (6th Cir. 2000); *Stuart Circle Hosp. Corp. v. Aetna Health Mgmt.*, 995 F.2d 500 (4th Cir. 1993); see also Korobkin, *supra* note 24, at 512.

76. *Prudential Ins. Co. of Am. v. Nat’l Park Med. Ctr.*, 154 F.3d 812 (8th Cir. 1998);

The Supreme Court, this time in a unanimous decision, followed the circuit courts that gave a broader reading to the savings clause and held that ERISA does not preempt the Kentucky statutes. In so doing, the Court renounced the complicated, multi-part test for determining whether a state law “regulates insurance” that it introduced in *Metropolitan Life* and replaced it with a simpler, easier-to-satisfy test. Under the rule enunciated in *Kentucky Ass’n*, in order to qualify as a law that regulates insurance and therefore receives protection from the savings clause, the state law in question need only be “specifically directed” at the insurance industry (as opposed to being a law of general applicability) and “substantially affect” an insurer’s insurance practices (as opposed to being a law that affects insurance companies only in their non-insurance-related capacities).⁷⁷

4. *The Implications of the Court’s Savings Clause Jurisprudence*

Since ERISA was enacted more than a quarter-century ago, the Supreme Court’s jurisprudence has tilted, on balance, in favor of preemption.⁷⁸ Going back as far as 1983, the Court has read the preemption clause broadly, finding that ERISA preempts a wide range of state laws because they either have a “reference to” or have a “connection with” EHBPs.⁷⁹ In its 1995 decision in *New York State Conference of Blue Cross and Blue Shield Plans v. Travelers Insurance Co.*,⁸⁰ the Court narrowed the scope of ERISA preemption somewhat, but also suggested that the scope of preemption would continue to be broad. The Court’s decisions in *Rush Prudential HMO* and *Kentucky Ass’n* reinforce this reading of *Travelers*, because the Court declined to address the underlying assumption of its holdings that the state laws at issue did in fact “relate to” ERISA plans, and thus were the subject of preemption.

Texas Pharmacy Ass’n v. Prudential Ins. Co. of Am., 105 F.3d 1035 (5th Cir. 1997); see also Korobkin, *supra* note 24, at 512.

77. *Kentucky Ass’n of Health Plans v. Miller*, 123 S. Ct. 1471, 1479 (2003).

78. The Court has drawn its share of criticism for this. See, e.g., Farrell, *supra* note 5; Catherine L. Fisk, *The Last Article About the Language of ERISA Preemption? A Case Study of the Failure of Textualism*, 33 HARV. J. ON LEGIS. 35 (1996); Korobkin, *supra* note 27; Saya, *supra* note 5, at 160; Susan J. Stabile, *Preemption of State Law by Federal Law: A Task for Congress or the Courts?*, 40 VILL. L. REV. 1, 31-35 (1995); Deborah J. Massaro, Comment, *Removal of the ERISA Preemption Shield: Will the Third Circuit’s Approach Make a Difference?* – In re U.S. Healthcare, Inc., 26 DEL. J. CORP. L. 585, 592-93 (2001); Nicole Weisenborn, Note, *ERISA Preemption and Its Effect on State Health Reform*, 5 KAN. J.L. & PUB. POL’Y 147 (1995).

79. *Shaw v. Delta Air Lines, Inc.*, 463 U.S. 85, 96-97 (1983).

80. 514 U.S. 645 (1995).

But while the Court continues to read the preemption clause broadly, it also continues to read the savings clause broadly. As a general statement, it is fair to say that the Court has promoted federal authority through the preemption clause, while simultaneously protecting state authority through the savings clause. An unintended consequence of this doctrinal approach is that it maximizes the gap in treatment that employees in self-insured and insured plans receive under ERISA.

II. EXPLOITING THE DEEMER CLAUSE LOOPHOLE

Part I described how a broad interpretation of ERISA's savings clause juxtaposed with ERISA's deemer clause creates a significant incentive for EHBPs to self-insure their members' health care costs rather than purchase third-party health insurance. No matter how great this incentive, however, EHBPs will not choose to self-insure if they cannot afford to assume the risk of catastrophic medical care claims in a given year. Stop-loss insurance protects EHBPs from catastrophic losses, thus making self-insurance feasible for even small employers and thereby facilitating widespread avoidance of state insurance regulations.

By using stop-loss insurance to minimize insurance risk while simultaneously avoiding state regulation, EHBPs exploit a loophole in ERISA's statutory structure. To the chagrin of supporters of greater regulation of health insurance, however, this loophole is consistent with the plain language of ERISA, and it is not inconsistent with ERISA's structure. Courts thus have properly refused to heed the calls of regulation supporters to close the loophole, although their analyses often make the issue much more complicated than it should be.

A. *The Economics of Self-Insurance*

A managed care organization or a traditional indemnity insurance company that sells third-party health insurance provides two distinct services. First, it administers the insurance plan, which includes establishing contracts with medical care providers and reviewing and paying covered claims. Second, it assumes the risk that in any given year its customers will incur medical care costs that are higher than their actuarial, average expected cost. This latter service is often described as bearing "insurance risk."⁸¹ Insurance companies, of course, do not provide such services for free. The premiums they charge can be understood as consisting of the customer's expected medical care cost, plus an extra

81. Butler, *supra* note 11, at 62.

amount to cover the costs of administration and insurance risk, including the company's profit margin.⁸²

By self-insuring rather than purchasing insurance, employers, in theory, can avoid the costs of paying a third-party insurer to provide administrative services. In reality, many employers contract with third parties to serve as TPAs and administer their health care benefits;⁸³ this is true even of extremely large companies with tens of thousands of employees.⁸⁴ Presumably, this is because TPAs' expertise in administration makes it cheaper for self-insured EHBPs to contract for administrative services rather than to provide them "in-house." This suggests that EHBPs are unlikely to save substantially, if at all, on administrative costs by self-insuring.⁸⁵

Whether it is in an EHP's interest to self-insure, then, depends on the extent of the insurance risk that it would undertake. An EHP with few beneficiaries can expect a large variance in annual medical care costs.⁸⁶ The costs incurred by a single member who suffers a catastrophic illness could be far greater than a plan's actuarially expected medical costs, resulting in severe cash-flow problems or even insolvency. The risks associated with extreme annual fluctuations decline as the size of the EHP increases; that is, the more members in an EHP, the lower the expected annual variance of the plan's expenses.⁸⁷ Insurance companies have an advantage relative to individual employers in managing insurance risk because they pool the individual risks of a large number of customers.⁸⁸ Extremely large employers, however, have a pool of individual

82. Laurence Baker, *Managed Care and Social Welfare: What Has Managed Care Really Done to the U.S. Health Care System?*, in *THE POLITICAL ECONOMY OF HEALTH CARE REFORMS* 35-39 (Huizhong Zhou ed., 2001).

83. Butler, *supra* note 11, at 90.

84. *Id.*

85. *Cf.* Jensen et al., *supra* note 45, at 187 (finding that "research suggests that . . . administrative costs for self-insured plans are actually higher than those of purchased plans containing the same coverage").

86. Risk in insurance pools is $1/N$ times the variance of each individual (N is the number of members in the pool). The larger the pool and the more diverse the population, the lower the variance in risk. CHARLES E. PHELPS, *HEALTH ECONOMICS* 331 (2003).

87. 1 ERIC MILLS HOLMES & MARK S. RHODES, *HOLMES' APPLEMAN ON INSURANCE*, 2D § 1.2 (1996) [hereinafter *APPLEMAN*]; ROBERT E. KEETON & ALAN I. WIDISS, *INSURANCE LAW: A GUIDE TO FUNDAMENTAL PRINCIPLES, LEGAL DOCTRINES, AND COMMERCIAL PRACTICES* 13 (Student ed. 1988).

88. *APPLEMAN*, *supra* note 87, § 2.18 (commenting that generally, only large corporations fully self-insure).

risks sufficiently large to minimize its insurance risk, reducing the value of purchasing third-party insurance.⁸⁹ The consequence is that the EHBPs of only very large employers should routinely self-insure, while most other EHBPs should be willing to pay an insurance company a premium for the service of bearing their insurance risk.⁹⁰

By allowing self-insured EHBPs to avoid state regulation, however, ERISA provides EHBPs three additional incentives to self-insure. First, a self-insured EHBP with members in multiple states can provide a single set of benefits for all of its employees and avoid the cost of conforming to conflicting regulations and mandates in different jurisdictions. Second, that set of benefits can be more limited, and thus cheaper to provide, than the set of benefits the EHBP would have to provide should it purchase third-party insurance encumbered by state mandates. Third, by self-insuring, EHBPs can avoid premium taxes on health insurance purchases imposed by most states (usually for the purpose of subsidizing state insurance pools to cover the uninsured or difficult-to-insure),⁹¹ as well as other state regulatory requirements that can be costly, inconvenient, or both, such as regulations concerning what information insurers must be provided to consumers.⁹² While minimizing administrative burdens of multi-state employers is a goal of ERISA, there is no indication that ERISA's drafters affirmatively desired to protect EHBPs from all costs associated with state insurance regulation. If that were Congress's goal, the statute presumably would not have included the savings clause.

Measuring the absolute popularity of self-insurance among employers at any given time is a notoriously inexact science, because understandings

89. See, e.g., Butler, *supra* note 11, at 62.

90. An exception to this rule might be some employers with very young workforces in states that require certain types of insurers (often HMOs) to sell coverage at "community rates"—that is rates that do not discriminate based on the demographics or claims experience of particular employer groups. Cf. David Reich-Hale, *Big Employers Self-Funding HMO Costs*, NAT'L UNDERWRITER: LIFE & HEALTH / FIN. SERVICES EDITION, Oct. 11, 1999, at S-21 (reporting that in 1998 sixty-three percent of employees in HMOs were in community-based plans). These employers might find self-insurance a particularly desirable arrangement because they can avoid paying insurance company rates that substantially exceed their expected claims experience. See, e.g., Michael Prince, *Self-Funded Health Plans Not Expanding Ranks*, BUS. INS., Feb. 21, 2000, at 3 ("Employers in community-rated HMOs can generally save money by going into a self-funded HMO if their claims experience is better than that of the overall group insured by the HMO.").

91. See, e.g., Prince, *supra* note 90, at 3.

92. See, e.g., Karl Polzer & Patricia A. Butler, *Employee Health Plan Protections Under ERISA*, 16 HEALTH AFF. 93, 94-95 (1997).

of what constitutes self-insurance vary among employers and because employers often offer employees a choice of plans,⁹³ some of which might be insured and others self-insured. Assessing the impact of ERISA's favorable regulatory treatment of self-insured EHBPs on employers' decisions about whether to self-insure is even more problematic because many exogenous factors can affect the relative benefits of self-insuring versus insuring.⁹⁴ Notwithstanding these notes of qualification, however, there is no doubt that self-insurance has become more popular among employers, by many fold, over the last three decades, and that the desire to avoid state benefits mandates and premium taxes can explain at least some of this increase in popularity.

Employers shifted from insured to self-insured EHBPs in large numbers in the 1980s. According to one study, only four percent of employee health benefits were paid for by self-insured plans at the time of ERISA's enactment, while forty-seven percent of EHBPs self-insured at least their primary health benefits plan in 1986.⁹⁵ According to other studies, that number rose to sixty-seven percent in 1992,⁹⁶ and forty-six percent of all employees who received health coverage as an employment benefit were enrolled in self-insured EHBPs by that year.⁹⁷

93. A recent Kaiser Family Foundation study reported that nearly half of all employees covered by employment-based health plans had a choice of three or more different plans, while only thirty-eight percent were offered only one plan. THE KAISER FAMILY FOUND. & HEALTH RESEARCH & EDUC. TRUST, EMPLOYER HEALTH BENEFITS: 2003 ANNUAL SURVEY 64 (2003) [hereinafter KFF/HRET].

94. For example, high interest rates give employers an incentive to self-insure, because self-insuring allows them to keep cash until an employee needs care rather than paying a premium to the insurer at the beginning of the year. See, e.g., Daniel M. Fox & Daniel C. Shaffer, *Health Policy and ERISA: Interest Groups and Semipreemption*, 14 J. HEALTH POL. POL'Y & L. 239, 252 (1989). Self-insurance also enables employers to collect detailed claims data in order to try to manage employee benefit costs that insurance companies often will not provide because they fear that employers with favorable claims experience might shop for cheaper insurance or decide to self-insure. Michael Prince, *Health Plans Shifting Approach as Costs Climb*, BUS. INS., Feb. 25, 2002, at 16.

95. Steve Kalmeyer, *ERISA and State Health Reform*, HEALTH POL'Y MONITOR, Spring 1997, at 1.

96. A. FOSTER HIGGINS & CO., FOSTER HIGGINS HEALTH CARE BENEFITS SURVEY 19 (1992), cited in Jeffrey Lenhart, *ERISA Preemption: The Effect of Stop-Loss Insurance on Self-Insured Health Plans*, 14 VA. TAX REV. 615, 615 n.1 (1995).

97. BUREAU OF LABOR STATISTICS, U.S. DEP'T OF LABOR, EMPLOYEE BENEFITS IN MEDIUM AND LARGE PRIVATE ESTABLISHMENTS (1993); see also Gregory Acs et al., *Self-Insured Employer Health Plans: Prevalence, Profile, Provisions, and Premiums*, 15 HEALTH AFF. 266-78 (1996) (estimating that forty percent of private sector employees and their dependents were

The strong correlation between the rise of managed care financing arrangements and the corresponding explosion of state insurance regulation on the one hand and the steep rise in employer self-insurance on the other does not, of course, itself prove that the former caused the latter. Many self-insured employers offer a rich set of benefits, suggesting that many factors affect an employer's decision to self-insure, not just the desire to save money by offering fewer benefits. Still, it seems clear that at least some and probably much of the increase in self-insurance can be attributed to the desire to use ERISA to avoid state regulations of one type or another.⁹⁸ As an illustration, one study attempting to explain the causes of employer shifts to self-insurance in the 1980s found that the desire to avoid the costs of state insurance mandates and premium taxes explained about two-thirds of the increase in employer self-insurance observed in the early part of that decade.⁹⁹

Although self-insurance rates have fluctuated in the last decade,¹⁰⁰ at least half of workers with employment-based health care benefits are probably in self-insured plans today. One recent study conducted by the Kaiser Family Foundation (KFF) reports that, in 2003, fifty-two percent of workers with employment-based health care benefits were in self-insured plans, down slightly from the fifty-six percent figure reported by a KPMG study in 1996.¹⁰¹ By further increasing the differential susceptibility to state regulation of insured and self-insured EHBPs, the Supreme Court's expansion of ERISA's savings clause over the last three years suggests that this percentage is likely to increase in coming years.

enrolled in self-insured plans).

98. Cf. Donald T. Bogan, *Protecting Patient Rights Despite ERISA: Will the Supreme Court Allow States to Regulate Managed Care?*, 74 TUL. L. REV. 951, 1005 (2000) (claiming a causal relationship between state regulation and the increasing number of self-insured EHBPs).

99. Jensen et al., *supra* note 45, at 208. In the authors' sample, state regulation could not explain further increases in self-insurance later in the decade, which does suggest other factors are also at play. The authors hypothesize that the different results for the different time periods analyzed might be the result of greater state regulatory activity early in the decade or employers most concerned with avoiding the costs of state regulation having already converted to self-insurance early in the decade. *Id.* at 210-11.

100. Relatively small fluctuations in the rate of self-insurance can be due to factors unrelated to the different regulatory treatment of insured and self-insured EHBPs, such as the extent of price competition in the insurance market and changes in the popularity of different types of managed care, some of which are easier to finance through self-insurance than others. For a good discussion, see Jon R. Gabel et al., *Self Insurance in Times of Growing and Retreating Managed Care*, 22 HEALTH AFF. 202 (2003).

101. KFF/HRET, *supra* note 93, at 125 exhibit 10.1.

B. Stop-Loss Insurance

The regulatory benefits of self-insurance created by ERISA help to explain why many EHBP's would like to self-insure, but the insurance risk traditionally associated with self-insurance should remain a major disincentive for all but the largest employers. What is most astonishing about the extent to which employers self-insure is that, although very large employers remain the most likely to self-insure their EHBP's, even small to mid-sized employers self-insure in significant numbers. According to the General Accounting Office, in 1992, thirty-two percent of employees working for companies with more than 100 employees were covered by a self-insured plan.¹⁰² KFF reports that in 2003 only ten percent of covered employees in firms with fewer than 200 workers were in self-insured plans, down from the twenty-four percent reported by KPMG in 1996, but KFF also reports that fully fifty percent of covered employees in mid-sized firms (200-999 employees) currently receive their benefits from self-insured plans.¹⁰³

The surprising popularity of self-insurance among small and mid-sized employers can be attributed significantly, although not entirely, to a product known as "stop-loss" insurance (or, when issued to an EHBP, sometimes known as "medical stop-loss" insurance).¹⁰⁴ Although the details can often be complicated, the basic concept of stop-loss insurance is simple. The EHBP pays for its employees' covered medical care expenses from a trust fund established for that purpose or from current revenues.¹⁰⁵ At the same time, the EHBP purchases third-party stop-loss insurance for itself—not for its members—that covers losses suffered by the plan as the result of members' catastrophic claims against it.¹⁰⁶ The stop-loss insurance

102. U.S. GEN. ACCOUNTING OFFICE, GAO/HEHS-95-167, *EMPLOYER-BASED HEALTH PLANS: ISSUES, TRENDS, AND CHALLENGES POSED BY ERISA* 49 (1995) [hereinafter GAO, *ISSUES AND TRENDS*].

103. KFF/HRET, *supra* note 93, at 125 exhibit 10.1.

104. A large majority of self-insured EHBP's purchase stop-loss coverage, and nearly all small and medium-sized employers that self-insure purchase stop-loss coverage to cap their exposure. A. FOSTER HIGGINS & CO., *supra* note 96, at 19 (reporting sixty-four percent of self-insured employers with more than one thousand employees purchased stop-loss coverage, and ninety-six percent of self-insured employers with fewer than one thousand employees purchased stop-loss coverage).

105. Paredes, *supra* note 6, at 249.

106. *Id.*; see also Deborah Shalowitz Cowans, *Employers Have Various Options in Covering Catastrophic Care*, BUS. INS., Aug. 2, 1999, at 3 ("For the most part, . . . self-insured employers . . . rely on stop-loss insurance to fund and manage catastrophic health care

pays the EHBP when the plan's losses in a given year exceed a predetermined amount, known as the "attachment point."

Stop-loss policies can have either "specific" attachment points, "aggregate" attachment points, or both. If the policy has a specific attachment point, usually a dollar amount, the stop-loss insurance reimburses the EHBP for any individual employee's medical costs in excess of the attachment point. If the policy has an aggregate attachment point, usually expressed as a percentage of the EHBP's actuarially determined expected annual cost, the insurance is tapped if the EHBP pays out more than that amount for total covered member medical care costs.¹⁰⁷

Stop-loss coverage as a risk management tool for EHBPs is not a new innovation. Just as traditional insurance providers usually "reinsure" part of their insurance risk in order to minimize exposure to catastrophes, even large EHBPs with an actuarially sound risk pool purchase stop-loss insurance just in case an unexpected scourge has a disproportionately catastrophic effect on its members.¹⁰⁸ An innovation that made possible the vast expansion of self-insured EHBPs in the 1980s and 1990s, especially among smaller employers,¹⁰⁹ was the sale of stop-loss policies with such low attachment points—some as low as \$500¹¹⁰—that the EHBP maintained little insurance risk or none at all.¹¹¹ For example, in terms of the insurance

claims.”).

107. Paredes, *supra* note 6, at 249.

108. One mid-1990s study found that sixty-one percent of all large employers self-fund with stop-loss insurance, while only thirteen percent self-fund without stop-loss insurance. Ken McDonnell, *Questions and Answers on Health Insurance Benefit Issues*, in EBRI ISSUE BRIEF No. 164, at 12 (1995).

109. *Cf.* Jerry Geisel, *ERISA Showdown Likely over State Stop-Loss Bill*, BUS. INS., May 3, 1999, at 1 (reporting that small employers favor stop-loss policies with low attachment points, whereas employers with more than 500 employees usually purchase policies with higher attachment points).

110. *See Key Patients' Protections: Lessons from the Field: Hearing Before the Senate Comm. on Health, Education, Labor and Pensions of the United States*, 106th Cong. 11 n.9 (1999) (statement of Kathleen Sebelius, Kansas Insurance Commissioner) (observing that small employers will often “self-fund only a very small dollar amount (\$500) [of their employees' health benefits] and then buy stop-loss insurance for the rest of their liability”).

111. *See, e.g., Md. Bars 'Stop-Loss' Policies*, INS. ACCOUNTING, Nov. 1, 1999, at 1 (quoting Maryland Insurance Commissioner describing the terms of one insurer's stop-loss policy). Anecdotal evidence indicates that stop-loss policies with extremely low attachment points became prevalent in the mid-1990s. *See* Polzer & Butler, *supra* note 92, at 98 (noting that many state insurance regulators reported an “increasing number of small businesses are ostensibly self-insuring while also purchasing stop-loss policies covering individual claims exceeding \$500 or \$1,000”).

risk it maintains, there is no difference between an EHBP that purchases third-party insurance policies for each employee with a \$500 annual deductible and an EHBP that offers a self-insured benefit plan with a \$500 deductible and maintains stop-loss insurance with a \$500 specific attachment point per employee—in both examples, the employer retains no insurance risk at all. By hiring a TPA and purchasing stop-loss coverage with low attachment points (sometimes from the same company), self-insured EHBPs can virtually eliminate all of the costs of self-insurance while taking advantage of the beneficial regulatory treatment provided to them by virtue of ERISA.

In light of the Supreme Court's recent expansive interpretations of ERISA's savings clause, the incentives for EHBPs to self-insure and purchase stop-loss insurance rather than purchase third-party health insurance are now more compelling than ever, although unrelated market forces that have caused sharp increases in stop-loss insurance premiums in the last few years could temporarily dampen this incentive.¹¹²

C. Legal Challenges

The late 1980s and early 1990s brought a number of legal challenges to the use of self-insurance arrangements coupled with stop-loss insurance by EHBPs attempting to avoid state insurance regulations. All of these challenges failed, as they should have, although the courts' reasoning was not always as precise as it might have been.

In the typical challenge, a self-insured EHBP that carries stop-loss insurance attempts to enforce a provision of the plan that is contrary to state law, or a plaintiff attempts to enforce a provision of state law that is contrary to the terms of a plan against an EHBP that carries stop-loss insurance. For example, in *United Food & Commercial Workers v. Pacyna*,¹¹³ a state anti-subrogation law prevented insurance companies from recovering benefit payments made to insured members who collected duplicate

112. See, e.g., Karen Cutts, *Using RRGs To Fund Stop-Loss Exposures in Self-Funded Medical Plans*, NAT'L UNDERWRITER: PROP. & CASUALTY / RISK & BENEFITS MGMT. EDITION, Apr. 7, 2003, at 33 (reporting current increases in medical stop-loss insurance premiums of twenty to fifty percent); Michael Prince, *Employers To Feel Bite of Health Reinsurance Rate Hikes*, BUS. INS., Oct. 23, 2000, at 30 (describing a number of trends, in addition to health care inflation, driving up stop-loss insurance rates); Michael Prince, *Reinsurers Shifting More Health Risks to Buyers*, BUS. INS., Oct. 28, 2002, at 10 (reporting that the price of medical stop-loss policies increased twenty to thirty percent in 2002 as a consequence of insurers suffering losses in other product lines and the reduction of overall capacity in the market).

113. 801 F.2d 1157 (9th Cir. 1986).

benefits from another party, such as a tortfeasor.¹¹⁴ When an EHBP sought subrogation according to the plan's coverage terms, the member, relying on the Supreme Court's distinction between self-insured and insured plans drawn in *Metropolitan Life* and *FMC Corp.*, argued that the plan's subrogation clause was unenforceable because the plan's stop-loss coverage rendered it "insured" and thus not shielded from state regulation by the deemer clause.

The Ninth Circuit in *United Food*, like the Fourth¹¹⁵ and Sixth¹¹⁶ Circuits in similar cases, ruled that the deemer clause did in fact protect the EHBP from the state regulation in question, notwithstanding the fact that the EHBP purchased stop-loss insurance.¹¹⁷ The usual justification is that the purchase of stop-loss insurance fails to render an EHBP "insured" for ERISA purposes because the EHBP maintains direct liability to plan members, while the stop-loss insurer is liable only to the plan, not to individual members.¹¹⁸ Therefore, as some courts have explained, an EHBP with stop-loss insurance would be liable to plan members for the cost of their health care even if the stop-loss insurer were to become insolvent, and conversely, members would have no claim against a stop-loss insurer if the employer went bankrupt, while an EHBP that purchases third-party health insurance for its members would have no liability to those members

114. *Id.* at 1159.

115. *Talquin v. Thompson*, 928 F.2d 649 (4th Cir. 1991).

116. *Lincoln Mut. Cas. Co. v. Lectron Prods., Inc.*, 970 F.2d 206 (6th Cir. 1992). *Lincoln Mutual* explicitly overruled an earlier, contrary decision in *Northern Group Services, Inc. v. Auto Owners Ins. Co.*, 833 F.2d 85 (6th Cir. 1987). *Lincoln Mutual*, 970 F.2d at 210 n.3. An earlier Sixth Circuit decision in *Michigan United Food & Commercial Workers Unions v. Baerwaldt*, 767 F.2d 308 (6th Cir. 1984), held that an EHBP that purchased stop-loss coverage must abide by a state insurance regulation, *id.* at 313, and is thus sometimes cited as conflicting authority, *see, e.g., Paredes, supra* note 6, at 256-57. The *Baerwaldt* decision, however, was based on the court's understanding from the plaintiff's complaint that the insurance company "will pay all benefits in excess of claims liability limit under the group policies"—that is, that the stop-loss insurance insured the plan members health care costs directly rather than insuring the plan itself. 767 F.2d at 313. Thus, while the court might have misunderstood the nature of the EHBP's arrangement with the insurance company, based on its understanding that benefits were provided to plan members by an insurance company rather than by the EHBP itself, its holding is not inconsistent with *Lincoln Mutual*, *United Food*, or *Talquin*.

117. 801 F.2d at 1161-62.

118. *See Talquin*, 928 F.2d at 653 (noting that *Talquin's* plan is directly liable to *Talquin's* employees and the stop-loss insurance covers the plan); *United Food*, 801 F.2d at 1161-62 (noting that "no insurance is provided to the participants").

if the insurance company became insolvent.¹¹⁹

Whether or not this distinction is functionally significant, however, it is irrelevant under ERISA. The relevant distinction between insured and self-insured EHBPs for ERISA's purposes is that members of an insured plan have a contract with an insurance company, whereas members of a self-insured plan have a contract only with the plan. ERISA distinguishes between insurance companies, which states may regulate, and EHBPs themselves, which states may not regulate, but the statutory text never makes a distinction between insured and self-insured plans per se. The plain language of ERISA requires courts to ask only whether the terms of a plan member's health insurance contract are provided by an EHP or by a third-party insurance company—the deemer clause, recall, states only that an employee benefit plan may not be regulated like an insurance company.¹²⁰ If an insurance company covers the member, state law may override terms of the insurance contract and substitute different or additional terms. If the EHP itself covers the member, state law governing insurance companies may not override the terms of the insurance contract because states may not regulate EHBPs as insurance companies. So, for example, if an insurance company issues a policy to an EHP member that excludes coverage for mental health care, a state mental health care mandate can nullify that exclusion and effectively rewrite the contract between the individual and the insurance company. On the other hand, if an EHP enters into an equivalent contract with a plan member, the deemer clause prohibits the same state mandate from overriding the exclusion, even though the EHP behaves functionally like an insurance company vis-à-vis its members.

In *Metropolitan Life*, the Supreme Court observed that a consequence of ERISA, in the context of the issues raised in that case, is that insured and self-insured plans receive different treatment. This observation is correct. Unfortunately, lower courts have sometimes misinterpreted this observation as being equivalent to a statement that the distinction between insured and self-insured plans is itself doctrinally relevant, which it is not—these terms of art never appear in ERISA's text. This error has led to a serious analytical tangle, as courts struggle to determine whether EHBPs that directly insure their members' health care costs and purchase stop-loss insurance for themselves are functionally "insured" or "self-insured." Such

119. See *Am. Med. Sec., Inc. v. Bartlett*, 111 F.3d 358, 364 (4th Cir. 1997); *Georgetown Univ. Hosp. v. Reliance Standard Life Ins. Co.*, No. 97-1912, 1999 U.S. App. Lexis 7803, at *5-*6 (4th Cir. 1999).

120. 29 U.S.C. § 1144(b)(2)(B) (2000).

EHBPs appear to be self-insured, in the sense that they bear the insurance risk of their members' illnesses and injuries rather than paying a third-party to bear that risk. On the other hand, they appear to be insured from a functional perspective, in the sense that they purchase insurance coverage to protect themselves from losses.

Some courts have responded to the confusion over the terms "insured" and "self-insured" by attempting to determine whether an EHBP is predominantly insured or self-insured. For example, in *Brown v. Granatelli*,¹²¹ the Fifth Circuit found that an EHBP with high-attachment point stop-loss coverage was self-insured and protected from state law by ERISA's deemer clause.¹²² It suggested in dicta, however, that an EHBP with low-attachment point stop-loss insurance likely would be considered an "insured" plan subject to state regulation.¹²³ Other courts have followed the Fifth Circuit in suggesting that whether an EHBP is subject to state regulation might depend on the specific level of its stop-loss insurance's attachment point.¹²⁴

This type of analysis is fundamentally misguided, because whether an EHBP maintains the actual insurance risk associated with employee illness bears no direct relevance to the question of whether the deemer clause, according to its text, prohibits state regulation of its members' health insurance contracts. Courts need only ask which entity promises to pay the health care costs incurred by plan members. If the EHBP must pay these costs, and thus acts as an insurer of its employee's health care, the state may not regulate the provisions of the employee-EHBP contract, and the plan is therefore "self-insured" according to the *Metropolitan Life* dichotomy. If a third-party insurance company bears the insurance risk of the employee's health care, the state may regulate the insurance contract, and the plan is therefore "insured" under *Metropolitan Life*. Whether a self-insured plan does or does not purchase stop-loss insurance, or whether that stop-loss insurance has a low or high attachment point, is simply irrelevant, at least under a close reading of ERISA's text.

121. 897 F.2d 1351 (5th Cir. 1990).

122. *Id.* at 1355.

123. *Id.*

124. See *Bricklayers Local No. 1 Welfare Fund v. La. Health Ins. Ass'n*, 771 F. Supp. 771, 774 (E.D. La. 1991); *Thompson v. Talquin Bldg. Prods. Co.*, C.A. No. 89-0082-H, 1990 U.S. Dist. LEXIS 11380 (W.D. Va. 1990); *Associated Indus. of Missouri v. Angoff*, 937 S.W.2d 277, 283 (Mo. Ct. App. 1996).

D. Should Courts Close the Loophole?

Many observers have criticized the differential treatment that self-insured and insured EHBP receive vis-à-vis state law.¹²⁵ Following the Supreme Court's lead in *Metropolitan Life*,¹²⁶ however, most attribute responsibility for the distinction to Congress, which drafted ERISA, rather than to the courts, whose job is only to interpret the statute.¹²⁷ Wise or not as a matter of policy, it would be impossible for courts to eliminate this distinction entirely without reading the deemer clause out of the statute—a result that would be inconsistent with any mainstream view of proper statutory interpretation.

Employer attempts to evade both state regulation and insurance risk by purchasing stop-loss insurance with low attachment points has elicited a somewhat different reaction; some courts and commentators suggest that courts should deny deemer clause protection to such plans.¹²⁸ Although such a judicial approach would require courts to ignore ERISA's text, it arguably would be consistent with a "purposive" view of statutory interpretation¹²⁹ if EHBP use of stop-loss insurance undermines the legislative goals implicit in ERISA.

The problem with this approach is that the extensive use of stop-loss insurance by EHBP does not undermine ERISA. This is not to say that EHBP that purchase low-attachment point stop-loss coverage are not exploiting a loophole in ERISA's text—they clearly are. It is unlikely that

125. See, e.g., Pitsenberger, *supra* note 5; Strain & Kinney, *supra* note 5.

126. See *Metropolitan Life Ins. Co. v. Massachusetts*, 471 U.S. 724, 747 (1985) (stating that by recognizing differential treatment received by members of insured and self-insured plans, the Court only recognizes a "distinction created by Congress").

127. See, e.g., Edward Albuero Morrissey, *Deem and Deemer: ERISA Preemption Under the Deemer Clause as Applied to Employer Health Care Plans with Stop-Loss Insurance*, 23 J. LEGIS. 307, 314 (1997); cf. Robert N. Covington, *Amending ERISA's Preemption Scheme*, KAN. J.L. & PUB. POL'Y, Winter 1999, at 1 (suggesting the need to amend the statute); Mark Alan Edwards, Comment, *Protections for ERISA Self-Insured Employee Welfare Benefit Plan Participants: New Possibilities for State Action in the Event of Plan Failure*, 1997 WIS. L. REV. 351, 368-69 (suggesting that "[t]he most logical way to reform the preemption effects of a federal statute . . . is to amend the statute itself").

128. *Brown v. Granatelli*, 897 F.2d 1351, 1355 (5th Cir. 1990); Sebelius, *supra* note 110, at 11 n.9 (arguing that self-insured employers with low attachment point stop-loss insurance should not be considered "self-insured" and should be required to follow state mandates).

129. HENRY C. BLACK, *HANDBOOK ON THE CONSTRUCTION AND INTERPRETATION OF THE LAWS* § 33 (1911); RONALD B. BROWN & SHARON J. BROWN, *STATUTORY INTERPRETATION: THE SEARCH FOR LEGISLATIVE INTENT* § 4.5 (2002).

the members of Congress who enacted ERISA in 1974 anticipated that EHBPs would be able to avoid the indirect effects of state insurance regulation while avoiding most insurance risk and would do so in large numbers. But the Congressional purpose inherent in ERISA's savings and deemer clauses is not to prevent small employers from avoiding state mandates. Rather, the purpose of ERISA's complicated structure is to balance traditional state authority to regulate insurance with employers' interest in avoiding the burden of complying with conflicting state laws.¹³⁰ In striking that balance, ERISA creates two categories of EHBPs—those that purchase third-party insurance for their employees and those that do not—and allows the EHBPs themselves to choose their category. By creatively identifying ways of making it less costly to choose one category rather than the other, EHBPs act consistently rather than inconsistently with ERISA structure.

III. EXPLOITING THE SAVINGS CLAUSE LOOPHOLE

As Part II explained, ERISA's text in no way suggests that whether an EHP purchases stop-loss insurance has any relevance to the question of whether the terms of health care coverage it provides its members are subject to state regulation, regardless of whether the attachment point of the stop-loss coverage is high or low. Additionally, there is no justification for courts to ignore ERISA's text in an effort to vindicate its purpose, because employers' use of stop-loss insurance does not undermine ERISA's attempt to balance competing policy goals. These conclusions do not suggest, however, that proponents of greater state regulation of health insurance must concede that they have been outmaneuvered by crafty employers. Just as the deemer clause creates a loophole that employers can exploit in an effort to minimize the reach of state regulation, the savings clause creates a loophole that states may exploit in an effort to maximize their regulatory reach.

ERISA permits EHBPs to arrange their business affairs in such a way that maximizes the benefits of self-insuring relative to purchasing third-party insurance. But ERISA also permits states to use their authority under the savings clause to maximize the benefits to EHBPs of purchasing third-party insurance relative to self-insuring. If states are unhappy that EHBPs use stop-loss insurance to make self-insuring a relatively more attractive option than purchasing state-regulated third-party insurance, their best

130. See, e.g., *Ingersoll-Rand*, 498 U.S. at 142 (describing ERISA's intent to minimize burdens on employers of conflicting state regulations).

response is to regulate stop-loss insurers in a way that undermines that advantage.

To date, some states have attempted to exploit this loophole, but their success has been limited in two ways. First, the Second Circuit's decision in *Travelers Insurance Co. v. Cuomo*¹³¹ and the Fourth Circuit's decision in *American Medical Security v. Bartlett*¹³² invalidated two states' attempts to exploit this loophole, casting doubt on its legality. Both cases were incorrectly decided when issued, however, and the Supreme Court's recent opinion in *Kentucky Ass'n* further undermines them. Therefore, those decisions should not deter states—certainly those outside of the Second and Fourth Circuits—from taking advantage of the loophole. Second, since no state has yet exploited the loophole as fully as is possible, its full effect has never been tested. This Part describes the savings clause loophole, analyzes its legal status, explains how states can expand it, and considers the policy consequences of doing so.

A. Exploiting the Loophole with Minimum Attachment Points

In 1995, the National Association of Insurance Commissioners (NAIC) enacted a model statute requiring that stop-loss insurance policies sold to EHBP to protect against excessive health care expenses include minimum attachment point levels.¹³³ The NAIC model calls for specific attachment points to be a minimum of \$20,000, and minimum aggregate attachment points to be 110 to 120 percent of the EHBP's expected annual claims, depending on the size of the EHBP covered.¹³⁴ Currently, at least fifteen states have adopted statutes or promulgated administrative regulations along the lines of the NAIC model.¹³⁵ The purpose of such state regulations

131. 14 F.3d 708 (2d Cir. 1993), *rev'd in part by* N.Y. State Conference of Blue Cross & Blue Shield Plans v. Travelers Ins. Co., 514 U.S. 645 (1995).

132. 111 F.3d 358 (4th Cir. 1997).

133. STOP LOSS INS. ACT (Nat'l Ass'n of Ins. Comm'rs 2004).

134. *Id.*

135. State statutes follow the structure of the NAIC model, although they sometimes change the aggregate and specific attachment points. *See* ALASKA STAT. § 21.42.145 (2004) (requiring stop-loss policies in Alaska to have at minimum a specific attachment point of \$10,000 and an aggregate attachment point for small employers of the greatest of \$4000 times the number of individuals, 120% of expected claims, or \$20,000); GA. CODE. ANN. § 33-50-5 (2002) (giving the Georgia Insurance Commissioner ability to review stop-loss policies); COLO. REV. STAT. § 10-16-119 (2002) (requiring a \$15,000 minimum specific attachment point and 120% minimum aggregate attachment point for policies issued in or after 2003); MINN. STAT. § 60A.235 (2002) (requiring stop-loss policies in Minnesota to have

is no secret. As the Maryland Insurance Commissioner explained when initially promulgating that state's regulation, the goal of such rules is to prevent EHBPs from substituting stop-loss coverage for third-party health insurance in order to avoid the costs of state mandated benefits and other regulations while continuing to shift the insurance risk of employee illness to third parties.¹³⁶

The deemer clause clearly prevents states from regulating EHBPs, but the savings clause just as clearly allows states to regulate insurance as long as the regulations do not extend to EHBPs. Stop-loss coverage providers, like health insurers, are in the business of insurance. If states may regulate the terms and conditions of health insurance policies, as *Metropolitan Life* clearly established that they can, states may also regulate the terms and conditions of stop-loss policies. When states require health insurance policies to include specific benefits, the option of purchasing third-party health insurance becomes less attractive to some EHBPs relative to the option of self-insuring. When states require stop-loss policies to include minimum attachment points, the option of purchasing third-party health insurance becomes more attractive to some EHBPs relative to the option of self-insuring. The fact that the savings clause gives states a tool to encourage EHBPs to purchase third-party insurance is no doubt an

a minimum specific attachment point of \$10,000 and a minimum aggregate attachment point higher than the sum of 140% of the first \$50,000 of expected claims, 120% of the next \$450,000 of expected claims, and 110% of the remaining expected claims); N.J. STAT. ANN. § 17B:27A-17 (West 2003) (requiring stop-loss policies in New Jersey to have a minimum specific attachment point of \$25,000 and a minimum aggregate attachment point of 125% of expected claims); OR. REV. STAT. § 742.065 (2001) (requiring stop-loss policies in Oregon to have a minimum specific attachment point of \$10,000 and a minimum aggregate attachment point of 120% of expected claims); 191 IOWA ADMIN. CODE § 35.20(2)(g) (requiring stop-loss policies in Iowa to have minimum aggregate coverage at 125% of actuarially projected claims); NEV. ADMIN. CODE ch. 689B § 350 (2003) (requiring stop-loss policies in Nevada to have a minimum specific attachment point of \$10,000; a minimum aggregate attachment point for groups not more than fifty people that is lower than the greater of \$4000 times the number of group members, 120% of expected claims, or \$10,000; for groups of more than fifty people an aggregate attachment point not lower than 110% of expected claims); 31 PA. CODE § 89.472 (2003) (requiring stop-loss policies in Pennsylvania to have a minimum specific attachment point of \$10,000 and a minimum aggregate attachment point of \$100,000).

136. See *Am. Med. Sec.*, 111 F.3d at 362; see also *Van Enters. v. Avemco Ins. Co.*, 231 F. Supp. 2d 1071, 1084 (D. Kan. 2002) (citing a bulletin issued by the Kansas Insurance Commissioner describing a desire to regulate stop-loss insurance because some "self-funded arrangements [we]re being formed for the purpose of avoiding compliance with Kansas' recent health insurance reform legislation").

unintended consequence of ERISA's structure. But then, so is the fact that the deemer clause allows EHBPs to use self-insurance coupled with low-attachment point stop-loss coverage to avoid the costs of state mandates without retaining any real insurance risk.

B. The Loophole Closed? Two Circuits Invalidate Stop-Loss Regulations

1. Travelers Insurance Co. v. Cuomo

In 1993, the Second Circuit considered challenges to a New York statute imposing hospital rate surcharges that differed based on the identity of the payer of hospital charges and to a New York Insurance Department regulation of stop-loss insurance contracts.¹³⁷ The court held that ERISA preempted both the surcharges and the stop-loss regulations.¹³⁸ The Supreme Court granted certiorari on the question of whether the surcharges were preempted¹³⁹ and then, in a landmark decision that narrowed the scope of ERISA's "relates to" clause, reversed.¹⁴⁰ The Court left unreviewed, however, the portion of the Second Circuit's opinion concerning stop-loss insurance regulation.

The stop-loss regulation at issue required stop-loss insurers to "undertake to ensure that statutorily mandated benefits be covered" by the underlying EHP.¹⁴¹ The *Travelers* court held that the regulation related to employee benefit plans and was not protected by the savings clause because it did not qualify as an insurance regulation. The court's analysis, in relevant part, consisted of the following points: (1) the provision was "not limited just to the stop-loss layer of insurance but appl[ied] generally to the entire" EHP; and (2) the regulation did "not have the effect of transferring or spreading risk between a self-funded plan and its stop-loss insurer."¹⁴² Although both descriptions are fair characterizations of the New York regulation's effects, neither supports the conclusion that the provision does not constitute an "insurance regulation" protected by the savings clause. Any law governing to whom and under what conditions an

137. *Travelers*, 14 F.3d at 711.

138. *Id.*

139. N.Y. State Conference of Blue Cross & Blue Shield Plans v. Travelers Ins. Co., 514 U.S. 645, 654 (1995).

140. *Id.* For a detailed analysis of the Supreme Court's decision and its effect the jurisprudence of ERISA's "relates to" clause, see Korobkin, *supra* note 27, at 488-90.

141. *Travelers*, 14 F.3d at 724.

142. *Id.*

insurance company may sell insurance products to customers constitutes a “regulation of insurance” under any common-sense understanding of that term.

The Supreme Court’s decision last term in *Kentucky Ass’n* jettisoned the Court’s prior complicated test for whether a state law constitutes an insurance regulation for savings clause purposes, replacing it with a simpler approach consistent with this view. In so doing, it severely undermined the Second Circuit’s holding in *Travelers*. Recall that under the rule of *Kentucky Ass’n*, a state law qualifies for protection from ERISA preemption under the savings clause so long as the law is “specifically directed” at the insurance industry and it regulates insurance practices.¹⁴³ The New York regulation clearly satisfied both prongs of this test, as the following paragraphs explain.

In *Kentucky Ass’n*, petitioner health insurance companies argued that a state “any willing provider” (AWP) law preventing insurers from excluding health care providers from their networks was not specifically directed at insurers because it equally affected providers.¹⁴⁴ Disposing of this argument, the Court first observed that, by its terms, the statute imposed requirements only on insurers.¹⁴⁵ It then explained that the fact that a regulation of insurance entities has the consequence of affecting the choices available to other entities does not preclude savings clause protection for the regulation.¹⁴⁶ Like the Kentucky AWP law, the New York regulation at issue in *Travelers* was specifically directed only at insurance companies (specifically, those that sell stop-loss insurance policies). By essentially forbidding insurance companies from selling policies to EHBP that do not provide the full range of benefits that the state mandates of health insurers, the regulation certainly affected the range of contracting options available to New York EHBP, but not in a qualitatively different way than the *Kentucky Ass’n* AWP law impacted the range of contracting options available to Kentucky doctors. More to the point, all insurance regulations affect the market choices available to third parties who wish to contract with insurance companies;¹⁴⁷ the New York stop-loss regulation was

143. *Kentucky Ass’n of Health Plans v. Miller*, 123 S. Ct. 1471, 1475 (2003); *see also supra* Subsection I.B.3.

144. *Kentucky Ass’n*, 123 S. Ct. at 1475.

145. *Id.*

146. *Id.* at 1475-76.

147. *Id.* at 1476 (“Regulations ‘directed toward’ certain entities will almost always disable other entities from doing, with the regulated entities, what the regulations forbid; this does not suffice to place such regulation outside the scope of ERISA’s savings clause.”).

a quite ordinary insurance regulation in this respect.

The *Kentucky Ass'n* petitioners also argued that the AWP laws at issue did not regulate the insurance practices of insurance companies because those laws did not directly affect the allocation of risk between insurers and insured members.¹⁴⁸ Rejecting this argument, the Court explained that the laws affected insurance practices by limiting the “scope of permissible bargains” between insurers and potential customers, as contrasted with a regulation governing how much insurance companies must pay janitors to clean their offices, which would be directed at insurance companies but have nothing to do with the insurance function.¹⁴⁹ Similarly, the New York stop-loss regulation is directed at insurance companies qua insurance companies, rather than insurance companies qua purchasers of office supplies, insurance companies qua landlords, or insurance companies acting in some other role unrelated to the provision of insurance.

2. *American Medical Security, Inc. v. Bartlett*

In 1995, the Maryland Insurance Commissioner promulgated a regulation providing that an insurance product sold to an employer that insures against the cost of claims that result from employees’ sickness or accidents would be characterized as “stop-loss insurance” only if it has a minimum specific attachment point of \$20,000 and a minimum aggregate attachment point of at least 125% of the expected annual claims cost.¹⁵⁰ A group of employers with self-funded EHBPs sought an injunction against Maryland’s enforcement of the regulation,¹⁵¹ and the regulation became the test case for whether states possessed the power to exploit ERISA’s savings clause loophole.

In *American Medical Security, Inc. v. Bartlett*,¹⁵² the Fourth Circuit upheld a district court ruling that ERISA preempted the regulation on the ground that it “attempt[ed] to mandate benefits that certain self-insured plans may offer.”¹⁵³ *American Medical Security* remains today the leading federal court

148. *Id.* at 1477.

149. *Id.*

150. MD. REGS. CODE tit. 9 § 31.02 (1995), *repealed by* Am. Med. Sec., Inc., v. Bartlett, 111 F.3d 358 (4th Cir. 1997). In a subsequent revision of the regulation, Insurance Commissioner changed the minimum specific attachment point for stop-loss insurance to \$10,000 and the minimum aggregate attachment point to 115% of expected claims costs. 23 Md. Reg. Issue 2, Jan. 19, 1996.

151. *Am. Med. Sec.*, 111 F.3d at 360-61.

152. 111 F.3d 358.

153. *Id.* at 365.

decision on the subject of state attempts to regulate stop-loss insurance for the purpose of making self-insuring a less attractive option for EHBP.¹⁵⁴ This is unfortunate because the opinion is badly flawed. At the time it was issued it was inconsistent with ERISA's text and structure as well as existing Supreme Court decisions interpreting the statute. In addition, the Supreme Court's *Kentucky Ass'n* decision now provides further support for the contention that *American Medical Security* was wrongly decided.

Maryland's attempt to exploit the savings clause made for a bad test case from the start because that state's regulation was at best poorly drafted and at worst substantively incoherent. Rather than establishing minimum attachment points for stop-loss insurance, effectively prohibiting the sale of stop-loss insurance with lower attachment points, as the NAIC model statute does, the Maryland regulation provided that an insurance policy with a low attachment point that protected an EHBP against losses resulting from employee health care costs would be "considered to be a policy or contract of health insurance."¹⁵⁵

The most natural reading of this regulation is that low-attachment-point stop-loss policies would be required to provide coverage for state-mandated health care benefits.¹⁵⁶ The problem with such a requirement is that it is logically incoherent to require a stop-loss insurer to cover mandated health benefits because a stop-loss insurer's customers—EHBPs—need not provide state-mandated benefits to plan members in the

154. Although *American Medical Security* postdates the Second Circuit's decision in *Travelers*, the *Travelers* decision has been largely overlooked by commentators, most likely because the majority of that opinion—but *not* the portion relevant to this discussion—was subsequently reversed by the Supreme Court. Only *Travelers* and *American Medical Security* have addressed state attempts to exploit the savings clause loophole on the merits. The issue has been raised in a handful of other cases, but these other courts have resolved their cases on procedural grounds without opining on the substantive question. *See, e.g.*, *Associated Indus. of Mo. v. Angoff*, 937 S.W.2d 277, 284-85 (Mo. Ct. App. 1996) (overruling MO. CODE REGS. ANN. tit. 20 § 400-2.150, a regulation of stop-loss insurance, because the insurance commissioner needed statutory authority or actuarial data to set the minimum attachment point for stop-loss insurance); *Van Enter., Inc. v. Avemco Ins. Co.*, 231 F. Supp. 2d 1071, 1087-88 (D. Kan. 2002) (overruling a decision by the insurance commissioner to place minimum aggregate and specific attachment points on stop-loss insurance because he did not have statutory authority to do so).

155. MD. CODE REGS. ANN. tit. 9 § 31.02.

156. *Cf. Brown v. Granatelli*, 897 F.2d 1351, 1356-58 (5th Cir. 1990) (Brown, J., dissenting) (arguing that a stop-loss policy covering an EHBP is in fact a group health insurance policy under Texas law and therefore required to provide state health benefits mandates).

first instance. Consider, for example, a Maryland mandate that health insurers provide coverage for skilled nursing home facilities.¹⁵⁷ What would it mean for Maryland to require an insurance policy issued to an *EHBP* to include such coverage? Arguably, this requirement could be interpreted as preventing the stop-loss insurer from excluding the *EHBP*'s costs of providing skilled nursing care to its members from its calculation of whether an *EHBP*'s losses have reached the policy's attachment point. Such an interpretation would be unobjectionable but mostly beside the point, because the insurance commissioner's concern in enacting the regulation was with *EHBP*s that *do not* provide skilled nursing home benefits, not with stop-loss carriers that refuse to reimburse *EHBP*s for the costs of nursing home care. If an *EHBP* excludes skilled nursing home care from the benefits it promises to its members, a stop-loss insurance company will never find itself obligated to pay costs incurred as a result of plan members receiving skilled nursing home care. Because the *EHBP* does not pay nursing home costs, the fact that a plan member incurs such costs will never result in the *EHBP* making a claim against its stop-loss insurance policy. A variation of this understanding of the regulation is that it requires a stop-loss insurer to reimburse employees directly for skilled nursing home costs that the employees incur. This interpretation makes little sense either, because stop-loss insurers have no contractual obligations of any kind to employees.

A less natural, but still plausible, interpretation of the Maryland regulation is that by labeling certain stop-loss insurance policies "health insurance," the state would consider the *EHBP* purchasers of those policies to be "insured" rather than "self-insured" plans for ERISA purposes and therefore required to provide all state-mandated benefits to their members.

The problems with the Fourth Circuit's decision in *American Medical Security* begin with its failure to make clear which of these readings (i.e., stop-loss insurers must pay for mandated benefits or *EHBP*s that purchase stop-loss insurance must provide mandated benefits) it gave to the regulation under scrutiny. At one point the court "recognize[s] that the regulations are carefully drafted to focus directly on insurance companies issuing stop-loss insurance and not on the [*EHBP*s] themselves."¹⁵⁸ This statement implies the former construction. In the very next paragraph, however, the court asserts that the regulation "seek[s] to require self-

157. MD. CODE ANN., INS. § 15-801 (2003).

158. *Am. Med. Sec.*, 111 F.3d at 363.

funded plans to offer coverage consistent with state insurance law.”¹⁵⁹ This statement implies the latter interpretation.

Which of these two statements reflects the court’s interpretation of the regulation is critical. If the Maryland regulation is interpreted to mandate that self-funded EHBP’s with low-attachment-point stop-loss coverage provide specific benefits to their members, then the rule clearly would be preempted, but the court’s opinion would be of trivial importance because it merely follows well-established deemer-clause precedent—states may not regulate EHBP’s as if they are insurance companies, even if they serve an insurance function.¹⁶⁰ If the Maryland regulation places requirements only on stop-loss insurance companies, as it appears to according to its text, however, the resolution of the case takes on a great deal of importance. This court’s confusion is understandable in light of the regulation’s incoherence, but it makes it quite difficult to divine the court’s holding in the case.

The court’s failure to clearly state its interpretation of what the regulation at issue actually requires suggests a lack of understanding on its part that its precise resolution of this question is important to the case. The best explanation of why *American Medical Security* explains the court’s understanding of the Maryland regulation so poorly is that the court determined that ERISA preempts the regulation regardless of its precise meaning. The court’s opinion repeatedly emphasizes that the Maryland regulation had the “purpose and effect” of influencing the behavior of self-funded plans.¹⁶¹ It asserts that the deemer clause prohibits Maryland law from “aiming at the plan-participant relationship.”¹⁶² It also concludes that such purpose and effect calls into doubt whether the savings clause protects Maryland regulation because it arguably fails the Supreme Court’s savings clause requirements of being a state law directed at the insurance industry and being integral to the insured-insurer relationship.¹⁶³ Thus, the court appears to believe that ERISA’s deemer clause, and perhaps also its savings clause, prohibit state regulations enacted with the intent or effect of increasing the attractiveness to EHBP’s of purchasing third party health insurance for their members, *regardless of whether the state directly regulates*

159. *Id.* at 363-64.

160. *See supra* Section II.C.

161. *Am. Med. Sec.*, 111 F.3d at 363.

162. *Id.* at 364.

163. *Id.* at 363 (“[T]he complications of the second and third *Metropolitan Life* factors [concerning the savings clause] together with the ‘deemer clause’ provide the core difficulty with the state’s regulation of stop-loss insurance policies issued to ERISA plans.”).

what choices EHBP's must make or, alternatively, regulates the products third-party insurance companies may sell.

There is no text or precedent that supports this interpretation of ERISA. States' power to regulate under the savings clause is limited only by the deemer clause's prohibition against applying insurance regulations to EHBP's themselves. *American Medical Security's* conclusion that it is "impermissibl[e]" for state regulations to affect ERISA plans' "costs and choices" finds no support in ERISA's language and is inconsistent with the statute's structure,¹⁶⁴ which allows for the preemption of a state law that "relates to" EHBP's and then the saving of that same law as an insurance regulation. The court's sweeping statement effectively reads the savings clause out of the statute. The court's conclusion also is inconsistent with the Supreme Court's ruling in *Metropolitan Life*. State benefits mandates increase the costs associated with purchasing health insurance for EHBP's and limit the choices available in the insurance market. Stop-loss insurance regulations merely have similar effects on an EHBP's decision to self-insure.

The Fourth Circuit's analysis begins to go awry when it asserts that "state insurance regulation may not directly or indirectly regulate self-funded ERISA plans"¹⁶⁵—a statement of the law that is, at best, misleading. As explained above, whether an EHBP is insured or self-insured is not the operative question under ERISA. The statute prohibits states from directly regulating EHBP's, whether they purchase third-party insurance or not. But a state regulation of insurance may indirectly affect options available to EHBP's, and it may therefore make self-insurance more or less attractive to EHBP's. If a state law having such an effect is considered an "indirect regulation," then the Fourth Circuit's statement of the law is incorrect.

As authority for its "directly or indirectly" statement, *American Medical Security* cites to the Supreme Court's decision in *FMC Corp.*,¹⁶⁶ but the relevant passage in *FMC Corp.* is actually itself a quotation from the Supreme Court's earlier opinion in *Metropolitan Life*. In the original statement, the Court said: "We are aware that our decision [upholding a mental health mandate as applied to third-party insurance companies] results in a distinction between insured and [self-insured] plans, leaving the former open to indirect regulation while the latter are not."¹⁶⁷ This

164. *Id.* at 364.

165. *Id.* at 361.

166. *Id.* (citing *FMC Corp. v. Holliday*, 498 U.S. 52, 62 (1990)).

167. *Metropolitan Life Ins. Co. v. Massachusetts*, 471 U.S. 724, 747 (1985), *quoted in FMC Corp.*, 498 U.S. at 62.

sentence recognizes that an EHBP that purchases third-party health insurance for its employees is indirectly subject to benefits mandates (because it cannot buy insurance that does not include the specified benefits), whereas an EHBP that self-insures its employees' health care costs is not affected at all—indirectly or otherwise—by the state law. The mere observation that state benefits mandates do not have even an indirect effect on self-insured EHBPs does not logically imply that other types of state insurance regulations that do have an indirect effect on self-insured EHBPs are therefore prohibited, as it would have to for *Metropolitan Life* to support the Fourth Circuit's conclusion in *American Medical Security*. In fact, *Metropolitan Life* suggests exactly the opposite: A state's direct regulation of insurance may permissibly have the effect of skewing the cost benefit analysis of an EHBP deciding whether to self-insure its members' health care costs or purchase third-party health insurance policies for that purpose.

Although the *American Medical Security* court's decision finds no support in ERISA's language or in the relevant Supreme Court decisions on related issues, the court's decision could be defensible nonetheless if the specific type of state law in question undermines ERISA's intent.¹⁶⁸ The problem is that the argument that states should be prohibited from regulating stop-loss insurance in a way that makes self-insurance less attractive to EHBPs runs into precisely the same trouble as does the argument that EHBPs should be prohibited from using stop-loss insurance with low attachment points to make self-insurance more appealing.¹⁶⁹ ERISA balances the value of allowing states to regulate insurance with the value of allowing employers to avoid inconsistent state laws. To effectuate this balance, Congress gave states the right to regulate insurance companies under the savings clause and gave EHBPs a safe harbor under the deemer clause to avoid such regulation by not purchasing third-party insurance. Congress did not intend for ERISA to make it particularly easy or cost-free for EHBPs to opt to finance their member benefits without third-party insurance any more than it intended to make self-insuring particularly burdensome. State attempts to use stop-loss insurance regulation to make self-insurance less attractive to EHBPs exploit a loophole, but such exploitation—like EHBPs exploitation of the deemer clause loophole—is consistent, rather than inconsistent, with ERISA's structure.

Immediately after losing in the Fourth Circuit, Maryland enacted

168. BROWN & BROWN, *supra* note 129, § 2.1; DICKERSON, *supra* note 129, at 67-102.

169. See *supra* Section II.C.

legislation with the identical purpose to the regulation that was struck down¹⁷⁰ but with language that followed the NAIC's Model Rule and avoided the confusion surrounding the regulation's definition of some stop-loss policies as health insurance.¹⁷¹ The statute prohibited insurance companies from selling stop-loss policies with specific attachment points lower than \$10,000 or aggregate attachment points lower than 115% of expected annual claims.¹⁷² American Medical Security claimed that the new statute was substantively no different than the regulation prohibited by the Fourth Circuit;¹⁷³ the Maryland Insurance Commissioner claimed that the statute was protected from preemption by the savings clause.¹⁷⁴ Both sides were correct.

American Medical Security asked a federal district court to enjoin enforcement of the statute, but the court ruled that since the statute was newly enacted, the insurance companies would have to proceed with a new challenge on the merits.¹⁷⁵ The company initially promised a new court fight on the ground of ERISA preemption,¹⁷⁶ but it later chose instead to withdraw from the Maryland insurance market and drop its challenge.¹⁷⁷ The company's general counsel said he was "not sure if anyone will want to put the money behind a challenge to the law."¹⁷⁸ To date, no one has.

ERISA is a complicated statute, to be sure, but the operation of the savings clause and the deemer clause are spelled out rather clearly: States may regulate the sale of insurance, but their insurance regulations may not extend to self-insured EHBPs that serve an insurance function vis-à-vis their members but are not otherwise in the insurance business. The flawed analysis of the Second and Fourth Circuits notwithstanding, this statutory

170. Maryland's Deputy Insurance Commissioner explained that the statute prevents insurance companies from selling stop-loss policies to self-insured employers that "are really a sham to avoid state regulations." Dennis Kelly, *Maryland Has Begun Enforcing*, BESTWIRE, July 26, 1999 (quoting Deputy Insurance Commissioner Dennis Carroll); see also *Maryland Regulators Enforcing Law on Stop-Loss Attachments*, BUS. INS., July 26, 1999, at 1.

171. MD. CODE ANN., INS. § 15-129 (1999).

172. *Id.*

173. See Jerry Geisel, *Stop-Loss Enforcement OK'd: But Maryland Likely to Face Additional Legal Challenges*, BUS. INS., Sept. 6, 1999, at 2.

174. See *id.*

175. See *id.*

176. See *id.*

177. See Jerry Geisel, *Stop-Loss Battle To End: Insurer To Withdraw from Maryland, Ending Challenge to Law*, BUS. INS., Nov. 1, 1999, at 2.

178. *Id.* (quoting Tim Moore, General Counsel and Senior Vice President of American Medical Security).

balance permits states to place limits on what types of stop-loss insurance products insurance companies may sell and to whom they may sell them.

*C. Further Exploiting the Savings Clause Loophole:
“Underlying Coverage” Requirements*

Since the Second Circuit’s decision in *Travelers*, all of the states that have attempted to exploit the savings clause loophole have employed what might be labeled a restrained approach. Specifically, those states have required only that stop-loss insurance have minimum attachment points so that EHBPs cannot seamlessly replace third-party health insurance with stop-loss insurance and avoid the costs of state benefits mandates and insurance taxes without sacrificing any of the benefits of third-party insurance. The reason for such regulatory restraint is understandable: Only EHBPs that purchase stop-loss insurance with very low-attachment points are purchasing such insurance for the obvious purpose of exploiting the deemer clause loophole and dodging the cost of state law requirements. If states wish to use the savings clause loophole to neutralize the deemer clause loophole, focusing only on low-attachment point stop-loss insurance is appropriate.

Mandating minimum attachment points for stop-loss insurance is a restrained regulatory approach because, although it makes the decision to self-insure (and thereby avoid state health insurance mandates) less desirable from the perspective of EHBPs, it makes that decision only marginally less desirable. With such stop-loss insurance regulations enacted, an EHP that wishes to avoid the costs of state insurance mandates must maintain some insurance risk, but it is still able to cede most of the insurance risk to a third-party by purchasing stop-loss insurance with the minimum permissible attachment points. Given this set of choices, some EHBPs that would choose to self-insure and purchase stop-loss insurance with very low attachment points are likely to decide to purchase third-party health insurance instead, but many would choose to purchase stop-loss insurance with higher attachment points and continue to avoid the consequences of state insurance regulations, including benefits mandates and premium taxes.

The minimum attachment point approach to regulation fails to recognize that the savings clause loophole can be exploited to far greater effect. Specifically, following the New York regulation improperly struck down in *Travelers*, states could enact legislation or promulgate regulations that prohibit insurance companies from selling any stop-loss coverage for

losses associated with health care costs unless the underlying coverage provided by the EHBP to its employee members includes all of the state-mandated benefits that insurance companies must provide. Such laws might be called “underlying coverage requirements” for stop-loss insurance.

Much like minimum attachment point requirements, underlying coverage requirements would be directed at the insurance practices of insurance companies and should be protected from preemption under the savings clause on that basis. It is true that such requirements would have the indirect effect of preventing other entities (here, EHBPs) from doing in concert with an insurance company what the law prohibits insurance companies from doing. This fact, however, does not vitiate savings clause protection for the restriction placed on insurance companies.¹⁷⁹ And because such a law would not place any requirements on EHBPs, it would not interfere with the safe harbor provided by the deemer clause.¹⁸⁰

An underlying coverage requirement would render the option of self-insuring members’ health care costs far less attractive to EHBPs that currently self-insure their members’ medical costs, purchase stop-loss insurance to reduce insurance risk, and provide their members a menu of benefits that does not include all state mandated coverage. Such EHBPs would have to choose between (1) maintaining their stop-loss insurance and expanding the benefits they provide to their members to include all state-mandated benefits, (2) purchasing third-party health insurance for their members that includes all state-mandated benefits, or (3) functionally self-insuring their employees’ medical costs by retaining the entire insurance risk. There is little doubt that many EHBPs that currently self-insure and offer a limited set of benefits to their members would choose either the first or the second option, especially if they do not have extremely large risk pools. In other words, a likely consequence of a state instituting an underlying coverage requirement would be that employees who enjoy health care coverage and work for all but the largest employers would receive the benefits mandated by the state. Such a result would be consistent with the expectations of ERISA’s drafters, who envisioned that the type of large employers that operate in multiple states could develop a single EHBP for all of their employees free from inconsistent state regulations, but that states would otherwise continue to regulate health insurance.¹⁸¹

179. See *Kentucky Ass’n of Health Plans v. Miller*, 123 S. Ct. 1471, 1475 (2003).

180. See *FMC Corp. v. Holliday*, 498 U.S. 52, 61-64 (1990).

181. See generally GAO, ISSUES AND TRENDS, *supra* note 102.

D. Drawbacks to Exploiting the Savings Clause Loophole

From the perspective of state regulators, there are three primary drawbacks to exploiting the savings clause loophole, which vary in their nature and severity: (1) some employers might end their sponsorship of EHBPs; (2) employers operating in multiple jurisdictions might be able to escape the impact of the state's efforts; and (3) financially precarious employers might drop stop-loss insurance coverage, increasing their insolvency risk. Each of these drawbacks affects both the minimum attachment point and the underlying coverage requirement regulatory approach, although each is likely to have a more significant impact on underlying coverage requirements. The first drawback should not be a significant concern in light of policy choices already made by the state; the second should also not be a significant concern because it could reduce the effectiveness of attempts to exploit the loophole but would not have independently undesirable consequences; the third should give regulators significant pause, especially in the case of underlying coverage requirements.

1. Loss of Benefits

The first drawback to exploiting the savings clause loophole is that doing so runs the risk of causing some employers that currently sponsor EHBPs to stop providing any health care coverage to their employees. No federal or state law (with the exception of Hawaii) requires employers to provide health care coverage as a fringe benefit;¹⁸² thirty-four percent of the nation's employers—and forty-five percent of employers with fewer than nine workers—do not.¹⁸³ The cost of mandates varies from state to state, of course, but few doubt that they are substantial, and some estimates suggest they can account for up to nearly one-fourth of health care claims

182. Hawaii law mandates employer-provided health coverage, *see* HAW. REV. STAT. § 393 (2004), but the state received an exception from the federal government for its law. *See* 29 U.S.C. 1144 (2000). A recently enacted California statute would have mandated that employers with more than fifty employees provide health care coverage or pay into a state pool to fund the cost of providing such coverage. Health Insurance Act of 2003, CA. LAB. CODE §§ 2120-2210 (West Supp. 2004). California voters blocked the law's implementation, however, by defeating a ballot proposition in November 2004. Jordan Rau & Evan Halper, *Election 2004*, L.A. TIMES, Nov. 4, 2004, at B1. Whether the measure would have survived ERISA preemption analysis is unclear.

183. KFF/HRET, *supra* note 93, at 40.

costs.¹⁸⁴ State insurance premium taxes also increase the cost to employers of providing third-party insurance for their employees. Faced with a choice between providing more expensive health care coverage through the purchase of third-party insurance, continuing to self-insure but doing so without the safety-net of stop-loss insurance or with higher-attachment-point stop-loss insurance, or simply dropping health care coverage from their menu of fringe benefits entirely, some employers will—and many might—select the last option. The ironic effect could be that the very employees state regulators desire to protect could be left worse off.

This concern, however, is one that is broader than the issue of employers who self-insure in order to avoid state benefits mandates or premium taxes. State regulators face the same risk when they decide to impose any mandate or tax on the state's insurance companies.¹⁸⁵ Each mandate or tax increases the cost to employers of providing third-party health insurance, thus increasing the risk of marginal employers dropping their sponsorship of health insurance. By one estimate, one-fifth of small employers that do not currently offer health care benefits to their employees would do so if there were no benefits mandates.¹⁸⁶

A state that mandates that health insurance companies provide specific benefits presumably has already decided that the gains to employees whose employers provide increased coverage to meet the minimum requirements outweigh the costs to employees whose employers elect to eliminate health care coverage altogether. Having already accepted this trade-off, it is not clear why a state would hesitate to impose an underlying coverage mandate on stop-loss insurance companies (except, perhaps, if the state accepted the risks associated with imposing benefits mandates only as a result of its knowledge that cost-conscious employers could avoid those mandates without dropping coverage by becoming self-insured and purchasing stop-loss coverage). Put another way, if a state fears that imposing an underlying coverage mandate will cause many employers to cancel their EHBPs, it

184. According to a GAO study, Virginia's mandated benefits account for twelve percent of group health insurance claims, Maryland mandated benefits account for twenty-two percent of claims, and Iowa mandated benefits only account for only five percent of claims. U.S. GEN. ACCOUNTING OFFICE, GAO/HEHS-96-161, *HEALTH INSURANCE REGULATION: VARYING STATE REQUIREMENTS AFFECT COST OF INSURANCE* 11 (1996).

185. They face a similar concern when they decide whether to raise the state minimum wage, knowing that some employers might lay-off low-wage employees rather than increase wages to comply with the new minimum. Robert A. Hillman, *The Rhetoric of Legal Backfire*, 43 B.C. L. REV. 819, 852 (2002); Daniel Shavero, *The Minimum Wage, the Earned Income Tax Credit, and Optimal Subsidy Policy*, 64 U. CHI. L. REV. 405 (1997).

186. See Jensen, *supra* note 42.

should reconsider the efficacy of its benefits mandates.

2. *Jurisdiction Jumping*

A second drawback to regulating stop-loss insurance providers is that, to the extent that only some states choose to enact such regulations, they might be relatively easy for EHBPs of some multi-state employers to avoid.

States' authority to regulate insurance allows them to set the terms and conditions of insurance policies issued in their state. A typical state statute prohibits any person from transacting "a business of insurance in" or "relative to a subject of insurance resident, located or to be performed in" the state without complying with applicable provisions of state law.¹⁸⁷ Thus, State A can require that insurance companies that sell health insurance policies covering employees residing in that state comply with applicable state requirements.¹⁸⁸ But State A lacks authority over the terms and conditions of policies an insurer sells to customers that reside in State B.¹⁸⁹ If an employer's entire business operation is located in State B, State A should have little concern with the terms of a stop-loss insurance policy that the employer purchases. Difficult questions arise, however, when an employer's business operates in multiple states. An employer with headquarters in State A, its major plant in State B, and employees spread out over States A, B, C, D, might respond to the imposition of an underlying coverage mandate on stop-loss insurers in State A by purchasing its stop-loss policy in State B from a company duly licensed there. It is unclear whether, and in what precise cases, State A could assert regulatory authority over the terms and conditions of the stop-loss policy under these circumstances—complicated choice of law questions are involved.¹⁹⁰

187. See, e.g., ARIZ. REV. STAT. § 20-107 (2004); FLA. STAT. ch. 624.11 (2004); IDAHO CODE § 41-113 (2004); NEV. REV. STAT. 679A.150 (2004); WYO. STAT. ANN. § 26-1-103 (2004).

188. See *Guardian Life Co. of Am. v. Ins. Comm'r*, 446 A.2d 1140 (Md. App. 1982) (holding that a group health insurance policy sold by a Rhode Island insurance company to a Maryland employer and covering Maryland employees must comply with Maryland insurance requirements notwithstanding that the master policy was delivered by the insurance company to a Rhode Island trustee rather than to the employer in Maryland).

189. *Rizzo v. Price*, 294 A.2d 541 (Conn. 1972); *Standard Ins. Co. v. Sturdevant*, 566 P.2d 52 (Mont. 1977). For statutes, see CONN. GEN. STAT. §§ 1-1, 38-69, 38-71, 38-72 (2003); MASS. GEN. LAWS ANN. ch. 175, § 162; (1999); and N.Y. INS. LAW § 2114, 2115(a), 2116 (2004).

190. See generally Banks McDowell, *Choice of Law in Insurance: Conflicts Methodology To Minimize Discrimination Among Policyholders*, 23 CONN. L. REV. 117 (1990) (discussing the complex choice of law problems that arise when multiple states have an interest in the law

The possibility that employers operating in multiple states might avoid the effects of stop-loss insurance in one of those states by contracting for stop-loss insurance in a different jurisdiction suggests that the savings clause loophole might be somewhat narrower than regulators wish it would be. But the possibility that exploiting the savings clause loophole might not be a complete response to all employers that exploit the deemer clause loophole does not logically provide an argument against regulators taking action. At the very least, minimum attachment point regulations and underlying coverage mandates would have their desired effect on employers whose operations are located in a single state, and employers with multiple-state operations best able to avoid the consequences of such regulations will tend to be larger employers less likely to use low-attachment-point stop-loss insurance for the sole purpose of avoiding state benefits mandates.

3. Risk of Insolvency

The third drawback to regulating stop-loss insurance is that some employers might choose to continue to operate a self-insured EHBP but without stop-loss coverage. This risk is especially great if the state institutes an underlying coverage requirement rather than merely a minimum attachment point. Self-funded employers without stop-loss insurance expose their employees to an increased risk of plan insolvency. ERISA imposes no federal financial solvency requirements on EHBPs,¹⁹¹ and the deemer clause prevents states from imposing any on them. A catastrophic medical expense incurred by a participant in a self-funded EHBP without stop-loss coverage could cause the plan to become insolvent, leaving plan members without coverage or increasing the financial burden on any state-sponsored fund that might insure such insolvencies or provide health benefits to the uninsured.¹⁹² That the potential for self-insured EHBPs to

governing an insurance contract).

191. In contrast, ERISA does include detailed requirements for employer-sponsored pension plans to insure their solvency. Donald T. Bogan, *Protecting Patient' Rights Despite ERISA: Will the Supreme Court Allow States To Regulate Managed Care?*, 74 TUL. L. REV. 951, 975 (2000).

192. To cover members of an insured EHBP that becomes insolvent, states provide health guaranty associations. All states have enacted statutes based on an NAIC Model Law that requires insurance companies to join as a condition of transacting business in the state. The association is run by a board of directors, who determine the action necessary to cover individuals who lose health care coverage as a result of their company's insolvency. This action can range between assuming coverage for individuals from the guaranty fund to

become insolvent is of concern to state regulators is demonstrated by the fact that all states have solvency requirements for insurance companies.¹⁹³ This concern is also demonstrated by the fact that many states require self-insured employee health benefit plans that are exempt from ERISA preemption for various reasons *to purchase stop-loss insurance*.¹⁹⁴

Unlike the drawbacks of regulating stop-loss insurance described above, the increased risk of EHBP insolvency is an independent negative consequence of regulation that regulators must balance against the benefits of encouraging more employers to provide state mandated benefits as part of their self-insured plans or choosing to purchase third-party insurance subject to those state mandates.

CONCLUSION

For better or worse, ERISA creates a two-tiered structure for the regulation of employer-provided health insurance benefits. Employers who

dispersing the individuals between member insurance companies. LIFE AND HEALTH INS. GUARANTY ASS'N MODEL ACT (Nat'l Ass'n of Ins. Comm'rs 2003). For examples of relevant state regulations, see ARIZ. REV. STAT. §§ 20-681 to 20-695 (2004); CAL. INS. CODE § 1067.18 (2004); COLO. REV. STAT. §§ 10-20-101 to 10-12-120 (2004); 215 ILL. COMP. STATS. 5/531-01 to 5/531-19 (2004); MD. CODE ANN., INS. §§ 9-401 to 9-419 (2004); MO. REV. STAT. §§ 376.715- 376.758 (2004); N.Y. INS. LAWS §§ 7701- 7718 (2004).

193. See, e.g., ALASKA ADMIN. CODE tit. 3, § 21.241 (2003) (placing limitations on the type of investments and the percentage of total investments that health insurers can place in those investments); *id.* § 21.320 (regulating the amount and type of reserves that insurance companies must retain); CONN. AGENCIES REGS. § 38a-214-4 (2003) (requiring health insurers to retain a certain amount in contingency reserves); FLA. ADMIN CODE r. 4-137.001 (describing insurer reporting requirements); IOWA ADMIN. CODE r. 191-41.11(514B) (establishing minimum net worth for HMOs); MD. REGS. CODE tit. 31, § 12.01 (setting HMO fiscal requirements).

194. For example, Georgia, South Carolina, New Mexico, Minnesota, among others, require multiple-employer health benefits plans, which are exempt from ERISA's preemption protection against state laws, to buy stop-loss insurance. GA. COMP. R. & REGS. r. 120-2-50-.05 (2003) ("[A] multiple employer self-insured health plan is required to obtain individual and aggregate stop-loss coverage from an insurer authorized to transact business in Georgia"); MINN. R. 2765.1300 (2002) (requiring every joint self-insurance arrangement to carry both individual and aggregate stop-loss insurance); N.M. ADMIN. CODE tit. 13 § 19.4.16 (2004) ("[E]very MEWA shall have at all times individual and aggregate excess stop-loss coverage from an insurer authorized to transact insurance in the state of New Mexico."); S.C. CODE ANN. REGS. 69-42. Other states, Wisconsin, require local government units to obtain stop-loss insurance to cover their self-insured plans. WIS. ADMIN. CODE INS. 8.11 (2003).

purchase third-party health insurance are indirectly subject to state regulation because the savings clause exempts state regulation of insurance companies from federal preemption. Employers who self-insure their EHBP are not subject to state level requirements, because the deemer clause prevents state regulators from treating EHBPs like insurance companies. This statutory structure, reinforced by recent Supreme Court rulings expanding the scope of the savings clause, provides an incentive for employers who wish to avoid state mandated benefits and premium taxes but who do not wish to bear insurance risk to self-insure their EHBPs and purchase low attachment point stop-loss insurance. This approach exploits a loophole in ERISA, but one that is unavoidable in light of the text of the deemer clause; it would be improper for courts to judicially close the loophole in the light of ERISA's text and structure.

The savings clause, however, provides regulators with a loophole of their own that they may exploit to reduce the desirability to employers of exploiting the deemer clause loophole. Properly interpreted, the savings clause gives states the right to prevent insurance companies from selling low attachment point stop-loss policies to EHBPs, or even from selling stop-loss policies with any attachment point to EHBPs that do not provide state-mandated benefits for health insurance programs in general. Rather than decrying the consequences of courts properly interpreting the deemer clause, regulators need to decide whether the benefits of exercising their power over stop-loss insurance providers granted by the savings clause outweigh the corresponding costs of doing so. One good loophole deserves another.

Breathing Life into the Framework Convention on Tobacco Control: Smoking Cessation and the Right to Health

Benjamin Mason Meier, J.D., LL.M.*

The harms of smoking are global in scope, and states must act multilaterally to repel this global threat to public health. Embodying this cooperative spirit, the member states of the World Health Organization (WHO) have banded together to challenge tobacco through international law. While successful in its execution, this international effort to control smoking neglects cessation interventions, thereby offering little salvation to those whose health is at greatest risk—those already addicted to tobacco. Addressing these forgotten victims requires a new paradigm for tobacco control: the human right to health.

The WHO's Framework Convention on Tobacco Control (FCTC)¹ has created general principles of cognitive and normative consensus for international public health, challenging the globalization of smoking through the globalization of tobacco control.² Based upon a

* American Legacy Foundation Fellow, Columbia University. This Article is dedicated to the memory of Dean Lee Teitelbaum (1941-2004), whose struggle with tobacco robbed us of his legal vision. This research was made possible by a grant from the American Legacy Foundation. Its contents are solely the responsibility of the author and do not necessarily represent the official views of the American Legacy Foundation. The author wishes to thank Professors Donna Shelley and Ronald Bayer and the staffs of Columbia University's Center for Applied Public Health and the World Health Organization's Tobacco Free Initiative for their thoughtful comments on previous drafts of this Article.

1. WHO Framework Convention on Tobacco Control, WHA Res. 56.1, World Health Assembly, 56th Ass., 4th plen. mtg, Agenda Item 13, Annex, WHO Doc. A56.VR/4 (May 21, 2003), http://www.who.int/tobacco/fctc/text/en/fctc_en.pdf [hereinafter FCTC].

2. See David P. Fidler, *International Law and Global Public Health*, 48 U. KAN. L. REV. 1, 2 (1999) (noting a "globalization of public health" to oppose harms to health resulting from economic globalization); Derek Yach & Douglas Bettcher, *Globalisation of Tobacco Industry Influence and New Global Responses*, 9 TOBACCO CONTROL 206, 206 (2000) (describing the "globalisation of public health," through which "a risk culture is emerging with the realisation that many problems are global, and that states cannot deal with these problems on their own").

“convention/protocol approach” to treaty-making, the member states of the WHO intend the broad obligations of the FCTC to be supplemented by several individualized protocols, which, once ratified, will develop specific obligations for the respective aspects of tobacco control addressed by the FCTC.³ By first gaining the support of states for the minimal commitments of the framework convention, the drafters of the FCTC have assured that the Conference of the Parties for the FCTC will engage in a continuing dialogue on the specifics of international tobacco control as protocols are introduced, negotiated, and ratified.⁴ Despite this successful, albeit incremental, multilateral approach to tobacco control, neither the FCTC nor any currently proposed protocol adequately addresses the subject of smoking cessation.⁵

While emphasizing measures that indirectly reduce the demand for

3. Intergovernmental Negotiating Body on the WHO Framework Convention on Tobacco Control, *Future Protocols: Note by the Secretariat*, WHO Doc. A/FCTC/INB6/INF.DOC./2 (Jan. 18, 2003) [hereinafter Intergovernmental Negotiating Body] (noting that “the negotiation of a framework convention is not a complete process, but the beginning of one that will include the formulation of one or more protocols”); DANIEL BODANSKY, *FRAMEWORK CONVENTION ON TOBACCO CONTROL: THE FRAMEWORK CONVENTION/PROTOCOL APPROACH* 11 (1999) (“The framework convention/protocol approach allows law-making to proceed incrementally, beginning with a framework convention that establishes a general system of governance for an issue area, and then developing more specific commitments and institutional arrangements in protocols.”); *International Law and Health, Two Approaches: The World Health Organization’s Tobacco Initiative and International Drug Controls, Summary of Remarks by Laurence Boisson de Chazournes*, 94 AM. SOC’Y INT’L L. PROC. 193, 194 (2000) (“In deciding that it would take the form of a framework convention, member states have indicated that the legislative process to be used will be of a continuing nature.”); Luk Joossens, *From Public Health to International Law: Possible Protocols for Inclusion in the Framework Convention on Tobacco Control*, 78 BULL. WORLD HEALTH ORG. 930, 930-31 (2000).

4. At the time of publication, an Open-Ended Intergovernmental Working Group for FCTC is currently negotiating the rules of procedure under which the Conference of the Parties will convene. Once established, the Conference of the Parties, in fulfilling its oversight role for the Convention, “shall keep under regular review the implementation of the Convention and take the decisions necessary to promote its effective implementation and may adopt protocols, annexes and amendments to the Convention . . .” FCTC, *supra* note 1, art. 23; *see also infra* note 52 and accompanying text (discussing the role of the Conference of the Parties in the adoption of “A Protocol on the Treatment of Tobacco Dependence”).

5. Although this Article refers repeatedly to “smoking” and “smoking cessation,” this phraseology is used primarily for rhetorical convenience and is not intended to exclude the use of cessation interventions for other forms of tobacco.

tobacco, the FCTC fails to place firm mandates on states to address clinical smoking cessation, thus abandoning the millions already addicted to nicotine and vulnerable to the morbidity and mortality of smoking. In Part I, this Article begins by examining the scope and harms of the tobacco pandemic, explaining the processes that led states to recognize the magnitude of this global threat and to draft the FCTC. In doing so, this Article highlights the unfulfilled promise of smoking cessation for stemming the tobacco pandemic, critically assessing the FCTC's failure to mandate clinical cessation interventions. Article 14, the only section of the FCTC to address cessation, obviates state responsibility to provide any clinical interventions for those addicted to nicotine. Although the WHO initially proposed a "Protocol on the Treatment of Tobacco Dependence," member states quickly abandoned this regulatory mechanism in favor of the less-obligatory policy recommendations of the FCTC.

Part II argues that such neglect—turning a blind eye to a dangerous and often deadly addiction—violates the international human right to health. After defining the scope of the right to health, Part II analyzes affirmative obligations on states to address smoking cessation pursuant to this right, laying out a hierarchy of resource-dependent options that states might employ in fulfilling their obligations to palliate the effects of the tobacco pandemic. Applying this analysis to the FCTC, Part III proposes that states party to the FCTC reengage a protocol to address nicotine addiction and clinical tobacco cessation interventions. This Article concludes that a FCTC cessation protocol would revitalize the right to health and give states the formalistic tools necessary to curb smoking, prevent disease, and promote public health.

I. SMOKING AND THE FCTC

A. Tobacco and Its Discontents

Countless others have elucidated the enormous public health ramifications of the tobacco pandemic.⁶ Today, over 1.1 billion people worldwide smoke.⁷ Approximately one-quarter of all lifelong smokers will die in their middle age (between the ages of thirty-five and sixty-nine) as a

6. *E.g.*, JUDITH MACKAY & MICHAEL ERIKSEN, *THE TOBACCO ATLAS* (2002); *TOBACCO CONTROL IN DEVELOPING COUNTRIES* (Prabhat Jha & Frank Chaloupka eds., 2000).

7. WORLD BANK, *CURBING THE EPIDEMIC, GOVERNMENTS AND THE ECONOMICS OF TOBACCO CONTROL* 13 (1999).

result of smoking, losing between twenty and twenty-five years of life.⁸ Another quarter of these smokers will die in their latter years as a result of smoking.⁹ Compounding this massive death toll and morbidity is the debilitating effect of passive inhalation of environmental tobacco smoke, so-called “second-hand smoke,” which affects not only the individual smoker but also those family members, coworkers, and others whose lives place them in close proximity to a smoker.¹⁰ Globally, this “quiet pandemic” claims the lives of approximately five million persons per year,¹¹ a figure that will rise to ten million by 2030, with the burden of death increasingly being felt by developing states.¹² With globalization’s dismantling of trade barriers permitting the burgeoning initiation of smoking in unsated developing states—particularly among the children and adolescents of these states—the global death rate from tobacco is expected to increase exponentially, causing approximately 150 million

8. Richard Peto & Alan D. Lopez, *Future Worldwide Health Effects of Current Smoking Patterns*, in CRITICAL ISSUES IN GLOBAL HEALTH 154, 155 exhibit 18.1 (C. Everett Koop et al. eds., 2001).

9. *Id.*

10. MACKAY & ERIKSEN, *supra* note 6, at 34-35 (depicting the myriad harms caused to adults and children by passive smoking of environmental tobacco smoke). In addition to heightened mortality and morbidity of those passively exposed to smokers through environmental tobacco smoke, fetuses may be exposed to smokers through mother to child transmission of nicotine and other chemicals.

11. WORLD HEALTH ORG., THE WORLD HEALTH REPORT: REDUCING RISKS, PROMOTING HEALTHY LIFE (2002). The death toll attributed to tobacco represented 8.8% of global deaths in 2000. Kenji Shibuya et al., *WHO Framework Convention on Tobacco Control: Development of an Evidence Based Global Public Health Treaty*, 327 BRIT. MED. J. 154, 155 (2003).

12. Peto & Lopez, *supra* note 8, at 157 (citing WORLD HEALTH ORG., TOBACCO OR HEALTH: A GLOBAL STATUS REPORT (1997)) (separating the causes of tobacco-related death by region); *International Law and Health, Two Approaches: The World Health Organization’s Tobacco Initiative and International Drug Controls, Summary of Remarks by Allyn L. Taylor*, 94 AM. SOC’Y INT’L L. PROC. 193, 193-94 (2000) (“At current levels of consumption, the tobacco epidemic is expected to kill up to ten million people per year by 2030, with 70 percent of these deaths occurring in developing nations.”) (emphasis added). For reasons underlying the tobacco industry’s influence in the developing world and the birth of the “global smoker,” see Jeff Collin, *Think Global, Smoke Local: Transnational Tobacco Companies and Cognitive Globalization*, in HEALTH IMPACTS OF GLOBALIZATION: TOWARDS GLOBAL GOVERNANCE 61, 71-75 (Kelley Lee ed., 2003) (noting that “TTCs [transnational tobacco companies] have sought to present the rise of cigarette sales as an indicator of modernity and symbol of economic progress within low-income countries”); and Melissa E. Crow, *Smokescreens and State Responsibility: Using Human Rights Strategies To Promote Global Tobacco Control*, 29 YALE J. INT’L L. 209, 210-11 (2004).

deaths in the next twenty-five years and one billion total deaths throughout the twenty-first century.¹³ Combined with the detrimental micro- and macroeconomic consequences of tobacco cultivation and cigarette consumption¹⁴—exploiting entire populations in vicious cycles of poverty, malnutrition, and death—tobacco use has become a threat to the prosperity of the state itself.¹⁵ This threat to global public health and human security, projected soon to become the world's leading cause of avoidable death, cannot conscionably be ignored.

B. Importance of Cessation

There is clear evidence that smoking cessation interventions can decrease the risk of premature morbidity and mortality.¹⁶ In fact, the earlier a smoker quits, the more dramatic this decrease in risk of premature sickness and death.¹⁷ Considering the pervasiveness of the tobacco pandemic, quitting smoking is the most efficient means of saving lives—“offer[ing] the only realistic way in which widespread changes in smoking status can prevent large numbers of tobacco deaths over the next

13. Jeff Collin et al., *The Framework Convention on Tobacco Control: The Politics of Global Health Governance*, 23 *THIRD WORLD Q.* 265, 273 (2002) (“Trade liberalisation has led to increased consumption of tobacco, but while it has no substantive effect on higher income countries, it has a large and significant impact on smoking in low-income countries and a significant, if smaller, impact on middle-income countries.”); Peto & Lopez, *supra* note 8, at 158, 160; *see also* WORLD HEALTH ORG., MAYO REPORT: ADDRESSING THE WORLDWIDE TOBACCO EPIDEMIC THROUGH EFFECTIVE, EVIDENCE-BASED TREATMENT (1999), http://www.who.int/tobacco/health_impact/mayo/en/ [hereinafter MAYO REPORT] (“By 2020, smoking will cause about one in three of all adult deaths, up from one in six adult deaths in 1990.” (quoting Dr. Gro Harlem Brundtland)).

14. Martin Bobak et al., *Poverty and Smoking*, in TOBACCO CONTROL IN DEVELOPING COUNTRIES, *supra* note 6, at 41, 56-58 (analyzing the socio-economic gradient in smoking to determine causal processes underlying the correlation between poverty and smoking).

15. *See generally Tobacco and Poverty: A Vicious Circle*, WHO Doc. WHO/NMH/TFI/04.01 (2004) (noting the impoverishing effects of tobacco on individuals and states).

16. *See generally* U.S. DEP’T OF HEALTH & HUMAN SERVS., THE HEALTH BENEFITS OF SMOKING CESSATION: A REPORT OF THE SURGEON GENERAL (1990).

17. WORLD BANK, *supra* note 7, at 27 (noting that the earlier a smoker quits, the better his or her probability of survival); WORLD HEALTH ORG., MONOGRAPH: ADVANCING KNOWLEDGE ON REGULATING TOBACCO PRODUCTS 10 (2001) (noting the dose-response relationship between tobacco use and most tobacco-related causes of death); Jack E. Henningfield & John Slade, *Tobacco-Dependence Medications: Public Health and Regulatory Issues*, 53 *FOOD & DRUG L.J.* 75, 79 (1998).

half century.”¹⁸ Indeed, the WHO has recognized the importance of cessation, noting that if “the goal for smoking control is a reduction of smoking-related mortality, special emphasis must be given to maximizing the number of individuals who quit smoking.”¹⁹ As compared with prevention, which averts death only in the distant future, cessation offers the promise of lowering morbidity and mortality in the short-term, validating tobacco control programs with tangible, life-saving results.²⁰

Overall, some seventy-five to eighty percent of smokers want to stop smoking.²¹ Yet quitting is not easily accomplished. Although approximately one-third of smokers worldwide attempt to quit each year (often without knowledge or use of cessation interventions²²), a mere one to three percent of all those attempting to quit remain tobacco-free even six months later.²³

18. Peto & Lopez, *supra* note 8, at 158; *id.* at 159 exhibit 18.2 (noting that “halving global cigarette consumption per adult by the year 2020 . . . would prevent about one-third of the tobacco deaths in 2020 and would almost halve tobacco deaths in the second quarter of the century”).

19. See WORLD HEALTH ORG., GUIDELINES FOR CONTROLLING AND MONITORING THE TOBACCO EPIDEMIC 18 (1998). Following commencement of the FCTC, WHO “urged governments to include anti-addiction treatments as part of comprehensive tobacco control programs.” *WHO Calls for War on Tobacco To Include Anti-Addiction Treatment*, DRUG WEEK, Aug. 29, 2003, at 263.

20. Vera Luiza da Costa e Silva, *Foreword*, WORLD HEALTH ORG., POLICY RECOMMENDATIONS FOR SMOKING CESSATION AND TREATMENT OF TOBACCO DEPENDENCE, at x (2003) (“Evidence has shown that cessation is the only intervention with the potential to reduce tobacco-related mortality in the short- and medium-term.”); Peto & Lopez, *supra* note 8, at 156 (noting that “the number of young adults who are taking up smoking around the year 2000 will strongly influence the number of deaths from tobacco around the year 2050 (and beyond)”); Martin Raw, *Fighting Tobacco Dependence in Europe*, 7 NATURE MED. 13, 14 (2001) (explaining that “adolescents suffer smoking related disease 40-60 years in the future, whereas for middle-aged adults it is 10-30 year [sic] away or less”); Kenneth E. Warner, *Reducing Harm to Smokers: Methods, Their Effectiveness, and the Role of Policy*, in REGULATING TOBACCO 111, 111-12 (Robert L. Rabin & Stephen D. Sugarman eds., 2001) (“Any reduction in tobacco-produced mortality over the next three decades necessarily must come from reductions in the risks current smokers face.”).

21. Michael Russell, *Help and Treatment for Addicted Smokers*, 16 WORLD HEALTH F. 19, 20 (1995).

22. Costa e Silva, *supra* note 20, at xvi. Knowledge of the risks of smoking and benefits of cessation significantly increases smokers’ efforts to quit. *Id.*

23. ROYAL COLLEGE OF PHYSICIANS, NICOTINE ADDICTION IN BRITAIN: A REPORT OF THE TOBACCO ADVISORY GROUP OF THE ROYAL COLLEGE OF PHYSICIANS (2000); PAN AM. HEALTH ORG., NICOTINE ADDICTION AND SMOKING CESSATION (1999), http://165.158.1.110/english/hpp/wntd_policy.htm. For a country-by-country assessment of the demographics of

Among those who quit temporarily, “the majority persist in tobacco use for many years and typically cycle through multiple periods of relapse and remission.”²⁴ Considering tobacco’s pharmacologically addictive qualities and the tobacco industry’s psychologically manipulative advertising (totaling well over \$10 billion per year²⁵), it comes as no surprise that the rate of unaided smoking cessation, burdened by a chronically high rate of relapse, remains low.²⁶ Because of the addictive effects of nicotine, regulatory reliance on education of the risks alone cannot be successful for many smokers. Clinical cessation interventions, when combined with other forms of institutional support, can significantly increase the number of attempts to quit and the likelihood of success at each attempt, dramatically improving the chances of breaking entrenched tobacco dependence.²⁷

Despite the proven efficacy and cost-effectiveness of cessation

quitting, see MACKAY & ERIKSEN, *supra* note 6, at 94-101 (noting, where available, the percentages of people who had quit smoking in a given country by 2002).

24. The Tobacco Use and Dependence Clinical Practice Guideline Panel, Staff, and Consortium Representatives, *A Clinical Practice Guideline for Treating Tobacco Use and Dependence: A US Public Health Service Report*, 283 JAMA 3244, 3246 (2000).

25. MACKAY & ERIKSEN, *supra* note 6, at 58 (“While there is no reliable estimate of global cigarette marketing expenditures, it is clearly in the tens of billions of US dollars a year. In the USA alone over \$10 billion is spent a year on marketing cigarettes, and this at a time when advertising is prohibited on television and radio, when there are limitations on certain types of outdoor advertising and sponsorship, and when cigarette sales are falling.”).

26. EUR. P’SHP TO REDUCE TOBACCO DEPENDENCE, WHO EVIDENCE BASED RECOMMENDATIONS ON THE TREATMENT OF TOBACCO DEPENDENCE 3 (2001).

27. David P. Hopkins et al., *Reviews of Evidence Regarding Interventions To Reduce Tobacco Use and Exposure to Environmental Tobacco Smoke*, 20 AM. J. PREVENTATIVE MED. 16, 33-40 (2001) (surveying success rates for various combinations of clinical cessation interventions); Thomas E. Novotny et al., *Smoking Cessation and Nicotine-Replacement Therapies*, in TOBACCO CONTROL IN DEVELOPING COUNTRIES, *supra* note 6, at 287, 288 (noting that “[t]he availability of effective cessation therapy might also help move smokers from pre-contemplation and contemplation stages to action and maintenance” increasing the number of quit attempts); The Tobacco Use and Dependence Clinical Practice Guideline Panel, Staff, and Consortium Representatives, *supra* note 24, at 3246 (“Although only about 7% of smokers achieve long-term success when trying to quit on their own, updated guideline analyses revealed that success rates can be increased to 15% to 30% by using guideline-recommended treatments.”). For example, clinical smoking cessation efforts, combined with changing social norms, have helped to lower the prevalence of smoking in the United States from forty-seven percent in 1965 to twenty-two percent in 1999. WORLD HEALTH ORG., *supra* note 20, at 1 (citing NAT’L CANCER INST., POPULATION-BASED SMOKING CESSATION: PROCEEDINGS OF A CONFERENCE ON WHAT WORKS TO INFLUENCE CESSATION IN THE GENERAL POPULATION (2000)).

interventions,²⁸ a paucity of cessation programs exist at the state level,²⁹ as “smoking cessation is not seen as a public health priority” by national politicians.³⁰ Vera da Costa e Silva, the WHO’s Director for Tobacco Control, has lamented that “the public health sector in many countries is not investing in smoking-cessation services, and in most countries only limited steps have been taken to provide treatment, train health care providers, and release financial resources.”³¹ Although tobacco cessation programs are cost-effective and health benefits are apparent in the short-term, states nevertheless resist these interventions because they still bear some initial cost, the benefits of which are not *immediately* demonstrable.³² Without states engaging smoking cessation as a legislative priority, those

28. EUR. P’SHIP TO REDUCE TOBACCO DEPENDENCE, *supra* note 26, at 6 (“Because tobacco dependence treatment is so cost effective, it should be provided by public and private health care systems.”); WORLD BANK, *supra* note 7, at 77-78; Raymond Niaura & David B. Abrams, *Smoking Cessation: Progress, Priorities, and Prospectus*, 70 J. CONSULTING & CLINICAL PSYCHOL. 494, 502 (2002). For a discussion of the efficacy and cost-effectiveness of specific cessation interventions, see *infra* notes 123-167 and accompanying text.

29. EUR. P’SHIP TO REDUCE TOBACCO DEPENDENCE, *supra* note 26, at 4 (noting the lack of tobacco support and treatment programs in European health care systems); see also Raw, *supra* note 20, at 13 (noting the difficulty of finding funds to work internationally in smoking cessation); Costa e Silva, *supra* note 20, at xi (“[D]espite the availability of cost-effective treatment for tobacco dependence, the public health sector in many countries[] is not investing in smoking-cessation services, nor in the development of an infrastructure that will motivate smokers to quit and support them in doing so.”).

30. Costa e Silva, *supra* note 20, at xi. Even in the United States, tobacco control efforts had been faulted for their emphasis on political expedience rather than public health. Clive Bates et al., *The Future of Tobacco Product Regulation and Labeling in Europe: Implications for the Forthcoming European Union Directive*, 8 TOBACCO CONTROL 225, 230 (1999) (noting that the U.S. Federal Trade Commission’s regulation of tobacco content is flawed, legitimizing the myth of low-tar cigarettes).

31. WHO Calls for War on Tobacco To Include Anti-Addiction Treatment, *supra* note 19, at 16.

32. Gro Harlem Brundtland, *Achieving Worldwide Tobacco Control*, 284 JAMA 750, 750 (2000) (lamenting the limited impact of tobacco control by noting “that action is occurring too late, partially because policy makers have not been motivated to intervene in time”); Niaura & Abrams, *supra* note 28, at 502; cf. Collin et al., *supra* note 13, at 267 (“The paucity of regulation may reflect the importance of domestic interests, particularly in the small number of national economies that are heavily dependent on tobacco production.” (citing TOBACCO CONTROL IN DEVELOPING COUNTRIES, *supra* note 6)). In addition to the lack of immediate political reward for actions to reduce the prevalence of smoking, national politicians are also besieged by relentless attempts by transnational tobacco corporations to manipulate individual national policies. See *infra* notes 196-198 and accompanying text (analyzing the influence of transnational tobacco corporations at the national level).

who need programs to help them quit cannot obtain the institutional support they need.

C. Exposing the Silent Pandemic—The Framework Convention on Tobacco Control

Recognizing the catastrophic impact of smoking on global public health, the World Health Assembly, representing all WHO member states, adopted Resolution 49.17 on May 26, 1996.³³ The resolution called upon the WHO “to initiate the development of a framework convention [on tobacco control] in accordance with article 19 of the WHO Constitution.”³⁴ Although Resolution 49.17 met with substantial resistance both inside and outside the WHO,³⁵ international tobacco control took on renewed importance after the World Health Assembly elected Dr. Gro Harlem Brundtland, a staunch tobacco control advocate, as Director-General of the WHO.³⁶ Dr. Brundtland’s commitment to tobacco control was embodied in

33. *International Framework Convention for Tobacco Control*, WHA Res. 49.17, World Health Assembly, 49th Ass., 6th plen. mtg., WHO Doc. A49/VR/6 (May 25, 1996). The seeds for the FCTC were planted through conversations between Ruth Roemer, a professor of health law and advocate of legislative approaches to tobacco control, and Allyn Taylor, then a doctoral student at Columbia Law School. At that time, “[t]he idea of a convention that utilized international law to further public health was new.” Judith Mackay, *The Making of a Convention on Tobacco Control*, 81 BULL. WORLD HEALTH ORG. 551, 551 (2003). At no point in the advocacy for the FCTC did either Roemer or Taylor address the issue of cessation interventions for current smokers. See generally Allyn L. Taylor, *An International Regulatory Strategy for Global Tobacco Control*, 21 YALE J. INT’L L. 257 (1996) (advocating an international regulatory strategy for tobacco control, including fiscal measures, regulation of advertising, regulation of smoking in public places, disincentives for youth tobacco consumption, and mandatory health education).

34. *International Framework Convention for Tobacco Control*, *supra* note 33. Prior to the adoption of Resolution 49.17, the WHO had commissioned an independent feasibility study for addressing tobacco control at the international level. *An International Strategy for Tobacco Control*, WHA Res. 48.11, World Health Assembly, 48th Ass., 12th plen. mtg., Annex 1, WHO Doc. WHA48/VR/12 (May 12, 1995). As part of this feasibility study, the WHO contracted with Roemer and Taylor, out of which WHO published its *International Strategy for Tobacco Control*. See ALLYN L. TAYLOR & RUTH ROEMER, *INTERNATIONAL STRATEGY FOR TOBACCO CONTROL* 17 (1996).

35. Mackay, *supra* note 33, at 551.

36. Gro Harlem Brundtland, Director-General Elect, The World Health Organization, Speech to the Fifty-First World Health Assembly, at 7, WHO Doc. A51/DIV/6 (May 13, 1998) (noting, in her opening speech to the World Health Assembly, that the WHO would take a leading role in “a broad alliance against tobacco, calling on a wide range of partners to halt the relentless increase in global tobacco consumption”). Prior to her ascension to

the creation of the WHO's international campaign against tobacco, the Tobacco Free Initiative.³⁷ By May 24, 1999, the World Health Assembly's 191 members had unanimously agreed to establish a Framework Convention on Tobacco Control, despite the fact that the WHO had never before drafted a binding international treaty in its fifty-five year history.³⁸ Following the establishment of, and two extensive drafting sessions by, the WHO's Working Group and Intergovernmental Negotiating Body,³⁹ the World Health Assembly unanimously adopted the FCTC on May 21, 2003,⁴⁰

the Director-General position at WHO, Dr. Brundtland had served three terms as Prime Minister of Norway. Gavin Yamey, *WHO in 2002: Have the Latest Reforms Reversed WHO's Decline*, 325 BRIT. MED. J. 1107, 1107 (2002) (assessing the effectiveness of Dr. Brundtland's tenure at WHO).

37. See MAYO REPORT, *supra* note 13 (highlighting that the WHO launched its "Tobacco Free Initiative," the organizational precursor of the FCTC, on the day Dr. Brundtland took office).

38. Although Article 19 of the WHO Constitution authorizes WHO to adopt conventions or agreements, the WHO had never before used this power. *Tobacco Free Initiative, Report by the Director General*, WHA Res. 52.18, World Health Assembly, 52nd Ass., 9th plen. mtg., Agenda Item 13, WHO Doc. A52/7 (Mar. 18, 1999). Because of the ineffectiveness of the WHO, based upon its past reluctance to legislate its health strategies, and the "modest level of global commitment to tobacco control," various commentators recommended that any WHO attempts to address the international tobacco pandemic involve only incremental standard setting. *E.g.*, Allyn Lise Taylor, *Making the World Health Organization Work: A Legal Framework for Universal Access to the Conditions for Health*, 18 AM. J.L. & MED. 301, 303 (1992) (noting that the "WHO's traditional reluctance to utilize law and legal institutions to facilitate its health strategies is largely attributable to the internal dynamics and politics of the organization itself"). Now, in the wake of the FCTC, these same scholars look to the WHO's agenda-setting capacity, pushing it to leverage its role as a representative of the community of states to shape state behavior in resolving other issues of public health impervious to solution at the national level. See Allyn Taylor, *Global Health Governance and International Law*, 25 WHITTIER L. REV. 253, 261-62 (2003) ("I believe that the FCTC may signal a turning point—a new era in international health cooperation. The WHO's unconventional consideration of the role of international law and institutions in promoting public health policies suggests an expansion of the organization's traditional scientific, technical approaches to public health, and perhaps, an evolution of its traditional culture.").

39. The World Health Assembly established the Working Group to establish the FCTC's technical foundation and the Intergovernmental Negotiating Body to undertake the drafting components of the FCTC. See *Towards a WHO Framework Convention on Tobacco Control*, WHA Res. 52.18, World Health Assembly, 52nd Ass., 9th plen. mtg., Agenda Item 13, WHO Doc. A52/7 (Mar. 18, 1999) (establishing the Working Group and Intergovernmental Negotiating Body).

40. The World Health Assembly, encompassing delegates of all member states and

shifting implementation of convention provisions to the states.⁴¹ By June 29, 2004, the day the FCTC closed for signature, 155 states had signed the FCTC, with ten states having already ratified it.

The sheer adoption of the FCTC—enabling states to overcome domestic and collective action problems to achieve a common good—should be seen as a great leap forward for tobacco control. Prior to the advent of the FCTC, only select Western states had enacted comprehensive tobacco control efforts.⁴² While critical of the FCTC’s approach, the author cannot and will not minimize the monumental importance of this effort, which overcame significant tobacco industry resistance to become a valuable precedent for national and global solutions to safeguard public health and eradicate disease.

Despite its many successes, the FCTC fails to place affirmative obligations on states vis-à-vis clinical smoking cessation. The Convention focuses instead on the globalized aspects of tobacco supply and indirect limitations on global demand. Through broad regulations on tobacco advertising, warning labels, taxation, and smuggling, the Convention seeks to change the social environment for smoking without actively changing individual behavior.⁴³ That is, the FCTC discourages consumption without encouraging cessation. As a result, the FCTC—the first treaty drafted explicitly to protect public health—has been criticized for lacking a firm

meeting at annual or special sessions, acts to adopt WHO conventions or agreements by a two-third majority vote, with each member state having one vote in the Assembly. World Health Organization Constitution, July 22, 1946, arts. 59, 60, 62 Stat. 2679, 14 U.N.T.S. 185, *reprinted in* WORLD HEALTH ORG., BASIC DOCUMENTS (40th ed. 1994) [hereinafter WHO Constitution].

41. Shibuya et al., *supra* note 11, at 154 (noting that, following ratification of the FCTC, “further efforts are needed to establish national capacities to set the foundation for the later implementation of the treaty, to negotiate the protocols on specific subjects within the framework, and to implement effective interventions to reduce tobacco consumption globally”). In addition to states, over one hundred eighty civil society organizations from over seventy states, linked together through the Framework Convention Alliance, served a crucial role in developing the FCTC and will continue to serve as a resource in implementing and monitoring the FCTC. Collin et al., *supra* note 13, at 274, 278; Clive Bates, *Developing Countries Take the Lead on WHO Convention*, 10 BRIT. MED. J. 209, 209 (2001); R. Hammond & M. Assunta, Editorial, *The Framework Convention on Tobacco Control: Promising Start, Uncertain Future*, 12 TOBACCO CONTROL 241, 241 (2003); Crow, *supra* note 12, at 217.

42. Taylor, *supra* note 33, at 268 (suggesting national tobacco control strategies).

43. See FCTC, *supra* note 1, art. 13 (advertising); *id.* art. 10, 11 (warning labels); *id.* art. 6 (taxation); *id.* art. 15 (smuggling).

basis in public health.⁴⁴

Although the FCTC's Preamble recognizes "that cigarettes and some other products containing tobacco are highly engineered so as to create and maintain dependence . . . and that tobacco dependence is separately classified as a disorder in major international classifications of diseases,"⁴⁵ the Preamble uses neither the word "nicotine" nor the word "addiction," two words that form the public health basis of tobacco control.⁴⁶ Overall, the FCTC focuses on initiation of smoking but not cessation. Article 14, the only portion of the FCTC devoted to cessation, reads:

Demand reduction measures concerning tobacco dependence and cessation

1. Each Party shall develop and disseminate appropriate, comprehensive and integrated guidelines based on scientific evidence and best practices, taking into account national circumstances and priorities, and shall take effective measures to promote cessation of tobacco use and adequate treatment for tobacco dependence.

2. Towards this end, each Party shall endeavour to:

(a) design and implement effective programmes aimed at promoting the cessation of tobacco use, in such locations as educational institutions, health care facilities, workplaces and sporting environments;

(b) include diagnosis and treatment of tobacco dependence and counselling services on cessation of tobacco use in national health and education programmes, plans and strategies, with the participation of health workers, community workers and social workers as appropriate;

(c) establish in health care facilities and rehabilitation centres programmes for diagnosing, counselling, preventing and treating

44. See, e.g., Crystal H. Williamson, *Clearing the Smoke: Addressing the Tobacco Issue as an International Body*, 20 PENN ST. INT'L L. REV. 587, 611 (2002) (noting that "participants [in FCTC drafting] themselves pointed out (and attempted to regulate) some matters that had decidedly more to do with trade than with health concerns").

45. FCTC, *supra* note 1, pmbl.

46. In 1964, the WHO Expert Committee on Drug Dependence defined "dependence" as "a state, psychic and sometimes also physical, resulting from the interaction between a living organism and a drug, characterized by behavioural and other responses that always include a compulsion to take the drug on a continuous or periodic basis in order to experience its psychic effects, and sometimes to avoid the discomfort of its absence." WORLD HEALTH ORG., TECHNICAL REPORT SERIES NO. 407, WHO EXPERT COMMITTEE ON DRUG DEPENDENCE, SIXTEENTH REPORT 6 (1969); see also WORLD HEALTH ORG., MANUAL OF THE INTERNATIONAL STATISTICAL CLASSIFICATION OF DISEASES, INJURIES, AND CAUSES OF DEATH 198-99 (1977). For a description of the rhetorical evolution of "dependence," see Harold Kalant, *Nicotine as an Addictive Substance*, in NICOTINE AND PUBLIC HEALTH 117, 119-22 (Roberta Ferrence et al. eds., 2000).

tobacco dependence; and

(d) collaborate with other Parties to facilitate accessibility and affordability for treatment of tobacco dependence including pharmaceutical products pursuant to Article 22. Such products and their constituents may include medicines, products used to administer medicines and diagnostics when appropriate.⁴⁷

Even here, the use of nonobligatory language—e.g., “endeavour to” following “shall” in the second paragraph—trivializes the role of cessation in a comprehensive tobacco control program. The use of hortatory rather than legal statements, soft rather than hard law, denies Article 14 of any self-executing requirements, leaving treaty implementation solely at the discretion of individual states.⁴⁸ This lack of mandatory provisions, compounded by weak implementation mechanisms and state reporting requirements,⁴⁹ provides no incentive for change in state cessation policy. Thus, while the FCTC’s program initiatives may buttress smokers’ psychological motivations to quit through, *inter alia*, health education programs, cigarette taxation, and smoke-free air laws, it commits states to do relatively little to reduce the psychological and addiction-related barriers to smoking cessation.⁵⁰

Although the WHO had previously offered paeans to the importance of clinical cessation programs in tobacco control policy, member states did little to act on this belief in drafting the FCTC. The First Meeting of the

47. FCTC, *supra* note 1, art. 14.

48. See Thomas Michael McDonnell, *Defensively Invoking Treaties in American Courts—Jurisdictional Challenges Under the U.N. Drug Trafficking Convention by Foreign Defendants Kidnapped Abroad by U.S. Agents*, 37 WM. & MARY L. REV. 1401, 1475 n.352 (noting that the presence of “shall endeavour to” language in extradition treaties denies relevant provisions of self-executing status); Annie Petsonk, *Challenges to International Governance: International Land-Use Law*, 87 AM. SOC’Y INT’L L. PROC. 488, 498 (1993) (remarks by Ralph Osterwoldt) (recognizing a distinction between “hard law, by which I mean binding obligations set out in international treaties and agreements, which typically provide that ‘states shall do X,’ and in ‘soft law,’ meaning guidelines, principles and hortatory statements contained in conventions, including requirements that states shall ‘endeavor to cooperate, report, exchange information’”).

49. Crow, *supra* note 12, at 218-20. The author notes that these weaknesses are not confined to Article 14 but serve to stymie the enforcement of many provisions within the FCTC.

50. See WORLD HEALTH ORG., *supra* note 19, at 19 (noting that “smoking control policies should contain both activities to strengthen smokers’ motivation to quit (health education, public information, price policies, smoke-free policies, behavioural treatments, etc.) and activities to reduce dependence-related difficulties for smokers to quit (behavioural and pharmacological treatment)”) (emphasis added).

Working Group on the WHO Framework Convention on Tobacco Control, which convened in October 1999, agreed that the FCTC should focus on tobacco demand reduction strategies, including the treatment of tobacco dependence.⁵¹ In 2000, the Second Meeting of the Working Group on the WHO Framework Convention on Tobacco Control expanded this cessation mandate, with the WHO's Tobacco Free Initiative submitting "Possible Subjects of Initial Protocols" that included "A Protocol on the Treatment of Tobacco Dependence" (Proposed Dependence Protocol), reprinted herein as an annex to the present Article.⁵² However, by 2003, suggestions for future protocols by the Intergovernmental Negotiating Body on the WHO Framework Convention on Tobacco Control addressed only the subjects of "advertising, promotion and sponsorship; tobacco-product regulation; illicit trade; and liability,"⁵³ leaving out the issues of tobacco dependence and cessation.

It is unclear exactly why member states abandoned the Working Group's Proposed Dependence Protocol without serious consideration, although many disparate factors likely influenced their decision. First, during preliminary negotiations, when the success of the FCTC remained in doubt, many nongovernmental organizations and states, seeking international consensus over legislative comprehensiveness,⁵⁴ criticized the

51. See Allyn L. Taylor & Douglas W. Bettcher, *WHO Framework Convention on Tobacco Control: A Global "Good" for Public Health*, 78 BULL. WORLD HEALTH ORG. 920, 925 (2000) (citing World Health Org., WHO Framework Convention on Tobacco Control: Report of the First Meeting of the Working Group (1999) (unpublished manuscript)); Williamson, *supra* note 44, at 610 (noting, during the early drafting stages of the FCTC, that "[t]reatment is to become part of a broad strategy of the WHO, making cessation a key component of primary health care").

52. *Possible Subjects of Initial Protocols: Elaboration of Technical Components of Three Possible Protocols*, Working Group on the WHO Framework Convention on Tobacco Control, 2d mtg., Agenda Item 6, WHO Doc. A/FCTC/WG2/4 (Feb. 15, 2000) [hereinafter Second Meeting of the Working Group] (noting that "the treatment of tobacco dependence was generally supported as an important demand-reduction strategy to be addressed in a protocol"). Although WHO's Tobacco Free Initiative, as Interim Secretariat, developed the Proposed Dependence Protocol, the FCTC requires that any protocols for adoption be proposed by state parties at least six months prior to a session of the Conference of the Parties, which can only occur once the FCTC has entered into force (on the ninetieth day following the date of ratification by the fortieth state). FCTC, *supra* note 1, art. 33.

53. Intergovernmental Negotiating Body, *supra* note 3.

54. See Crow, *supra* note 12, at 213 ("Due to the uncertain political viability of obtaining consensus on a conventional treaty structure, WHO's governing body, the World Health Assembly (WHA), opted for a framework convention, which can be supplemented by specialized protocols." (footnotes omitted)).

protocol as legislative overreaching. Throughout the convention process, those involved in drafting the FCTC focused their legislative will on the international components of the tobacco epidemic,⁵⁵ often at the expense of costly domestic programs like cessation interventions. Further, many viewed a cessation provision as too great a boon for transnational pharmaceutical corporations, long derided for their close ties to the WHO, which would stand to gain enormous financial profit from the widespread distribution of smoking cessation products.⁵⁶ To alleviate such conflicts of interest, pharmaceutical corporations were not invited to the plenary drafting sessions of the FCTC, and lobbying for cessation was viewed with skepticism. Finally, many of the compromises reached by the WHO's Working Group and Intergovernmental Negotiating Body allow states to postpone economically painful decisions until a later date. For example, states financially dependent on tobacco exports face the short-term prospect of agricultural losses if cessation interventions are successful. Foregoing cessation programs minimizes the immediate impact on agricultural exports, alleviating the prospect of state public health

55. See PHYSICIANS FOR A SMOKE-FREE CANADA, COMMENTARY ON WORLD HEALTH ORGANIZATION PROVISIONAL TEXTS OF PROPOSED DRAFT ELEMENTS FOR A WHO FRAMEWORK CONVENTION ON TOBACCO CONTROL 15 (2000), http://www.smoke-free.ca/pdf_1/commentsondraftfctc.PDF ("This protocol contains not a single measure that is international in character. In fact, it contains some measures that are potentially end-runs around existing national drug regulatory mechanisms It is recommended that this draft protocol be dropped entirely from further consideration."); Action on Smoking and Health, ASH Briefing for the First Negotiations (Oct. 2000), http://www.ash.org.uk/html/international/html/ashfctcposition.html#_Toc496178643 ("In our view, this [Proposed Dependence Protocol] can only be a general 'plan and report' obligation, with a number of (strictly optional) measures that could be taken. Detail might be developed in the technical bodies. *There is therefore no need for a protocol.*") (emphasis in original); Framework Convention Alliance, Comments on the Chair's Text of a FCTC Joint New Zealand NGO Submission (Mar. 2001), <http://fctc.org/archives/INB2nzngo.shtml> ("We consider that personal treatment issues, ie treatment of tobacco dependence, need not have their own set of provisions but be included as a part of tobacco control programmes.").

56. See Collin et al., *supra* note 13, at 276-77 (noting pharmaceutical consortia interested in advising WHO on tobacco control); see also Raw, *supra* note 20, at 13 (noting sponsorship of the WHO European Partnership Project to Reduce Tobacco Dependence by cessation product distributors GlaxoWellcome, Novartis, Pharmacia, and SmithKline Beecham). In addition, multinational pharmaceutical corporations have invested heavily in supporting the academic underpinnings of pharmacological treatment for nicotine addiction. *E.g.*, INTERVENTIONS FOR SMOKERS: AN INTERNATIONAL PERSPECTIVE, at ix (Robyn Richmond ed., 1994) (thanking "Marion Merrell Dow Pharmaceuticals in the United States for generous support of this book").

ministries being overruled by finance ministries. For these and other reasons, states never seriously considered cessation interventions through the FCTC, viewing such efforts as a quixotic undertaking foisted upon state delegates by the WHO Secretariat.

Regardless of the precise reasons, the FCTC has effectively abandoned those addicted to tobacco. Even when the WHO has attempted to develop evidence-based policy recommendations to help states implement practical cessation interventions in accordance with adoption of the FCTC,⁵⁷ states have shown little interest in establishing such smoking cessation policy in the absence of strong normative consensus on the importance of cessation.⁵⁸ By failing to emphasize cessation interventions, member states have denied life-saving treatments to millions of smokers, acting in contravention of smokers' human right to health.

II. THE RIGHT TO HEALTH: A HUMAN RIGHTS APPROACH TO ARTICLE 14

A. An Introduction to the Right to Health

An individual's right to health is recognized as a fundamental international human right. Founded upon the non-derogable right to life,⁵⁹ the Universal Declaration of Human Rights (UDHR)⁶⁰ affirms in

57. Costa e Silva, *supra* note 20, at ix-x ("Treatment of tobacco dependence needs . . . to be part of a comprehensive tobacco-control policy along with measures such as taxation and price policies, advertising restrictions, dissemination of information and protection of non-smokers through the creation of smoke-free public places.").

58. Although the WHO Tobacco Free Initiative held a comparative strategy development meeting in 2002 "to explore and recommend potential avenues for progress in the areas of smoking cessation and treatment of tobacco dependence," this meeting garnered only thirty-one participants, with country representatives from only Brazil, Canada, Germany, Hong Kong, the Russian Federation, Seychelles, Thailand, the Philippines, Venezuela, and Qatar. Costa e Silva, *supra* note 20, at xii.

59. Universal Declaration of Human Rights, G.A. Res. 217A(III), art. 3, at 71-72, U.N. Doc. A/810 (1948) ("Everyone has the right to life, liberty and security of person.") [hereinafter UDHR]; Virginia A. Leary, *Implications of a Right to Health*, in HUMAN RIGHTS IN THE TWENTY-FIRST CENTURY 481, 487 (Kathleen E. Mahoney & Paul Mohoney eds., 1993) ("It does not strain imagination to consider the 'right to health' as implicit in the right to life.").

60. UDHR, *supra* note 59, art. 3. "Although the UDHR is not a legally binding document, nations (states) have endowed it with great legitimacy through their actions, including its legal and political invocation at the national and international levels." Jonathan M. Mann et al., *Health and Human Rights*, in HEALTH AND HUMAN RIGHTS 7, 9 (Jonathan M. Mann et al. eds., 1999).

Article 25(1) that “[e]veryone has the right to a standard of living adequate for the health and well-being of himself and his family, including . . . medical care and necessary social services”⁶¹ The United Nations legislatively embodied the economic and social parameters of this right in the International Covenant on Economic Social and Cultural Rights (ICESCR), which elaborates the right to health in Article 12.1 to include “the right of everyone to the enjoyment of the highest attainable standard of physical and mental health.”⁶² To achieve the full realization of this right, Article 12.2 of the ICESCR requires states to take affirmative steps necessary for “(b) [t]he improvement of all aspects of environmental and industrial hygiene; (c) [t]he prevention, *treatment*, and control of epidemic, endemic, occupational and other diseases; [and] (d) [t]he creation of conditions which would *assure to all medical service and medical attention* in the event of sickness.”⁶³ Thus, under the plain language of the

61. UDHR, *supra* note 59, art. 25.

62. International Covenant on Economic Social and Cultural Rights, G.A. Res. 2200, U.N. GAOR, 21st Sess., Supp. No. 16, art. 12(1), U.N. Doc. A/6316 (1966) [hereinafter ICESCR]. In addition, the right to life embodied in Article 6 of the International Covenant on Civil and Political Rights (ICCPR) obligates states “to take positive measures to ensure the right to life including steps to reduce infant mortality rates, prevent industrial accidents, and protect the environment.” Cancado Trindade, *Environmental Protection and the Absence of Restrictions on Human Rights*, in HUMAN RIGHTS IN THE TWENTY-FIRST CENTURY, *supra* note 59, at 561, 573. Nonetheless, just a few scholars have attempted to place health care obligations on states through the ICCPR. See, e.g., Crow, *supra* note 12, at 230 (arguing that the U.N. Human Rights Committee, the legal body established to monitor States Parties’ compliance with the ICCPR, should consider the human rights dimensions of tobacco control under, inter alia, the right to life); Alicia Ely Yamin, *Not Just a Tragedy: Access to Medications as a Right Under International Law*, 21 B.U. INT’L L.J. 325, 330-31 (2003) (“Given that medications can be indispensable for life, it is foreseeable that state policies likely to lead directly to diminished physical accessibility and affordability of certain medications will, in effect, deprive people of life.”); Jonathan Wike, Note, *The Marlboro Man in Asia: U.S. Tobacco and Human Rights*, 29 VAND. J. TRANSNAT’L L. 329, 353 (1996).

63. ICESCR, *supra* note 62, art. 12.2 (emphasis added). The Committee on Economic, Social, and Cultural Rights (CESCR), the monitoring and interpreting body for the ICESCR, has specified that Article 12.2’s requirements are included only by way of illustration and are not intended to be an exhaustive list of state obligations. *The Right to the Highest Attainable Standard of Health*, CESCR General Comment 14, U.N. CESCR, 22d Sess., Agenda Item 3, ¶ 13, U.N. Doc. E/C.12/2000/4 (2000) [hereinafter *General Comment 14*] (noting that Article 12.2 “gives specific generic examples of measures arising from the broad definition of the right to health contained in article 12.1”). In addition, scholars have noted that “a State party in which any significant number of individuals is deprived . . . of essential primary health care . . . is, prima facie, failing to discharge its obligations under

ICESCR, the right to health includes a right to health care. Beyond this, the Committee on Economic, Social and Cultural Rights (CESCR), the legal body charged in the ICESCR with drafting official interpretations of and monitoring state compliance with the ICESCR, has found that the reference in Article 12.1 of the Covenant to “the highest attainable standard of physical and mental health” is not confined to a right to health care. On the contrary, the drafting history and the express wording of Article 12.2 acknowledge that “*the right to health embraces a wide range of socio-economic factors that promote conditions in which people can lead a healthy life, and extends to the underlying determinants of health, such as food and nutrition, housing, access to safe and potable water and adequate sanitation, safe and healthy working conditions, and a healthy environment.*”⁶⁴

Further, in the context of elaborating the actions to be taken by states under Article 12.2 (b) through (d), the CESCR has delineated specific state obligations under (1) the right to a healthy natural and workplace environment to “discourage[] the abuse of alcohol, and the use of tobacco, drugs and other harmful substances;”⁶⁵ (2) the right to treatment and control of diseases to “make available relevant technologies;”⁶⁶ and (3) the right to health care facilities, goods, and services to provide “equal and timely access to base preventive, curative, rehabilitative health services and health education . . . appropriate treatment of prevalent diseases . . . [and] the provision of essential drugs.”⁶⁷ The CESCR has found that states bear the responsibility to protect persons from corporate infringements of Article 12, specifically assigning state responsibility for “failure to discourage . . . consumption of tobacco.”⁶⁸

Since the ICESCR entered into force, various other multilateral

the [ICESCR].” DAVID P. FIDLER, *INTERNATIONAL LAW AND INFECTIOUS DISEASES* 189 (1999) (quoting M.C.R. CRAVEN, *INTERNATIONAL COVENANT ON ECONOMIC, SOCIAL AND CULTURAL RIGHTS: A PERSPECTIVE ON ITS DEVELOPMENT* 141 (1995)).

64. *General Comment 14*, *supra* note 63, ¶ 4 (emphasis added). Despite its devastating impact on health, tobacco is conspicuously absent from General Comment 14’s list of examples of underlying determinants of health. *But cf.* Katherine Gorove, *Shifting Norms in International Health Law*, 98 AM. SOC’Y INT’L L. PROC. 18, 21 (2004) (criticizing the CESCR for “pulling out of nowhere an interpretation of what it believes to be the ‘legal obligations’ of states parties to the Covenant with respect to the right to the highest attainable standard of health”).

65. *General Comment 14*, *supra* note 63, ¶ 15.

66. *Id.* ¶ 16.

67. *Id.* ¶ 17.

68. *Id.* ¶ 51.

treaties have given credence to a right to health.⁶⁹ Moreover, individual “[s]tates have long recognized an obligation to protect their population from obvious risks and hazards to their health,”⁷⁰ often embodying this right within their national constitutions.⁷¹ To the degree that consistent

69. Convention on the Rights of the Child, Nov. 20, 1989, art. 24(2)(b), 28 I.L.M. 1457 (requiring states to ensure the provision of necessary medical assistance, with an emphasis on primary health care, to all children); African Charter on Human and Peoples’ Rights, June 27, 1981, art. 16(2), 1520 U.N.T.S. 217, 249 (requiring states to “take the necessary measures to protect the health of their people and to ensure that they receive medical attention when they are sick”); Convention on the Elimination of All Forms of Discrimination Against Women, Dec. 18, 1979, arts. 11.1(f), 12, 1249 U.N.T.S. 13, 18-19; International Convention on the Elimination of All Forms of Racial Discrimination, Mar. 7, 1966, art. 5(e)(iv), 660 U.N.T.S. 195, 222; European Social Charter, Oct. 18, 1961, art. 11(3), 529 U.N.T.S. 89, 104 (obligating states parties “to take appropriate measures designed *inter alia* . . . to prevent as far as possible epidemic, endemic and other diseases”); American Declaration of the Rights and Duties of Man, art. 11, OEA/Ser.L.V/II.82 doc. 6 rev.1, at 17 (1948) (“Every person has the right to the preservation of his health through sanitary and social measures relating to food, clothing, housing and medical care . . .”). Additionally, the obligation of states to protect the public health may be found through treaties protecting environmental health. David P. Fidler, *A Globalized Theory of Public Health Law*, 30 J.L. MED. & ETHICS 150, 156 (2002) (citing David P. Fidler, *Challenges to Humanity’s Health: The Contributions of International Environmental Law to National and Global Public Health*, 31 ENVIRON. L. REP. 10048 (2001)).

In defining the contours of the right to health, the Preamble to the WHO Constitution declares that “the enjoyment of the highest attainable standard of health is one of the fundamental rights of every human being.” It further notes that “governments have a responsibility for the health of their peoples which can be fulfilled only by the provision of adequate health and social measures.” WHO Constitution, *supra* note 40, pmbl. (defining health as “a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity”). Nevertheless, the WHO Constitution is not viewed as anything more than aspirational in defining the right to health. FIDLER, *supra* note 63, at 187 (citing Leary, *supra* note 59, at 489); Lawrence O. Gostin & Lance Gable, *The Human Rights of Persons with Mental Disabilities: A Global Perspective on the Application of Human Rights Principles to Mental Health*, 63 MD. L. REV. 20, 103 (2004) (“[I]f health is, in the World Health Organization’s words, truly “a state of complete physical, mental and social well-being,” then no one can ever achieve it.”).

70. Leary, *supra* note 59, at 486; *see also* Fidler, *supra* note 69, at 156 (“The frequency with which states have used international law for the purpose of protecting and promoting human health speaks not only to states’ legal powers to assure healthy conditions, but also to their respective duties to do so.”).

71. *The Right of Everyone to the Highest Attainable Standard of Physical and Mental Health: Report of the Special Rapporteur, Paul Hunt, submitted in accordance with Commission Resolution 2002/31*, U.N. ESCOR, 59th Sess., Agenda Item 10, ¶ 20, U.N. Doc. E/CN.4/2003/58

state practice under the aforementioned treaties and constitutions comports with a right to health, it has been advanced that these practices, followed out of a sense of legal obligation, have created norms of customary international law,⁷² binding states to uphold the right to health.⁷³ However, since the right to health is consistently set forth in general, aspirational language that describes the ultimate goal but not the “actions that member nations must take,”⁷⁴ the treaty language, and possible customary law deriving therefrom, provides little guidance as to the specific scope of states’ obligations under the right to health.⁷⁵

The right to health remains a principle seeking a consensus. Outside of these sweeping platitudes, what specific rights does the right to health include? While criticized for its ambiguity,⁷⁶ the right to health has been

(2003) (noting that either a right to health or a right to health care is codified in over sixty national constitutions).

72. Rights created through the general multilateral treaties transmute into universally applicable norms of customary international law when supported by widespread state practice upholding those norms. A. D’Amato, *Treaty-Based Rules of Custom*, in INTERNATIONAL LAW ANTHOLOGY 94 (A. D’Amato ed., 1994); see FIDLER, *supra* note 63, at 99 (“Typically, a rule of customary international law emanating from treaty-based practice originates in a multilateral treaty of general scope.” (citing A. D’Amato, *supra*, at 100)). Likewise, multilateral treaties may codify existing custom.

73. See Eleanor D. Kinney, *The International Human Right to Health: What Does This Mean for Our Nation and World?*, 34 IND. L. REV. 1457, 1464 (2001); Patrick L. Wojahn, Comment, *A Conflict of Rights: Intellectual Property Under TRIPS, the Right to Health, and AIDS Drugs*, 6 UCLA J. INT’L L. & FOREIGN AFF. 463, 493-96 (2002). But see Gostin & Gable, *supra* note 69, at 109 (“While the standard advanced by General Comment 14 may eventually become customary international law, the universal acceptance of a broad right to health may nonetheless not occur for an extended period of time.”); Stephen J. Powell, *The Place of Human Rights Law in World Trade Organization Rules*, 16 FLA. J. INT’L L. 219, 229 (2004) (noting the silence of customary international law on the right to health).

74. Taylor, *supra* note 38, at 327. But cf. Gostin & Gable, *supra* note 69, at 101 (noting that “[r]egional instruments provide more detailed right to health provisions that more specifically outline State obligations”).

75. FIDLER, *supra* note 63, at 188 (noting that “the text of [ICESCR] Article 12(2) is too general to provide insight into concrete actions States parties need to take”); David P. Fidler, “Geographical Morality” Revisited: *International Relations, International Law, and the Controversy over Placebo-Controlled HIV Clinical Trials in Developing Countries*, 42 HARV. INT’L L.J. 299, 348 (2001) (“No moral or legal standard exists that gives the right to health universal meaning.”).

76. FIDLER, *supra* note 63, at 197 (“[T]he right to health is an international human right because it appears in treaties, but the right is so broad that it lacks coherent meaning and is qualified by the principle of progressive realization.”); Lawrence Gostin & Jonathan Mann,

interpreted to include, at a minimum, basic provisions of health care necessary to save lives.⁷⁷ In 1978, the WHO International Conference on Primary Health Care issued “Health for All by the Year 2000,” which has come to be called the Alma Ata Declaration, creating a model of state responsibility for universal access to primary health care.⁷⁸ Under the Alma Ata Declaration, the WHO laid out the essential aspects of primary health care, including:

education concerning prevailing health problems and the methods of preventing and controlling them; promotion of food supply and proper nutrition; an adequate supply of safe water and basic sanitation; maternal and child health care, including family planning; immunization against the major infectious diseases; prevention and control of locally endemic diseases; *appropriate treatment of common diseases and injuries; and provision of essential drugs.*⁷⁹

Twenty years later, the WHO followed up this conference with a new health policy, *Health for All in the Twenty-First Century*,⁸⁰ which focused primarily on health care. After reaffirming the essentials of primary health care from the Alma Ata Declaration, *Health for All in the Twenty-First*

Toward the Development of a Human Rights Impact Assessment for the Formulation and Evaluation of Public Health Policies, in *HEALTH AND HUMAN RIGHTS*, *supra* note 60, at 54 (noting that “a human rights concept as the right to health has not been operationally defined”); Virginia Leary, *Concretizing the Right to Health: Tobacco Use as a Human Rights Issue*, in *RENDERING JUSTICE TO THE VULNERABLE* 161, 162 (Fons Coomans et al. eds., 2000) (“The efforts to clarify the right to health have often been either too theoretical or, alternatively, too detailed and unfocused, resulting in the widespread view that the right to health is an elusive concept and difficult to make operational.”). *But see* Yamin, *supra* note 62, at 336 (arguing that “it can no longer be argued that the content of the right to health is unduly vague for implementing legislation or enforcement, or that it sets out merely political aspirations”).

77. *See General Comment 14*, *supra* note 63, ¶ 36 (elaborating the specific state obligations necessary to fulfill the right to health under Article 12 of the ICESCR).

78. WORLD HEALTH ORG., *GLOBAL STRATEGY FOR HEALTH FOR ALL BY THE YEAR 2000* (1981). For an explanation of WHO’s organizational evolution through the “Health for All campaign,” see Taylor, *supra* note 38, at 329, 314-23, 328-32 (noting that, despite its legal capacity to draft legislation, “WHO has been unable to ensure that nations give adequate and appropriate consideration to their obligations pursuant to the right to health”).

79. ROSALIND POLLACK PETCHESKY, *GLOBAL PRESCRIPTIONS: GENDERING HEALTH AND HUMAN RIGHTS* 124 (2003) (quoting WORLD HEALTH ORG., *supra* note 78) (emphasis added).

80. *Health-for-All Policy for the Twenty-First Century*, WHA Res. 51.7, World Health Assembly, 51st Ass., Agenda Item 19, WHO Doc A51/VR/10 (May 16, 1998).

Century drew upon the right to health to recommit states “to strengthening, adapting and reforming, as appropriate, our health systems, including essential public health functions and services, in order to ensure universal access to health services that are based on scientific evidence, of good quality and within affordable limits, and that are sustainable for the future.”⁸¹

Based upon these foregoing sources of international law, it can be concluded that while health care is a necessary component of the right to health, the right to health includes far more specific mandates on states.⁸² In addition to disease prevention, the right to health requires states to address the treatments necessary for health promotion.⁸³ Individuals are entitled to certain “core elements” of the right to health, including the treatment of prevalent diseases, the provision of essential drugs, and safeguards against serious environmental health threats.⁸⁴ In assuring this individual right, states have affirmative obligations to provide universal access to health services and medications and to protect individuals from serious health infringements by third parties.⁸⁵

But most obligations are not absolute. Outside of the core minimum content of the right to health,⁸⁶ states need only take steps toward the “progressive realization” of the right. In accordance with the principle of

81. *Id.* art. III.

82. Mann et al., *supra* note 60, at 8; *see also* BRIGIT C.A. TOEBES, *THE RIGHT TO HEALTH AS A HUMAN RIGHT IN INTERNATIONAL LAW* 17-18 (1999) (comparing a “right to health” with a “right to health care” and finding the former to be more expansive and encompassing the latter).

83. WHO defines health promotion to include “the process of enabling people to increase control over, and to improve, their health.” Ottawa Charter for Health Promotion, Nov. 21, 1986, <http://www.who.dk/policy/ottawa.htm>; *see also* LAWRENCE GOSTIN & ZITA LAZZARINI, *HUMAN RIGHTS AND PUBLIC HEALTH IN THE AIDS PANDEMIC* 29 (1997). At a minimum, a state has a duty “within the limits of its available resources, to *ensure the conditions necessary for the health* of individuals and populations.” *Id.* (emphasis added).

84. TOEBES, *supra* note 82, at 284.

85. *Id.* at 337-38. Like all human rights, one aspect of state obligation under the right to health involves the obligation to protect, which “requires States to take measures that prevent third parties from interfering with article 12 guarantees.” *General Comment 14, supra* note 63, ¶ 33.

86. “In order for a State party to be able to attribute its failure to meet at least its minimum core obligations to a lack of available resources it must demonstrate that every effort has been made to use all resources that are at its disposition in an effort to satisfy, as a matter of priority, those minimum obligations.” *The Nature of States Parties Obligations* (Art. 2, Par. 1), *CESCR General Comment 3*, U.N. CESCR, 5th Sess., ¶ 10, U.N. Doc. E/1991/23 (1990).

progressive realization, legislatively enacted through the ICESCR, a state must take steps to uphold the right to health only “to the maximum of its available resources, with a view to achieving progressively the full realization of the rights.”⁸⁷ Thus, in complying with the ICESCR’s obligations under the right to health, states may justifiably differ in their actions based upon their respective political will, disease prevalence, and economic resources, so long as their compliance efforts “move as expeditiously and effectively as possible towards the full realization of article 12.”⁸⁸ As a result, emphasis must be placed—particularly in developing states seeking to uphold the right to health—on the most cost-efficient delivery of life-saving services to the greatest number of people.⁸⁹

B. Nicotine Addiction Implicates the Right to Health

The right to health does not include the right to be healthy, and, as such, it does not address an individual’s lifestyle choices, regardless of their effects on health.⁹⁰ Yet, “[d]efining tobacco as a justice issue can be

87. ICESCR, *supra* note 62, art. 2.

88. *General Comment 14*, *supra* note 63, ¶ 31; FIDLER, *supra* note 63, at 184 (“The principle of progressive realization stands, therefore, for two propositions: (1) the ability of States to fulfill the right to health differs because their economic resources differ; and (2) the different levels of economic development . . . mean that not all countries will enjoy an equivalent standard of health.”); Steven D. Jamar, *The International Human Right to Health*, 22 S.U. L. REV. 1, 52 (1994) (“Implementation involves policy driven allocative judgments which are not based solely on principles or policies, but which are based also on political and economic considerations.”).

89. Osita C. Eze, *Right to Health as a Human Right in Africa*, in *THE RIGHT TO HEALTH AS A HUMAN RIGHT* 76, 87 (1979) (“It is little use looking at the statistics to find out how many doctors and other auxiliary medical staff there are for a given number of the population; how many hospitals, clinics and beds are built or acquired every year, nor what percentage of the national budget is spent on providing health facilities to the population. It is necessary to ascertain how many benefit from these facilities.”). As noted by Dr. Thomas Adeoye Lambo,

The technologies to be used in achieving this transition [to the delivery of health care] should be capable of operations within the meagre financial and material resources of the poor communities of the Third World; be adapted to the available resources of human skills within the community; they should be socially and culturally acceptable and, lastly, be functionally efficient.

Id. at 78-79 (quoting A. Lambo, *The Health of Man in a Sick World*, paper presented at the 10th Anniversary Meeting of the Club of Rome).

90. *General Comment 14*, *supra* note 63, ¶ 8 (“The right to health is not to be understood as a right to be *healthy*.”); *id.* ¶ 9 (excluding “unhealthy or risky lifestyles” from protection

contentious because many people still believe that tobacco use is solely an individual behavior choice and tobacco illness a lifestyle disease.”⁹¹ For smoking cessation to fall under the right to health, it is vital that smoking not be viewed as a personal decision or a distasteful habit but rather as a chronic illness largely outside the control of the individual.

For years, transnational tobacco corporations have successfully “blamed the victim,” advancing the pseudo-scientific view that smoking is a lifestyle decision rather than a physiological addiction.⁹² In doing so, the tobacco industry has tried to co-opt human rights rhetoric, cynically employing the language of “choice” to advance its corporate objectives.⁹³ However, smoking is not simply the result of conscious choice but rather the culmination of pharmacological, sociocultural, and demographic factors exploited by rapacious transnational tobacco corporations. Because tobacco use has been proven to result in a powerful addiction that impairs

under the right to health in Article 12 of the ICESCR); Leonard S. Rubenstein, *Human Rights and Fair Access to Medication*, 17 EMORY INT’L L. REV. 525, 530 (2003) (noting that the right to health is “not a right to be healthy, since genetic make-up, *individual behavior* and other factors also affect health” (emphasis added)); Taylor, *supra* note 38, at 310 (“The right to health does not, however, constitute an entitlement to individual good health.”).

91. Cheryl Heaton & Kathleen Nelson, *Reversal of Misfortune: Viewing Tobacco as a Social Justice Issue*, 94 AM. J. PUB. HEALTH 186, 187 (2004).

92. *Id.* (“Tobacco marketers’ public relations strategies have long sought, falsely, to frame the issue of tobacco use as one of ‘freedom of choice’ and ‘smokers’ rights’ to downplay the nicotine-dependency argument.”).

93. Crow, *supra* note 12, at 225 (suggesting the use of international legal bodies as a means of “enabl[ing] the tobacco control community to reclaim the language of rights from the tobacco industry, which regularly uses this tactic to promote its own objectives”); Peter D. Jacobson & Soheil Soliman, *Co-opting the Health and Human Rights Movement*, 30 J.L. MED. & ETHICS 705, 708 (2002) (“Internal tobacco industry documents show that the industry was aware early on that [human rights rhetoric] would be a powerful strategy for combating regulation.”).

Despite this rhetoric of choice, the tobacco industry has been keenly aware of and exploited the commercial benefits of nicotine’s addictive properties since at least 1962. STANTON A. GLANTZ ET AL., *THE CIGARETTE PAPERS* 58-60 (1996). Compounding this deceitful rhetoric, the CEOs of every major tobacco corporation swore before the U.S. Congress as late as 1994 that they believed nicotine not to be addictive. See generally PHILIP J. HILTS, *SMOKE SCREEN: THE TRUTH BEHIND THE TOBACCO INDUSTRY COVER-UP* (1996); Allan M. Brandt & Julius B. Richmond, *Tobacco Pandemic*, WASH. POST, Jan. 15, 2004, at A21. Even today, as tobacco executives attempt “to extricate the companies from the cul-de-sacs into which they had placed themselves by their earlier denial,” they continue to deny the addictive power of nicotine, spuriously likening nicotine dependence to that of coffee and chocolate. Collin, *supra* note 12, at 77.

autonomous decision-making and impedes voluntary choice, an individual's decision to continue smoking cannot be said to be the result of a truly free and informed choice.⁹⁴ Through the addiction, "the freedom to commit obviously imprudent actions may have the consequence of creating conditions in which continuing autonomy can no longer be maintained."⁹⁵ As a result, tobacco control—once considered a private good, affecting only lifestyle choices—must now be reevaluated as a public good, requiring a public health based approach to treat involuntarily recalcitrant smokers.⁹⁶

Although nicotine is not the direct agent of harm, it is nevertheless the pharmacological basis of tobacco smoking, causing deadly consequences for smokers and those exposed to environmental tobacco smoke. It is now axiomatic that nicotine is a drug of addiction, inducing pharmacological and behavioral processes similar to those of heroin and cocaine.⁹⁷ Cigarettes and other tobacco products can therefore be viewed as highly engineered drug delivery vehicles for sating this nicotine addiction, which, even if used as directed, can cause death. As such, it becomes clear that "[t]he cigarette did for nicotine what crack did for cocaine: it made the drug relatively convenient and uniquely addictive by making nicotine easily and conveniently inhalable."⁹⁸ Transnational tobacco corporations have marketed to this addiction, with well over a billion people self-administering

94. ROBERT E. GOODIN, *NO SMOKING: THE ETHICAL ISSUES* 7 (1989) (arguing that "what we are being protected from is something that would deprive us of the capacity for autonomous choice"). Furthermore, an individual's initial decision to begin smoking is made frequently when he or she is too young to be truly informed about the risks of smoking and give meaningful consent to those risks.

95. Albert Weale, *Invisible Hand or Fatherly Hand? Problems of Paternalism in the New Perspective on Health*, 7 J. HEALTH POL. POL'Y & L. 784, 800 (1983) (detailing the conditions under which "free decisions are unlikely to be the best guide to a person's interests"). Ironically, transnational tobacco corporations have consistently marketed cigarettes as a means of expressing freedom and individuality. Collin, *supra* note 12, at 72.

96. See Taylor & Bettcher, *supra* note 51, at 925 ("Traditionally, prevention or treatment of noncommunicable diseases was considered to be mostly a private good, since the risk factors associated with such diseases, including use of tobacco, are related to individual choices in lifestyle.").

97. U.S. DEP'T OF HEALTH & HUMAN SERVS., *NICOTINE ADDICTION: THE HEALTH CONSEQUENCES OF SMOKING: A REPORT OF THE SURGEON GENERAL* (1988); WORLD HEALTH ORG., *supra* note 1950, at 18; E.W. Lee & G.E. D'Alonzo, *Cigarette Smoking, Nicotine Addiction, and Its Pharmacologic Treatment*, 153 ARCHIVES INTERNAL MED. 34 (1993).

98. Henningfield & Slade, *supra* note 17, at 81 (citing John Slade, *Nicotine Delivery Devices*, in *NICOTINE ADDICTION: PRINCIPLES AND MANAGEMENT* 3 (C. Tracy Orleans & John Slade eds., 1993)).

a highly addictive psychoactive drug to maintain their deadly “habit.” It is the nicotine addiction and withdrawal symptoms—not free choice—that prevent countless smokers from achieving and sustaining smoking cessation.⁹⁹ Thus, from a rights perspective, cessation interventions should be analyzed as nothing more than the clinical treatment of nicotine addiction and its concomitant manifestations of disease and death.

Nicotine addiction is a chronic illness, necessitating the state provision of medical resources to enhance individual autonomy in deciding whether or not to continue smoking. The WHO has recognized that nicotine addiction is a disease¹⁰⁰ and that “nicotine dependence is clearly a major barrier to successful cessation.”¹⁰¹ Viewing the right to health as a right that enhances autonomy and human dignity, states must prioritize health interventions to promote those treatments “most likely to increase autonomy amongst those least able to exercise it without outside help.”¹⁰² Treating those addicted to nicotine should be a priority. Yet the FCTC does not treat the addiction as a disease, denying tobacco the clinical diagnosis that would trigger obligations under the right to health.

C. A Right to Health Approach to Smoking Cessation

Although international treaties recognize a right to health, the right is frequently criticized for being “so broad that it lacks coherent meaning and is qualified by the principle of progressive realization.”¹⁰³ Because of this, the WHO has rarely approached health issues from a human rights

99. As noted by Drs. Henningfield and Slade in recognizing nicotine addiction as a disease unto itself:

The American Psychiatric Association has identified two medical disorders that pertain to nicotine addiction: 1) nicotine dependence, which is a “pattern of repeated self-administration that usually results in tolerance, withdrawal, and compulsive drug-taking behavior,” and 2) nicotine withdrawal, which causes “clinically significant distress or impairment in social, occupational, or other important areas of functioning.”

Henningfield & Slade, *supra* note 17, at 79.

100. Second Meeting of the Working Group, *supra* note 52 (recalling, in the preamble of the Proposed Dependence Protocol, that “tobacco dependence is classified as a disease under the International Classification of Diseases (ICD-10), and that nicotine addiction is classified as a disease under the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV”).

101. WORLD HEALTH ORG., *supra* note 19, at 19.

102. ALASTAIR V. CAMPBELL, *MEDICINE, HEALTH AND JUSTICE: THE PROBLEM OF PRIORITIES* 282 (1978).

103. FIDLER, *supra* note 63, at 197.

perspective.¹⁰⁴ In the present case, the FCTC never articulates the right to health as the normative justification for any of its obligations on states, robbing the FCTC of the moral authority necessary to enact comprehensive tobacco control programs.¹⁰⁵

Under the right to health, states have affirmative obligations to provide tobacco cessation interventions that are available, accessible, culturally acceptable, and medically appropriate.¹⁰⁶ The AIDS pandemic refocused the right to health, reengaging primary health care as a bedrock of public health. Global control of the AIDS pandemic initially ignored the right of the afflicted to humane treatment. As noted by George Annas, this global AIDS strategy was based on a “war-containment or escalation discourse (the ‘war on AIDS’ strategy), in which control is viewed as an end in itself and the infected body becomes a battlefield.”¹⁰⁷ Annas contrasted this with a “human rights discourse, in which our collective futures and the values of human flourishing and the right to humane treatment are paramount.”¹⁰⁸ After years of scholarship and advocacy, this rights-based

104. Leary, *supra* note 76, at 167 (noting that the WHO has “shown little interest in approaching health issues through the lens of human rights”). *But see also* Allyn L. Taylor, *Governing the Globalization of Public Health*, 32 J.L. MED. & ETHICS 500, 505 (2004) (recognizing that “notable strides were made to address the [WHO’s] historical neglect of the linkage between health and human rights” during Dr. Brundtland’s tenure as Director-General).

105. Although “[t]he possibility of including more direct references to the human rights implications of tobacco control in the FCTC was discussed at various points in the treaty’s evolution,” member states eventually relegated any mention of the right to health to the Preamble. Crow, *supra* note 12, at 222 n.78; *cf. International Law and Health, Two Approaches: The World Health Organization’s Tobacco Initiative and International Drug Controls, Summary of Remarks by Virginia Leary*, 94 AM. SOC’Y INT’L L. PROC. 193, 195 (2000) (suggesting that “focusing on the problem of tobacco consumption is a useful means of concretizing the ‘right to health’ and thus joining the human right community in an alliance with the public health community in implementing that right”). While most studies attempt to expand the right to health to include aspects of preventive medicine, *e.g.*, FIDLER, *supra* note 63, at 305-07 (proposing a “Framework Convention on Infectious Diseases”), this Article attempts to define obligations pursuant to the core “health care” component of the right to health.

106. *See infra* note 114 and accompanying text.

107. George J. Annas, *The Impact of Health Policies on Human Rights: AIDS and TB Control*, in *HEALTH AND HUMAN RIGHTS*, *supra* note 60, at 37 (exploring the effect of discourse on public health strategies).

108. *Id.*; Yamin, *supra* note 62, at 330 (“The fundamental premise underlying the notion of universal human rights is that people are not expendable; those people’s avoidable deaths are not just a tragic shame.”); *see also infra* notes 168-185 and accompanying text (discussing the application of the right to health in establishing an entitlement to life-saving

discourse for AIDS treatment has now become engrained in the clarion call for access to antiretroviral therapies, with scholars arguing that “access to medications has been recognized as implicating both the right to life and the right to health under international law.”¹⁰⁹ This paradigm shift has reinforced the normative content of the right to health, explicitly including a right to treatment for life-threatening disease. If humane medical treatment is to be found for smokers, it too may be found in the right to health.

Much like the inequity of focusing only on prevention while ignoring those suffering from AIDS, governmental focus solely on preventing the initiation of smoking violates the human rights of those already addicted to tobacco. In fulfilling obligations under the right to health with respect to tobacco control, states must develop intervention programs to treat addicted smokers.¹¹⁰ As interpreted by the CESCR, a state’s obligation to fulfill the right to health has three interrelated components:

The obligation to *fulfil (facilitate)* [the right to health] requires States inter alia to take positive measures that enable and assist individuals and communities to enjoy the right to health. States parties are also obliged to *fulfil (provide)* a specific right contained in the Covenant when individuals or a group are unable, for reasons beyond their control, to realize that right themselves by the means at their disposal. The obligation to *fulfil (promote)* the right to health requires States to undertake actions that create, maintain and restore the health of the population.¹¹¹

This tripartite framework requires states to establish a national policy to move progressively toward universal access to life-saving interventions.¹¹² Thus, states must intervene to provide access to tobacco cessation treatments—including, but not limited to, essential medications. In recognizing cessation under the right to health, it is imperative that states acknowledge cessation interventions as an essential treatment for the disease of addiction.

medications).

109. Yamin, *supra* note 62, at 370.

110. Wike, *supra* note 62, at 360 (noting that “one could easily find a *state duty to render health care for those affected by tobacco*, both smokers and nonsmokers, as well as to redistribute the social costs of tobacco’s ill effects”) (emphasis added).

111. *General Comment 14*, *supra* note 63, ¶ 37.

112. Yamin, *supra* note 62, at 357-59.

D. Developing Hierarchy from the Cessation Continuum

The FCTC has promulgated low-cost policy approaches to smoking cessation that serve only to foster a social climate and supportive environment for quitting. These public health measures have allowed for the creation of smoke-free workplaces, increased taxation of tobacco, packaging regulations, enhanced education, and smuggling prohibitions. Although these measures do promote smoking cessation indirectly—denormalizing the act of smoking itself—such measures alone are clearly insufficient to aid those smokers addicted to tobacco and unable to quit.¹¹³

To fulfill its obligations under the right to health, a state must provide facilities, services, and essential medications that are: (1) available in sufficient quantity, (2) accessible without discrimination, (3) culturally acceptable, and (4) medically appropriate and of good quality.¹¹⁴ While these aspects of the right to health are interrelated, each is essential to an equitable state cessation intervention. As science and technology have evolved, so too has the scope of each aspect of a state's obligations under the right to health.¹¹⁵ Using these principles as a guide, states should undertake an evidence-based comprehensive health systems approach to tobacco control in addition to the public health programs within the FCTC.¹¹⁶ This would allow states to take a more active role in smoking cessation by incorporating contemporary clinical best practices into their national health policy.¹¹⁷

113. *E.g.*, Heaton & Nelson, *supra* note 91, at 189 (noting that “[e]ven though cost may be an incentive to quit, tobacco addiction can be stronger than a rational financial decision”).

114. *General Comment 14*, *supra* note 63, ¶ 12.

115. Taylor, *supra* note 38, at 311.

116. The distinctions between a public health approach and health systems approach to smoking cessation are noted in matrices developed at the June 2002 WHO meeting on Global Policy for Smoking Cessation in Moscow, Russia. WORLD HEALTH ORG., *supra* note 20, at 7-10.

117. *See General Comment 14*, *supra* note 63, ¶ 36 (“The obligation to *fulfil* [the right to health] requires States parties, *inter alia*, to give sufficient recognition to the right to health in the national political and legal systems, preferably by way of legislative implementation, and to adopt a national health policy with a detailed plan for realizing the right to health.”). Clinical best practices refer to evidence-based guidelines of smoking cessation compiled through meta-analyses of published research. The two major clinical best practices reports on tobacco cessation are the U.S. Agency for Health Care Policy and Research’s *Clinical Practice Guideline for Smoking Cessation*, M. FIORE ET AL., U.S. AGENCY HEALTH CARE POL’Y & RESEARCH, SMOKING CESSATION: CLINICAL PRACTICE GUIDELINE NO. 18 (1996), *updated in* M. FIORE ET AL., U.S. DEP’T OF HEALTH & HUM. SERVS., TREATING TOBACCO

A health systems approach to smoking cessation includes both behavioral and pharmacological interventions to overcome an individual smoker's nicotine addiction.¹¹⁸ This combination of interventions buttresses the individual smoker's ability to progress through the psychological stages of quitting (pre-contemplation, contemplation, readiness, action, and maintenance) while deterring relapse to addictive smoking behaviors.¹¹⁹ Whereas providing a primary health care system is a core obligation of the right to health that cannot be deferred for lack of resources,¹²⁰ other resource-based obligations are to be assured through progressive realization over time.¹²¹ Thus, although the health system's combination of behavioral and pharmacological interventions offers the

USE AND DEPENDENCE: CLINICAL PRACTICE GUIDELINE (2000) [hereinafter M. FIORE ET AL. (2000)]; and the British Health Education Authority's *Smoking Cessation Guidelines for Health Professionals*, Martin Raw et al., *Smoking Cessation Guidelines for Health Professionals*, 53 THORAX 1 (1998).

118. Niaura & Abrams, *supra* note 28, at 499 (citing FIORE ET AL. (2000), *supra* note 117) (noting that "multicomponent programs enjoy greater efficacy compared with single component programs" and that "more is better"). The Proposed Dependence Protocol provides a preliminary definition of "tobacco dependence treatment," which "includes (singly or in combination) behavioural and pharmacological interventions such as education, brief counseling and advice, intensive support, administration of pharmaceuticals or other interventions that contribute to reducing and overcoming tobacco dependence in individuals and in the population as a whole." Second Meeting of the Working Group, *supra* note 52, at 6.

119. Multicomponent intervention outperformed either behavioral intervention or pharmacological intervention, when employed alone. MACKAY & ERIKSEN, *supra* note 6, at 82; John R. Hughes et al., *Recent Advances in the Pharmacotherapy of Smoking*, 281 JAMA 72, 75 (1999) (finding that pharmacological and behavioral interventions augment each other); Marcel E. Pieterse, *Effectiveness of a Minimal Contact Smoking Cessation Program for Dutch General Practitioners: A Randomized Controlled Trial*, 32 PREVENTIVE MED. 182, 188 (2001); Russell, *supra* note 21, at 20 ("When used as an adjunct to intensive support in specialized clinics NRT [nicotine replacement therapy] products are equally effective, with success rates averaging around 25-30% sustained, lapse-free, biochemically validated cessation throughout one year."). For theoretical hypotheses explaining the mechanisms through which behavioral and pharmacologic interventions augment each other, see John R. Hughes, *Combining Behavioral Therapy and Pharmacotherapy for Smoking Cessation: An Update*, in INTEGRATING BEHAVIOR THERAPIES WITH MEDICATION IN THE TREATMENT OF DRUG DEPENDENCE 92 (L.S. Onken et al. eds., 1995).

120. *General Comment 14*, *supra* note 63, ¶ 43; Rubenstein, *supra* note 90, at 531 (noting the importance of the obligation to provide essential drugs in assuring HIV antiretroviral therapies).

121. See *supra* notes 87-89 and accompanying text.

best hope for breaking tobacco dependence while remaining the least intrusive on other rights,¹²² such comprehensive interventions are not currently within the capacity of many states. Still, for states seeking to allocate health resources to maximize the health of all of their citizens,¹²³ smoking cessation, relative to other public health measures, can offer the greatest return (in lives saved) on a state's investment.¹²⁴ Further, states can maximize efficiency by coordinating mechanisms of behavioral and pharmacological interventions through public or private insurance schemes. In allocating these resources to achieve the progressive realization of the right to health, the following cessation intervention hierarchy would allow states to prioritize smoking cessation methods in accordance with the right to health while acknowledging national circumstances and resource availability.¹²⁵

122. WORLD HEALTH ORG., *supra* note 19, at 19 ("In preparing national tobacco control plans and strategies, planners may wish to encourage the provision of a broad range of smoking cessation strategies that would include combinations of the most effective group programmes of smoking cessation, physician advice and, where appropriate, nicotine replacement therapy."); WORLD HEALTH ORG., *supra* note 20, at 51 (noting that "a combination of behavioural and pharmacological treatment produces the best outcomes"); Henningfield & Slade, *supra* note 17, at 79; Warner, *supra* note 20, at 115 ("The combination of serious physician counseling with patient follow-up and use of pharmacotherapy can produce cessation rates in the vicinity of 30%.").

123. See Theo C. Van Boven, *The Right to Health: Paper Submitted by the United Nations Division of Human Rights*, in *THE RIGHT TO HEALTH AS A HUMAN RIGHT*, *supra* note 89, at 54, 63-64 (noting that the United Nations Division of Human Rights has investigated "[w]hether advanced medical techniques for the prolongation of life should be applied to a few patients as long as the cost involved curtails the provision of less sophisticated medical care . . . for the many . . . where the economy cannot accord to every sick person the entire range of available medical treatment from which he could benefit").

124. See Lawrence O. Gostin, *Public Health, Ethics, and Human Rights: A Tribute to the Late Jonathan Mann*, 29 J.L. MED. & ETHICS 121, 125 (2001) ("When public health authorities work in the areas of tobacco control, the environment, or occupational safety, for example, their belief is that everyone will benefit from smoking cessation, clean air, and safe workplaces."); Niaura & Abrams, *supra* note 28, at 502 ("[S]moking cessation interventions are arguably the most cost-effective of any preventive or other medical interventions. Moreover, interventions are cost-effective across a range of intensity, for example, from clinician advice to pharmacotherapy to specialized clinics" (citing Tammie O. Tengs et al., *Five Hundred Life Saving Interventions and Their Cost Effectiveness*, 15 RISK ANALYSIS 369 (1995))).

125. Whereas the Second Meeting of the Working Group articulates several of the cessation interventions analyzed herein, *see* Second Meeting of the Working Group, *supra* note 52, the Proposed Dependence Protocol fails to address how these mechanisms should

1. Behavioral Interventions

Behavioral interventions offer the best opportunity for states to control tobacco addiction at limited cost. Given that “[s]ocial support for quitting should be possible in all countries, even those with extremely limited resources,”¹²⁶ the right to health mandates that states undertake the lifesaving behavioral interventions discussed below without regard to state resources. More burdensome than the requirements of the FCTC, the following cessation programs require state action to establish a scientifically based institutional framework for behavioral interventions.

a. Physician Advice¹²⁷

A state health system can only succeed in meaningfully reducing smoking prevalence where individual physicians reach out directly to their patients who smoke.¹²⁸ Studies have shown that even brief advice from a physician can dramatically increase cessation rates, improving abstinence rates by up to thirty percent.¹²⁹ Because of the frequency with which smokers are forced into the health care system and the efficacy of physician advice, physician interventions—including information, services, and referrals—promise to be the most efficient cessation treatment in successfully influencing the greatest number of smokers motivated to quit.¹³⁰

Despite this, many physicians eschew treatment of tobacco addiction

be attained in the context of the principle of progressive realization, *see supra* note 88 and accompanying text.

126. WORLD HEALTH ORG., *supra* note 20, at 51 (citation omitted).

127. In this context, “physician advice” refers to any one-on-one cessation intervention delivered in the context of other medical services by any health care provider, including doctors, nurses, nurse practitioners, pharmacists, and dentists.

128. Russell, *supra* note 21, at 20 (“It is only through the primary care system that large enough numbers of smokers can be reached to produce a significant reduction in national prevalence.”).

129. ROYAL COLLEGE OF PHYSICIANS, NICOTINE ADDICTION IN BRITAIN: A REPORT OF THE TOBACCO ADVISORY GROUP OF THE ROYAL COLLEGE OF PHYSICIANS (2000); Niaura & Abrams, *supra* note 28, at 497 (noting that “there is a dose-dependent relationship between the intensity of person-to-person contact and successful cessation outcome”) (citations omitted); Pieterse, *supra* note 119, at 187.

130. David B. Abrams et al., *Integrating Individual and Public Health Perspectives for Treatment of Tobacco Dependence Under Managed Health Care: A Combined Stepped-Care and Matching Model*, 18 ANNALS BEHAV. MED. 290, 294 (1996) (defining efficiency as the product of population, reach, and efficacy).

because they lack the resources, motivation, and understanding necessary for effective intervention.¹³¹ Consequently, the WHO has advised that “[a]ll health professionals, including doctors, nurses and pharmacists, should be given both basic and in-service training so that they are capable of providing advice and treatment for tobacco dependence.”¹³² As noted in the FCTC, states should “include diagnosis and treatment of tobacco dependence and counselling services on cessation of tobacco use in national health and education programmes, plans and strategies, with the participation of health workers, community workers and social workers.”¹³³ Realizing this aspiration involves education in smoking and smoking cessation as part of the core curriculum of schooling and post-graduate training, with detailed education in smoking for physicians specializing in oncology, cardiovascular disease, obstetrics, and adolescent health.¹³⁴

By relating one-on-one with the patient, physicians can provide efficacious, culturally sensitive advice that is appropriately tailored to the patient’s individual smoking habits and quitting methods.¹³⁵ This “patient-

131. WORLD HEALTH ORG., *supra* note 20, at 52; *see also* L.H. Ferry et al., *Tobacco Dependence Curricula in US Undergraduate Medical Education*, 282 JAMA 825, 825 (1999); Niaura & Abrams, *supra* note 28, at 497 (listing the barriers that influence physician readiness to adopt smoking cessation interventions and recognizing that “effective strategies are needed to enhance the adoption of efficacious smoking cessation interventions within a population of primary care physicians and practices”); J.G. Spangler et al., *Tobacco Intervention Training: Current Efforts and Gaps in US Medical Schools*, 288 JAMA 1102, 1108 (2002). Although physicians generally are not trained for cessation services, or reimbursed for their counseling efforts, the availability of NRTs, *see infra* Subsection II.D.2.a, has given physicians a clinical reason to engage their patients about smoking. Warner, *supra* note 20, at 116. *But see* Hughes et al., *supra* note 119, at 75 (theorizing that “approval of OTC [over the counter] medications for smoking cessation may have prompted some physicians to become less interested in providing smoking cessation prescriptions”).

132. WORLD HEALTH ORG., *supra* note 19, at 19; *see also* WORLD HEALTH ORG., *supra* note 20, at 17 (“Efficacious and highly cost-effective treatments have been reviewed in many countries and institutions and they advocate that all health-care personnel and clinicians should consistently deliver smoking cessation interventions to their patients.”).

133. FCTC, *supra* note 1, art. 14(2)(b).

134. As noted by the WHO, this training could be accomplished “by working with international associations such as World Medical Associations, the World Organization of Family Practitioners, and the International Council of Nurses to develop model tobacco control curriculum and course outlines for basic training in delivering smoking-cessation therapies.” WORLD HEALTH ORG., *supra* note 20, at 54. In addition, states should initiate their efforts by lowering the prevalence of smoking among those in the health professions. EUR. P’SHIP TO REDUCE TOBACCO DEPENDENCE, *supra* note 26, at 6.

135. *See* Warner, *supra* note 20, at 116 (“[B]ehavioral scientists have developed financially

treatment matching” would allow assessment prior to treatment, improving both cost-effectiveness and overall efficacy.¹³⁶ While personalized to the individual smoker, this opportunistic system of treatment nevertheless should be based on firm guidelines, for example the United States’s Public Health Service’s Guidelines’ “five As” of individual smoking cessation: Ask about smoking at every opportunity and record smoking status; Advice the smoker to stop; Assess the smoker’s willingness to stop; Assist the smoker to stop through recommendation of treatments and referral to specialists; Arrange follow-up with the smoker.¹³⁷ This intervention strategy is adaptable to several cultures and has been proven efficacious in controlled trials. Specific physician assistance can include helping the smoker to set a quit date, suggesting behavioral strategies to prevent relapse, and prescribing pharmacotherapies to aid those for whom breaking the nicotine addiction requires more than educational and motivational help.¹³⁸

b. Counseling/Support Groups

In contrast to the brief advice of a physician, intensive counseling involves repeated behavioral interventions. As recognized in the FCTC, albeit with its nonobligatory language,¹³⁹ states “shall endeavour to” provide cessation counseling by “establish[ing] in health care facilities and rehabilitation centres programmes for diagnosing, counselling, preventing and treating tobacco dependence”¹⁴⁰ In practice, this will involve a

feasible means of tailoring cessation messages and strategies to the needs and desires of specific individuals.”); cf. Judith Mackay, *Combating Addiction in Developing Countries*, 16 WORLD HEALTH F. 25, 27 (1995) (noting that physicians often fail in promoting smoking cessation when they give the same advice to all smokers) (citing Professor Robyn Richmond).

136. Niaura & Abrams, *supra* note 28, at 499 (“The major theoretical advantage of matching is that smokers can be assessed according to some relevant, predictive dimension prior to treatment, be assigned to receive the treatment that is appropriate and adequate for them, and can avoid thereby the cumulative burdens of trial and failure.” (citations omitted)).

137. FIORE ET AL. (2000), *supra* note 117; see also EUR. P’SHP TO REDUCE TOBACCO DEPENDENCE, *supra* note 26, at 5 (recommending guidelines for physician interventions) (citing TJ GLYNN & MW MANLEY, HOW TO HELP YOUR PATIENTS STOP SMOKING: A NATIONAL CANCER INSTITUTE MANUAL FOR PHYSICIANS (1989); FIORE ET AL. (2000), *supra* note 117).

138. See *infra* Subsection II.D.2.

139. See *supra* note 48 and accompanying text (noting the nonobligatory use of “shall endeavour to” in Article 14 of the FCTC).

140. FCTC, *supra* note 1, art. 14(2)(c).

“smoking cessation specialist,” who works with either individuals or groups to offer coping skills and social support throughout smoking cessation.¹⁴¹ These specialists need not be physicians, thus allowing every state to provide this intervention in meeting its obligations under the right to health.¹⁴² However, because this intervention requires repeated, specialized interaction, it will be more expensive than physician counseling and less likely to appeal to the greatest number of smokers.¹⁴³ Consequently, this form of intervention is not the most advantageous primary means of tobacco cessation intervention, but may nevertheless prove cost-effective as a secondary means of cessation intervention for smokers unable to quit through other means.¹⁴⁴

To ease patient cost outside of traditional counseling formats, telephone help lines,¹⁴⁵ in addition to nascent Internet-based counseling,¹⁴⁶ offer promise for efficacious future treatments as these burgeoning techniques become more widely available and progress from low-cost self-help materials to easily accessible interactive tools. By offering alternative forms of cessation interventions, state programs may more easily reach the smokers least motivated to quit while continuing to allow the interpersonal delivery of services tailored to each individual’s needs.¹⁴⁷

2. Pharmacotherapies

Pharmacotherapies, including nicotine replacement therapy (NRT) and non-nicotine medications, are medically appropriate means of breaking addiction and saving lives. While not guaranteeing the success of

141. EUR. P'SHIP TO REDUCE TOBACCO DEPENDENCE, *supra* note 26, at 5-6.

142. See WORLD HEALTH ORG., *supra* note 20, at 51 (“All countries have lay persons who can provide informal social support for quitting and who can be trained to conduct more formal interventions.”).

143. Niaura & Abrams, *supra* note 28, at 495 (noting that “expensive and more efficacious treatments (e.g., combined pharmacologic and behavioral interventions delivered by smoking cessation specialists) are by definition less able to be disseminated widely and are less likely to appeal to most smokers”).

144. *Id.* (citing E. Lichtenstein & R.E. Glasgow, *Smoking Cessation: What Have We Learned over the Past Decade?*, 60 J. CONSULTING & CLINICAL PSYCHOL. 518 (1992)).

145. EUR. P'SHIP TO REDUCE TOBACCO DEPENDENCE, *supra* note 26, at 6 (noting that “[t]elephone help lines can be effective and are very popular with smokers”).

146. WORLD HEALTH ORG., *supra* note 20, at 21-22 (discussing QuitNet, at <http://www.quitnet.org>; and stop-tabac.ch, at <http://www.stop-tabac.ch>).

147. For advantages of patient-treatment matching through tailored cessation interventions, see *supra* note 136 and accompanying text.

every quit attempt,¹⁴⁸ pharmacotherapies nevertheless represent a clinically effective means of cessation. Pursuant to the right to health, states should take steps to ease the regulation of NRT and non-nicotine medication and make such options available without prescription through either public or private insurance coverage. To assure this, states must use their public health apparatus to assure that these medications are selected solely on the basis of clinical best practices and are accessible through reliable means of distribution to the greatest number of persons. Of course, any discussion of access to pharmacotherapies necessarily implicates the antagonism between international trade regimes and the right to health. As discussed in Section II.E, states should employ the public health exception to international trade agreements, making the production and importation of these treatments affordable, and thus economically accessible, to all. Eliminating institutional barriers to NRTs and non-nicotine medication use would spur pharmaceutical company investment in research and development of new pharmacotherapies.¹⁴⁹ To aid this effort, states may, in accordance with the right to health, collaborate in research regarding new NRT and non-nicotine therapies, incorporating these innovative pharmacotherapies into culturally appropriate cessation interventions.

a. Nicotine Replacement Therapy

In the form of patches, gums, sprays, lozenges, or inhalers,¹⁵⁰ NRT

148. See Niaura & Abrams, *supra* note 28, at 500 tbl.1 (comparing the efficacies of various smoking treatments (i.e., gum, patch, spray, inhaler, bupropion, and clonidine) relative to placebo). Because these pharmacotherapies, as distinguished from “essential” HIV antiretroviral therapies, are neither absolutely necessary nor clearly sufficient to save lives, it is unclear whether access to these treatments can be considered core obligations under the right to health. *Cf. infra* text accompanying notes 191-192 (discussing the implications of pharmacotherapies being labeled “essential drugs”).

149. See Henningfield & Slade, *supra* note 17, at 90 (noting that “decisions of corporate entities are based on all available sources of information—both the real and projected regulatory obstacles (including anticipated size of clinical trials), as well as past and projected marketing obstacles (including restrictions on claims)”).

150. While there are some comparative advantages to each form of NRT—mostly dealing with “preference, affordability and side effects”—there are only marginal differences in cessation efficacy among the various forms. Hughes et al., *supra* note 119, at 75 (advocating patient preference as the “primary basis” for choosing among NRTs); Niaura & Abrams, *supra* note 28, at 500. For a description of the comparative clinical advantages of the various forms of NRT, see Karl Olov Fagerström, *Nicotine-Replacement Therapies*, in NICOTINE AND PUBLIC HEALTH, *supra* note 46, at 199, 200-03; and Henningfield & Slade, *supra* note 17, at 82, 86-88.

allows nicotine maintenance or reduction while diminishing or eliminating the deleterious consequences associated with the use of tobacco products, allowing smokers to modify their behaviors without additionally having to combat the addictive hold of nicotine and its associated withdrawal symptoms.¹⁵¹ As such, NRTs disaggregate nicotine addiction from tobacco dependence, giving individuals the opportunity to abstain from tobacco without being forced to abstain additionally from nicotine.¹⁵² The clinical community regards NRT to be safe (in both the short and long term) and effective, “doubl[ing] the success rates of other cessation efforts, whether or not other interventions are used in parallel.”¹⁵³

NRTs are cost-efficient,¹⁵⁴ self-administrable,¹⁵⁵ and do not require continuous physician intervention.¹⁵⁶ The widespread use of NRTs could avert the deaths of millions of smokers and those exposed to

151. The use of NRTs is based on the theory that “tobacco users could use a safer form of nicotine delivery to break the nicotine-addiction cycle by enabling them to achieve and sustain abstinence from tobacco products while they established new behaviors to resist relapse.” Henningfield & Slade, *supra* note 17, at 85 (citing Jack E. Henningfield, *Nicotine Medications for Smoking Cessation*, 333 NEW ENG. J. MED. 1196 (1995)).

152. It is important to note again that the FCTC refers only to “tobacco dependence,” rather than “nicotine addiction.” See *supra* notes 45-46 and accompanying text. Although the medical community often uses these two terms interchangeably, the advent of NRT clearly implicates the distinction between the uses of these terms in devising tobacco cessation programs.

153. WORLD BANK, *supra* note 7, at 54, 55 tbl. 4.3 (citing Raw et al., *Smoking Cessation: Evidence-Based Recommendations for the Healthcare System*, 318 BRIT. MED. J. 182 (1999)).

154. WORLD HEALTH ORG., *supra* note 19, at 19 (“Although there can be an initially higher cost for NRT, it can be more cost-efficient in the long run for both individuals and governments.”); Novotny et al., *supra* note 27, at 302 (noting that “NRTs could cost about \$276 per disability-adjusted life-year (DALY) in low-income and middle-income countries,” below the cost-effectiveness limit set by the World Bank for these settings); cf. WORLD BANK, *supra* note 7, at 56 (“The cost-effectiveness of nicotine replacement therapy has not been studied widely, especially in the low-income and middle-income countries where most smokers live.”); Shibuya et al., *supra* note 11, at 156 tbl. (calculating the cost effectiveness of NRT by WHO geographic subregion).

155. At present, smokers may obtain many NRTs in varied doses of nicotine delivery, allowing them to self-adjust their nicotine intake. See Fagerström, *supra* note 150, at 200-02 (discussing the pharmacokinetics of nicotine gum, transdermal patch, nasal spray, oral inhaler, and sublingual tablet).

156. WORLD BANK, *supra* note 7, at 54 (noting the advantages of self-administration for smokers “in countries where there are limited resources for intensive support by health professionals”).

environmental tobacco smoke.¹⁵⁷ Yet NRTs are significantly more difficult and expensive to obtain than tobacco products, particularly in the developing world.¹⁵⁸ In upholding the right to health, states must lower marketing regulations on NRTs and subsidize them to the greatest extent possible through either private or public insurance mechanisms, thereby making NRTs as readily available as the addictive products they serve to counteract.¹⁵⁹ As empirically shown, NRTs would be most effective in reaching the largest number of smokers if they became, in order of importance: (1) available over the counter, rather than solely by prescription;¹⁶⁰ (2) sold in lower supply, rather than as a whole course of

157. *Id.* at 56.

158. As noted by the WHO,

whereas the tobacco industry ensures that tobacco products are readily available, attractive, and highly affordable, pharmacotherapy is frequently out of reach, available often only by prescription or from limited points of sale, and is often more expensive on a daily basis and point of sale basis because it is generally distributed in packages that include behavioural treatment guidance and sufficient units to discourage simple occasional use as a temporary substitute for tobacco.

WORLD HEALTH ORG., *supra* note 20, at 16; *see also* Henningfield & Slade, *supra* note 17, at 76 (noting that “proven effective tobacco-dependence treatments remain far more restricted in marketing (and thus far less appealing), and far less accessible than tobacco products”); Novotny et al., *supra* note 154, at 293, 299 (noting the availability of NRT products in various countries and concluding that “the regulation of pharmaceutical nicotine products is considerably more extensive than the regulation of cigarettes . . . giv[ing] cigarettes market advantages”). In addition to financial accessibility, NRTs are less clinically appealing than cigarettes as a vehicle for nicotine administration. Henningfield & Slade, *supra* note 17, at 83 (noting that “individuals will choose a product (e.g., cigarettes) that provides an immediate, neurologically-based reward, albeit with a substantial risk of disease in the future, over a product (e.g., nicotine medications) that provides little immediate reinforcement and the distant reduction of the risk of tobacco-related disease”).

159. *See* Henningfield & Slade, *supra* note 17, at 81 (noting that cessation interventions “may be viewed as countermeasures to the forces [of tobacco] (such as low unit purchase price, wide availability, ubiquitous advertising, images of glamorization, and comparatively high social acceptability relative to illicit drugs)”).

160. Fagerström, *supra* note 150, at 205 (“[A doctor’s prescription] can be a big obstacle for those who are not close to a physician or who do not have the resources to take time off to see a doctor and pay for the prescription. The need for a prescription may also reinforce an unfortunate notion that there are adverse effects with nicotine-replacement products, while cigarettes must be relatively safe because they are sold freely.”). *But see* Niaura & Abrams, *supra* note 28, at 500 (noting that “the efficacy of the gum and patch in [the OTC] environment is less than that observed in controlled clinical trials and probably depends to a significant degree on factors such as underdosing, ceasing use prematurely, using

treatment with a large initial payment;¹⁶¹ and (3) subsidized as part of a health care plan.¹⁶² Through such steps, NRTs have the potential to reach those who need them, dramatically decreasing tobacco-induced disease and death.

b. Non-Nicotine Medication

Bupropion hydrochloride, an antidepressant, has been approved in the United States for use as a first-line therapy for smoking cessation.¹⁶³ Marketed by GlaxoSmithKline as Zyban®, this sustained-release formulation of bupropion is the first non-nicotine-based medication to receive approval by the United States Food and Drug Administration. In clinical trials, bupropion doubled rates of cessation as compared to placebo.¹⁶⁴

In addition, the United States Food and Drug Administration has considered—but not yet approved—clonidine (an antihypertensive agent that alleviates withdrawal symptoms) and nortriptyline (an antidepressant) as effective pharmacotherapies for smoking cessation. While clonidine has proven efficacy, it is considered a second-line pharmacologic agent, “partly because of increased likelihood of side effects and rebound blood pressure problems on discontinuation of the drug,” and, thus, it is recommended for use only contingent upon bupropion failure.¹⁶⁵

Despite the early successes of non-nicotine-based medications, “[s]urprisingly little is know[n] about mechanisms of efficacy for

inappropriately, and having an (un)availability of supplemental behavioral treatment”). For a state-by-state analysis of NRT availability over-the-counter, see MACKAY & ERIKSEN, *supra* note 6, at 82-83.

161. Fagerström, *supra* note 150, at 205 (noting that although NRTs are comparable in price to cigarettes, “[m]uch of the perception of high price is related to the larger package sizes compared with cigarettes, which require a greater outlay of money at one time”).

162. If subsidized as part of a private or public health care plan, large-scale procurement would allow insurers to drive down the costs of therapy, using their bargaining leverage in ways similar to those used in obtaining vaccines. Christiane Poulin, *The Public Health Implications of Adopting a Harm-Reduction Approach to Nicotine*, in NICOTINE AND PUBLIC HEALTH, *supra* note 46, at 429, 432-33.

163. WORLD HEALTH ORG., *supra* note 20, at 34 (citing FIORE ET AL. (2000), *supra* note 117).

164. Richard D. Hurt et al., *A Comparison of Sustained-Release Bupropion and Placebo for Smoking Cessation*, 337 NEW ENG. J. MED. 1195, 1199 (1997).

165. Niaura & Abrams, *supra* note 28, at 500.

bupropion and other antidepressants such as nortriptyline.”¹⁶⁶ There is concern that the product development pipeline may dry up unless research partners collaborate to share the burdens of pharmacotherapy research.¹⁶⁷ Through the FCTC framework, states—including state research partnerships with the private sector—have an opportunity to develop these collaborations in researching the biochemical mechanisms of action employed by non-nicotine medications, improving their use and efficacy and engendering the development of new therapeutic compounds.

E. Access to Medications

By invoking the right to health in the context of the FCTC, states would have obligations to provide these vital autonomy-enhancing medications in fulfilling the human rights of those addicted to nicotine.¹⁶⁸ Given the importance of pharmacotherapies in treating tobacco addiction, it is of paramount importance that states make these products accessible. Yet for a state to make these medications accessible in compliance with the right to health will require that they be both physically and economically accessible to all who need them.¹⁶⁹ This cannot be done solely through cooperation with pharmaceutical corporations, whose profit motive often conflicts with public health. Fulfilling these human rights, i.e., making

166. *Id.*

167. *Id.* at 501-02. *But see* Brion J. Fox & Joanna E. Cohen, *Tobacco Harm Reduction: A Call To Address the Ethical Dilemmas*, 4 NICOTINE & TOBACCO RES. S81, S83 (2002) (noting that the majority of U.S. clinical drug trials are performed by pharmaceutical corporations, which forces researchers investigating NRT and non-nicotine therapies to work with, and possibly be controlled by, private corporations driven solely by profit).

168. In addition to the right to health, an argument may be made for universal access to NRTs pursuant to the ICESCR's guarantee of the right of everyone to enjoy the benefits of scientific progress and its applications. *See* GOSTIN & LAZZARINI, *supra* note 83, at 135 (noting that Article 15 of the ICESCR “aims to bring essential scientific advances to not only those who can pay for them”); Rubenstein, *supra* note 90, at 532 (arguing that the ICESCR implies a balance between human rights and intellectual property rights); Yamin, *supra* note 62, at 343-44 (advancing a right to antiretroviral therapies under, *inter alia*, ICESCR's right to the benefits of scientific progress). *Compare* ICESCR, *supra* note 62, art. 15(1)(b) (recognizing “the right of everyone . . . [t]o enjoy the benefits of scientific progress and its applications”), *with* ICESCR, *supra* note 62, art. 15(c) (recognizing “the right . . . [t]o benefit from the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he is the author”). *But see also* FIDLER, *supra* note 63, at 212 (“Within the context of the ICESCR, the right to enjoy the benefits of scientific progress seems to have received less attention than the right to health.”).

169. *General Comment 14*, *supra* note 63, ¶ 12(b).

medications affordable, will require states to combat the injurious mechanics of international trade, a confrontation intentionally avoided by the FCTC.¹⁷⁰ As with the medicalization of HIV treatment, expanded NRT access for states with limited resources will need to circumvent intellectual property protections provided for by the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS).¹⁷¹

In complying with their obligations under the right to health, it is possible for states to circumvent pharmaceutical patents without acting in willful contravention of international trade laws. Maneuvering within the TRIPS regime, Article 8 of the TRIPS Agreement permits a limitation on the TRIPS requirement that states establish standards for protecting intellectual property rights where noncompliance is “necessary to protect public health and nutrition, and to promote the public interest in sectors of vital importance to their socio-economic and technological development, provided that such measures are consistent with the provisions of this Agreement.”¹⁷² The World Trade Organization drafted this admittedly ambiguous exception to balance “the goal of providing incentives for future inventions of new drugs and the goal of affordable

170. Hammond & Assunta, *supra* note 41, at 242 (“The relationship between the FCTC provisions and international trade agreements—one of the most contentious issues in the negotiations—was left ambiguous in the final document, again a result of developed country pressure.”).

171. Agreement on Trade-Related Aspects of Intellectual Property Rights, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, LEGAL INSTRUMENTS—RESULTS OF THE URUGUAY ROUND vol. 31, 33 I.L.M. 81 (1994) [hereinafter TRIPS]. Some scholars have argued, in the case of HIV antiretroviral therapies, that the rights to life and health should take precedence over intellectual property agreements, negating any discussion of TRIPS in providing access to life-saving medications. Rubenstein, *supra* note 90, at 532; Zita Lazzarini, *Making Access to Pharmaceuticals a Reality: Legal Options Under TRIPS and the Case of Brazil*, 6 YALE HUM. RTS. & DEV. L.J. 103, 120-25 (2003). In so doing, these scholars advance the CESCR’s interpretation of the right to health as “clearly alluding to the core obligation to provide essential medications . . . ‘emphasiz[ing] that any intellectual property regime that makes it more difficult for a State party to comply with its core obligations in relation to health, food, education, especially, or with any other right set out in the Covenant is inconsistent with the legally binding obligations of the state party.’” Yamin, *supra* note 62, at 344 (quoting *Human Rights and Intellectual Property*, U.N. CESCR, 27th Sess., ¶ 12, U.N. Doc. E/C.12/2001/15 (2001)).

172. TRIPS, *supra* note 171, art. 8. Nevertheless, it remains unclear how this exception will apply in practice. See Rubenstein, *supra* note 90, at 533 (arguing that the WTO should “take the next step beyond a vague commitment to public health to assure that its interpretations of trade agreements are consistent with international human rights law, including the right to health and its requirement of making essential medicines available”).

access to existing drugs.”¹⁷³ In fact, Brazil, India, Thailand, and South Africa employed this very argument to allow for the manufacture, compulsory licensing, and parallel importation of generic HIV antiretroviral therapies.¹⁷⁴ In the wake of this multinational rebellion against TRIPS’s barriers to addressing the AIDS pandemic, the World Trade Organization has reaffirmed its commitment to the public health safeguard provisions, adopting at the 2001 Doha, Qatar conference the developing states’ position that TRIPS “can and should be interpreted and implemented in a manner supportive of WTO members’ right to protect public health and, in particular, to promote access to medicines for all.”¹⁷⁵ In so doing, the Doha Declaration recognizes access to life-saving medications as part of the human right to health.¹⁷⁶

From this human rights perspective, both within and apart from the TRIPS framework, developed states may have an obligation under the right to health to provide assistance to developing states in realizing their obligations for smoking cessation.¹⁷⁷ These developed states currently fail to respect or protect the right to health by restricting developing states’ access to medications, abusing TRIPS mechanisms in enforcing the rights of pharmaceutical corporations abroad.¹⁷⁸ In fact, these obligations on developed states, while not explicitly stated in rights discourse, underlie

173. WORLD HEALTH ORG. & WORLD TRADE ORG., WTO AGREEMENTS & PUBLIC HEALTH: A JOINT STUDY BY THE WHO AND THE WTO SECRETARIAT 41 (2002).

174. PETCHESKY, *supra* note 79, at 81 (noting that the threats of lawsuits and economic sanctions by the United States and multinational drug companies may itself violate the TRIPS regime).

175. *Ministerial Declaration*, WTO Ministerial Conf., 4th Sess., WT/MIN(01)/DEC/1 (Nov. 14, 2001), http://www.wto.org/english/thewto_e/minist_e/min01_e/mindecl_e.htm. Rosalind Petchesky attributes the success of this position at the Doha conference to, inter alia, weakened U.S. opposition on the subject as a result of the United States’s own consideration of compulsory licensing of the drug Cipro in the face of the U.S. anthrax deaths of 2001. PETCHESKY, *supra* note 79, at 106. However, Petchesky notes that the United States has systematically attempted to weaken consensus on the Declaration since the Doha conference. *Id.* at 107.

176. PETCHESKY, *supra* note 79, at 106.

177. See Rubenstein, *supra* note 90, at 53 (noting an “obligation of international assistance and cooperation” on developed states). Some have argued that developed states are currently failing to respect the right to health just by promoting tobacco and transnational tobacco corporations overseas, affirmatively causing harm to foreign citizens. Wike, *supra* note 62, at 359-60.

178. Yamin, *supra* note 62, at 353 (noting that “laws and regulations that would restrict access to medications by increasing prices—thereby decreasing access—would presumptively constitute a violation of the state party’s obligations under the ICESCR”).

the FCTC cessation provision, which encourages states to “collaborate with other Parties to facilitate accessibility and affordability for treatment of tobacco dependence including pharmaceutical products.”¹⁷⁹ In complying with Article 14 of the FCTC, such collaborative efforts might include, for example, direct aid to developing states, the establishment of a “global fund” for tobacco cessation, or preferential humanitarian pricing for low-income markets. Similar arguments were successfully made in gaining access to HIV anti-retroviral medications under the right to health.¹⁸⁰ Member states clearly were aware of the AIDS treatment analogy when they drafted the text of Article 14.¹⁸¹ Following the June 2002 WHO meeting on Global Policy for Smoking Cessation, the meeting’s policy recommendations advocated that:

It is critical to make cessation products more affordable to those who, so far, have been unable to afford them. It might be worthwhile to organize a campaign similar to that undertaken for AIDS treatment in Africa, which placed significant international pressure on pharmaceutical companies to reconsider their pricing policies for AIDS drugs in poor African countries where the pandemic was escalating. Similarly, there is an argument to be made for making available cheap generic variants of NRT and Zyban-like products and for the relaxation of patent laws for cessation products on the basis of the extremely high death toll exacted by smoking and other tobacco use.¹⁸²

Yet despite this convenient AIDS analogy, smokers—with a less compassion-inducing cause and a lack of stigma-induced cohesion—clearly

179. FCTC, *supra* note 1, art. 14(2)(d). This principle is advanced more specifically in the Proposed Dependence Protocol, which requires that “Parties shall take into account the particular needs of developing countries and assist in improving their national capacities and capabilities to participate in the measures [to treat tobacco dependence].” Second Meeting of the Working Group, *supra* note 52.

180. *E.g.*, WORLD HEALTH ORG., MACROECONOMICS AND HEALTH: INVESTING IN HEALTH FOR ECONOMIC DEVELOPMENT 86-90 (2001); Lazzarini, *supra* note 171, at 115-20. *But cf.* PETCHESKY, *supra* note 79, at 110 (criticizing the World Health Organization’s differential pricing and public-private partnership arrangements because they “work to preserve the system of patents, pre-empt compulsory licensing, construct price reductions as a voluntary or ‘charitable’ response, and thus protect the entire system of markets and capitalist profits”).

181. *See supra* note 179.

182. WORLD HEALTH ORG., *supra* note 20, at 57; *see also id.* at 58 (arguing that “consideration should . . . be given to the liberalization of trade rules where cessation products are involved”).

lack the mobilization apparati that have been a hallmark of the myriad organizations fruitfully demanding treatment for HIV.¹⁸³ As noted in Professor Rosalind Petchesky's assessment of HIV advocacy, "[t]here is no doubting the effective role that demonstrations and other forms of direct action have played in pressuring the US government and transnational drug companies to make significant concessions and in creating a broad public awareness of access to treatment as a human rights issue."¹⁸⁴ Although tobacco will cause more preventable deaths than AIDS over the next century, anti-tobacco advocates have not approached the mobilization or litigation efforts of the global campaign for access to essential medicines for HIV treatment.¹⁸⁵ This is due in part to nongovernmental tobacco control organizations' inability and unwillingness to engage in the human rights debate necessary to lobby for access to pharmaceutical treatments.¹⁸⁶ Consequently, cessation advocacy groups have not gained the public relations leverage necessary to galvanize public opinion for access to treatment. Thus, although the FCTC emphasizes the importance of financial assistance, steadfast resistance from developed states postponed discussion of funding mechanisms until the (currently ongoing) Intergovernmental Working Groups for the establishment of the Conference of the Parties.¹⁸⁷ Once the Conference of the Parties meets, it will have the opportunity to discuss both protocol development and financial assistance to developing states. It is imperative that states act now,

183. Cf. PETCHESKY, *supra* note 79, at 81-84 (chronicling the lobbying and legal strategies of the "global campaign for access to essential medicines" for HIV).

184. *Id.* at 85-104 (providing examples of direct action for HIV care and human rights in South Africa and Brazil).

185. For reasons underlying the intense mobilization of human rights activists around AIDS, see Yamin, *supra* note 62, at 326-27 (noting that "these diseases [AIDS, tuberculosis and malaria]—especially the HIV/AIDS pandemic—have garnered attention due to their economic and social consequences, as well as because of the human tragedies they represent"); and Salih Booker & William Minter, *Global Apartheid*, NATION, July 9, 2001, at 20 ("AIDS thus points to more fundamental global inequalities than those involving a single disease, illuminating centuries-old patterns of injustice.").

186. See Crow, *supra* note 12, at 222 n.78 (attributing the lack of rights language in the FCTC to, inter alia, "the lack of involvement of organizations with experience in rights-based approaches in the negotiations" and "the public health community's relative unfamiliarity with international human rights law" (citing Telephone Interview with Allyn Taylor, Former Senior Legal Adviser to WHO Tobacco Free Initiative (Aug. 19, 2003))); cf. Hammond & Assunta, *supra* note 41, at 241 (noting that the FCTC brought human rights organizations into the tobacco control movement).

187. Crow, *supra* note 12, at 217.

through human rights discourse, to assure that cessation is a part of any discussion on financing of tobacco control.

III. CESSATION PROTOCOL

The FCTC goes far in addressing the global tobacco pandemic, but it neglects the plight of those already addicted to nicotine, with this failure treading heavily upon the right to health. The FCTC is not enough. The convention-protocol approach of the FCTC possesses the inherent advantages and disadvantages of any incremental legislation: While states can assent to broad principles in the convention, this “enables nations to relieve some public pressure for action without resolving or committing to taking concrete steps to control tobacco production and consumption.”¹⁸⁸ A protocol, separately negotiated and ratified, would possess the same legal weight as the FCTC and could thereby create the subsequent obligations lacking in the convention itself. The FCTC is ineffective in addressing smoking cessation, failing to uphold the right to health, where it does not cause states to alter their behavior in line with evidence-based clinical best practices for cessation interventions.¹⁸⁹ A global tobacco control program can be effective only through strong protocol language that upholds a right to clinical cessation interventions and clarifies its substantive obligations.

The initial Proposed Dependence Protocol, though never fully considered by the WHO’s Intergovernmental Negotiating Body, offers valuable language for the development of a protocol to address the want of smoking cessation in the FCTC. Specifically, the Proposed Dependence Protocol obligates states to (1) take all practical, effective, and cost-effective measures to treat tobacco dependence within national health care and social welfare systems; (2) exchange information with and provide technical and financial support to other states; (3) survey and report on tobacco dependence treatments; and (4) support research and development into tobacco dependence treatments.¹⁹⁰ The Proposed Dependence Protocol provides a framework upon which a human rights based protocol might be drafted.

To address the deficiencies in the Proposed Dependence Protocol,

188. TAYLOR & ROEMER, *supra* note 34, at 17; *see also supra* note 3 and accompanying text (discussing the FCTC’s “convention/protocol approach” to treaty-making).

189. Taylor & Bettcher, *supra* note 51, at 923 (noting that “a measure of the agreement’s effectiveness is determined by the extent to which it causes the states to alter their behaviour in line with the national obligations contained in the treaty”).

190. *See infra* Annex; *see also supra* note 52 and accompanying text.

introducing many of the cessation strategies analyzed throughout this Article, it is vital that any protocol affirm member states' commitment to the right to health. A human rights basis for cessation would give credence to international regulation over that which is purely domestic in character. For example, simply by declaring NRTs to be "essential drugs" within the WHO Action Programme on Essential Drugs,¹⁹¹ the Proposed Dependence Protocol would trigger state obligations to make these products available in sufficient quantity to address the needs of smokers.¹⁹² Although tobacco cessation must be undertaken at the national level, it nevertheless requires that states band together in developing international solutions for these domestic problems. Reinstating the Proposed Dependence Protocol with explicit reference to human rights would create norms for tobacco cessation consistent with state obligations to protect and fulfill the right to health.

Tobacco cessation is not simply an issue confined to high-income developed states, but a globalized issue of universal importance. Transnational tobacco corporations have resisted international regulation, framing the FCTC as the "New Colonialism," a Western solution to a Western problem that has been forcibly imposed by Westerners on reluctant developing states.¹⁹³ However, belying the industry's argument, developing states have shown intense advocacy for transnational collaboration to address global tobacco, recognizing that they cannot each combat transnational tobacco corporations alone.¹⁹⁴ In light of this global

191. See *WHO Model List of Essential Drugs*, in 13 WHO DRUG INFO. 249 (1999).

192. *General Comment 14*, *supra* note 63, ¶ 12(a) (establishing that access to "essential drugs, as defined by the WHO Action Programme on Essential Drugs" is part of the minimum core obligations of Article 12 of the ICESCR).

193. See Collin, *supra* note 12, at 79 ("An increasingly significant area in which the tobacco industry seeks to structure debate, and of particular interest in the context of globalization, is the attempt to present tobacco control as an issue for high-income countries.").

194. Bates, *supra* note 41, at 209 (noting that "the most powerful response [at the second meeting of the Intergovernmental Negotiating Body] came from the developing countries . . . dispel[ing] the myth inspired by the tobacco industry that poor countries somehow have other, more important, matters to consider than the tobacco epidemic").

Although described by the tobacco industry as a "developed world obsession being foisted on the developing world," it was in fact developing countries which saved the FCTC from being gutted by a handful of developed countries which have no intention of ever implementing most of its provisions. Unlike other treaties, where developed countries dominate the debate, developing countries were vocal, spirited, and led the charge for most of the progressive provisions.

desire for international cooperation, the industry's argument that the FCTC is unresponsive to the needs of the poor appears to be nothing more than pretext for the true tobacco corporate strategy: By leaving individual states "free to develop the most appropriate policies for the specific circumstances of their country,"¹⁹⁵ transnational tobacco corporations can more easily "divide and conquer" in manipulating individual national policies.¹⁹⁶ Despite this, many nongovernmental tobacco control advocates have played into the tobacco industry's national tobacco control paradigm, eschewing cessation through the FCTC, in part because they lack the discursive skills to engage in the human rights debate.¹⁹⁷ The Proposed Dependence Protocol offers an opportunity, like the FCTC itself, to overcome this industry influence,¹⁹⁸ but, if the goal of member states is to

Hammond & Assunta, *supra* note 41, at 241 (citation omitted).

195. News Release, British American Tobacco, British American Tobacco Proposes "Quantum Leap" for Sensible Tobacco Regulation (Aug. 29, 2000), <http://www.bat.com>, quoted in Collin, *supra* note 12, at 79.

196. See Brundtland, *supra* note 32, at 750 ("[A]ttempts to control tobacco face concerted opposition. Some tobacco companies act to manipulate public opinion, deceive the public about their efforts to develop nicotine delivery devices, target children, and fund research merely to sow doubt about the scientific evidence of the risks of tobacco use." (citations omitted)); Collin et al., *supra* note 13, at 266 (recognizing "the ability of transnational corporations (TNCs) to undermine the regulatory authority of national governments"); Deborah Arnott, *The Killer's Lobbyists*, GUARDIAN (London), May 15, 2003, at 30 (noting the monumental influence of the tobacco lobby in the developing world).

197. See *supra* note 186 and accompanying text.

198. While drafting the FCTC, there was concern that "the transnational tobacco conglomerates, which have tenaciously opposed the development of national tobacco control regulations, w[ould] wield their considerable economic and political power to obstruct any international legislation on tobacco control." Taylor, *supra* note 33, at 285; see also Henry A. Waxman, *The Future of the Global Tobacco Treaty Negotiations*, 346 NEW ENG. J. MED. 936, 938 (2002) (arguing against the United States's opposition, on behalf of tobacco corporations, to various FCTC provisions). Nevertheless, in crafting the FCTC, the WHO noted that the "ability of international organizations through the treaty-making process to encourage and assist nations in overcoming powerful and organized industry resistance to regulation is evidence of the important role that international law-making could play in efforts to regulate the activities of transnational tobacco conglomerates." TAYLOR & ROEMER, *supra* note 34, at 15. In adopting the FCTC through the World Health Assembly, the member states of the WHO were able to overcome intense industry resistance, succeed together where individual states had failed, and create global norms of tobacco control. For an analysis of the role of transnational tobacco corporations in attempting to influence the FCTC, see WORLD HEALTH ORG., TOBACCO COMPANY STRATEGIES TO UNDERMINE TOBACCO CONTROL ACTIVITIES AT THE WORLD HEALTH ORGANIZATION (2000); Collin et al., *supra* note

generate international norms for smoking cessation, it is imperative that they employ the human right to health. Grounding the Proposed Dependence Protocol upon human rights, as a benefit to states rich and poor, would enhance its global effectiveness:

Within this rights-based approach to smoking cessation, the Proposed Dependence Protocol should address the cultural acceptability of cessation interventions in developing states. As recognized in the overview of the Proposed Dependence Protocol, “widely varying circumstances in Parties will not allow an identical approach or a perfectly harmonized regulatory framework for treatment products.”¹⁹⁹ Further, both the composition of the cigarette and the individual smoker’s habits and pharmacological reactions to that cigarette vary by state and culture.²⁰⁰ Thus, in addressing these cross-cultural concerns, the WHO should create technical assistance programs to fund research to (1) engage in a comparative analysis of state approaches to treatment of tobacco dependence between developed and developing states; (2) investigate culturally relative aspects of tobacco control, seeking to develop culturally appropriate standards for cessation interventions; and (3) define a range of state cessation interventions that are consistent with implementation of the Proposed Dependence Protocol and developing states’ economic, social, and cultural norms. The WHO’s Tobacco Free Initiative has already begun such efforts, contracting national experts throughout the world to create “specific report[s] about the successful use of effective access to tobacco dependence treatment in tobacco control.”²⁰¹ However, these disparate, sporadically published efforts, while serving as models of best practices in cessation interventions, lack the coordination and resource centralization that an international technical assistance program would provide. Continuing the WHO’s efforts through a

13, at 271 (“Tobacco companies sought to influence policy by building relationships with WHO staff, including gaining contacts through hiring or offering future employment to officials, and placing industry consultants in positions within WHO.”).

199. Second Meeting of the Working Group, *supra* note 52, at 5.

200. Collin, *supra* note 12, at 64 (“Cigarettes of the same brand, but produced for differing markets, may vary significantly, for example, with respect to tar, nicotine and nitrosamine content.”) (citing N. Gray et al., *Variation Within Global Cigarette Brands in Tar, Nicotine, and Certain Nitrosamines: Analytic Study*, 9 TOBACCO CONTROL 351 (2000)); Caryn Lerman et al., *Individualizing Nicotine Replacement Therapy for the Treatment of Tobacco Dependence: A Randomized Trial*, 140 ANNALS INTERNAL MED. 426 (2004) (noting the effect of ethnicity and race on cessation intervention efficacy).

201. Tobacco Free Initiative, Terms of Reference of Agreement for Performance of Work (APW): For the Country Reports on Effective Access to Tobacco Dependence Treatment (July 2004) (on file with author).

coordinated international program, states, in implementing treatment interventions based on culturally relative clinical best practices, could use a proven global model while tailoring their national programs to meet the needs of different groups, with heightened attention paid to relevant indigenous communities, ethnic groups, racial minorities, and women.

In addition, member states should not disregard the need for equitable pricing of and access to pharmaceutical cessation interventions within the Proposed Dependence Protocol. Such a protocol should reaffirm member states' commitment to prioritizing the right to health above the rigid trade parameters of global capitalist structures while still permitting the lawful manufacture and parallel importation of generic treatments without subversion of the international market structures within TRIPS.²⁰² Shifting the locus of cessation interventions from the private interests of pharmaceutical corporations to the public interests of states reinforces state responsibility for alleviating the burden of tobacco-related disease under the right to health. Further, this generic pharmaceutical strategy would obviate the need for states to engage in lopsided differential pricing negotiations with pharmaceutical corporations on an "ad hoc, drug-by-drug basis."²⁰³ By empowering states to uphold the right to health through their own national public health strategies, rather than relying solely on the ever-vacillating humanitarianism of developed states and pharmaceutical corporations, the Proposed Dependence Protocol would permit states to take an accountable, democratic role in addressing the needs of those affected most by tobacco.

To assist these developing states in financing generic cessation interventions, the Proposed Dependence Protocol should develop a global fund through the World Health Organization. Although generic pricing may lessen the burden on low-income markets, any large-scale access to cessation treatments will require large-scale funding that is not available solely through national financing. By unifying the donations of

202. See Press Release, Infact, NGOs Denounce New Draft of Tobacco Control Treaty as Too Weak To Reverse Global Tobacco Epidemic (Jan. 15, 2003), <http://www.infact.org/011503drft.html> (arguing that the FCTC fails to prioritize public health over trade); see also *supra* note 170 and accompanying text.

203. See PETCHESKY, *supra* note 79, at 112. Petchesky notes:

At the national level, the result of "differential pricing" is that "each price cut for each drug in each country is negotiated separately," or that countries must defend their right to seek cheaper alternatives in lengthy litigations in the national courts. Meanwhile, months and years go by and millions more die needlessly. *Id.*

nongovernmental organizations and developed states through mandatory assessments, the World Health Organization can overcome member states' collective action difficulties to achieve coordinated global cessation goals.²⁰⁴ Developing states have long advocated the establishment of a trust fund to assist their tobacco control efforts, and, at their insistence, the WHO's Tobacco Free Initiative is researching this idea, along with other matters of tobacco control financing, in preparation for discussions by the Conference of the Parties.²⁰⁵ States should incorporate this discussion into their consideration of the Proposed Dependence Protocol.

States can work together to solve issues of tobacco cessation, aiding each other in disseminating the results of basic science and translating these results into new behavioral treatments and pharmacological regimens. Through a process termed "leap-frogging," this method of scientific dissemination allows "the adoption of measures in a forerunner state to serve as models elsewhere."²⁰⁶ The WHO's Tobacco Free Initiative, as the interim (and likely permanent) Secretariat for the FCTC, has an opportunity to coordinate a global laboratory network to research and evaluate tobacco cessation programs.²⁰⁷ Applying this research to country-

204. While such public-private partnerships would still rely on the humanitarian will of donors, a global trust fund could institutionalize these voluntary donations, providing a long-term, sustainable outlook on global tobacco cessation.

205. *Potential Sources and Mechanisms of Support, Report by the Secretariat*, Open-Ended Intergovernmental Working Group on the WHO Framework Convention on Tobacco Control, Agenda Item 9, WHO Doc. A/FCTC/IGWG/1/INF.DOC./1 (May 27, 2004), http://www.who.int/gb/fctc/PDF/igwg1/FCTC_IGWG1_ID1-en.pdf (exploring potential sources and mechanisms for financing the implementation of the FCTC).

206. Collin, *supra* note 12, at 83 (citing Framework Convention Alliance, Briefing Paper for the 2nd Meeting of the Intergovernmental Negotiating Body of the Framework Convention on Tobacco Control: Comments on the Chair's Text (Mar. 2001), <http://www.fctc.org/FCTCfca.shtml>); Taylor, *supra* note 104, at 501 (noting that "rapid worldwide dissemination of recent advances in scientific knowledge and technology has advanced international agreement and action by providing the evidence base and the technological tools needed for effective national action and international cooperation").

207. The WHO has already experimented with global research consortia, bringing together scientists from around the world for its Scientific Advisory Committee on Tobacco Product Regulation, a group that has published six detailed recommendations on the technical aspects of regulating tobacco products. In addition, the WHO has recently published an ambitious agenda for global tobacco research under the auspices of the FCTC. WORLD HEALTH ORG., BUILDING BLOCKS FOR TOBACCO CONTROL: A HANDBOOK 274-79 (2004); *see also* FCTC, *supra* note 1, art. 22 ("The Parties shall cooperate directly or through competent international bodies to strengthen their capacity to fulfill the obligations arising from this Convention, taking into account the needs of developing country Parties and

specific policy interventions, the WHO, using rapid assessment procedures,²⁰⁸ may quickly assess regional tobacco cultures and help design culturally appropriate interventions. Through such collaborative cessation efforts, those global efforts originally envisioned by the framers of the FCTC, states can fight together in battling back against the scourge of tobacco.

CONCLUSION

Even though smoking cessation is not perceived to be the most pressing issue facing many states, it is—based upon its life-saving potential alone—a fundamental component of the right to health. Bolstered by the authoritative force of the FCTC, states have a unique opportunity to realize their obligations under the right to health to aid those addicted to nicotine. Cessation is the goal; the right to health is the key to achieving that goal. Adding a clear tobacco dependence protocol to the FCTC would give states direction in fulfilling their human rights obligations toward tobacco cessation. Yet the achievement of a protocol is not an end in itself; it is the beginning of a progressive evolution of the right to health to include obligations for tobacco cessation and life-saving interventions for other public health crises.

Even in its success, cessation is not a panacea for the ills of tobacco, but rather a synergistic complement to the other tobacco-control approaches employed by the FCTC. Smoking cessation can save millions of

Parties with economies in transition. Such cooperation shall promote the transfer of technical, scientific and legal expertise and technology, as mutually agreed, to establish and strengthen national tobacco control strategies, plans and programmes . . .”).

208. With a mixture of objective observations and brief interviews concerning a specific cultural issue, rapid assessment procedures (RAPS) allow health anthropologists to provide “timely qualitative descriptions of the research setting which can be used to assess the progress of the intervention programme.” Roberta L. Hall et al., *Rapid Assessment Procedures To Describe Tobacco Practices at Sites Managed by Indian Tribes*, 4 TOBACCO CONTROL 156, 156 (1995). For an example of a tobacco cessation RAP conducted in a developing state, see Deborah Ossip-Klein, Understanding the Culture of Tobacco Use in the Dominican Republic Using Rapid Assessment Procedures (RAPS), Lecture as part of the Tobacco Seminar Series, Mailman School of Public Health, Columbia University (Mar. 11, 2004) (on file with author); see also David Seddon et al., *Developing Guidelines for Policy Research: Recommendations for Future Approaches and Methods*, in CASE STUDY REPORT: GLOBAL ANALYSIS PROJECT ON THE POLITICAL ECONOMY OF TOBACCO CONTROL IN LOW AND MIDDLE-INCOME COUNTRIES 105 (J. Patrick Vaughan et al. eds., 2000) (developing a programme for political economic analysis in implementing tobacco policies pursuant to the FCTC).

lives and bring every human being closer to the enjoyment of the highest attainable standard of health. Without smoking cessation programs, the positive health effects of the FCTC will not be felt for at least a generation, with FCTC programs offering little salvation from the steady and sustained death of current smokers. But it is the near-term benefit of cessation, the denormalization of smoking, that makes such interventions so politically perilous, with effective cessation programs resulting in an immediate decrease in tobacco consumption and sales. Consequently, governments and nongovernmental advocates should expect no greater intransigence from transnational tobacco corporations than when they explore state and international cessation efforts. With transnational tobacco corporations using their corporate leverage to block such life-saving measures, effective international mobilization will be needed to thwart the impertinence of the tobacco industry.

The success of the FCTC has heralded new mechanisms for collective action to challenge global threats to public health. For this globalization of public health to take hold, the FCTC precedent cannot fail to protect those most vulnerable. The FCTC exposes the silent pandemic of tobacco by chronicling efforts states may take to discourage the underlying determinants of smoking. Yet the FCTC forsakes those addicted to nicotine, offering no positive message to those trapped by their dependence on tobacco. Through a cessation protocol to the FCTC, states can act pursuant to the right to health, develop interventions to encourage cessation, and create the conditions necessary to foster dignity and hope.

ANNEX: PROTOCOL ON THE TREATMENT OF TOBACCO DEPENDENCE²⁰⁹

PROPOSED TECHNICAL COMPONENTS OF A PROTOCOL ON THE TREATMENT OF TOBACCO DEPENDENCE: AN OUTLINE OF BASIC OBLIGATIONS AND CONTROL MEASURES

Overview

This Protocol should create a basic duty to establish treatment measures that are practical, effective, cost-effective and available to all who require them. However, widely varying circumstances in Parties will not allow an identical approach or a perfectly harmonized regulatory framework for treatment products. In order that the measures taken constitute a coherent and systematic approach, the Parties should formulate a national programme. The national programme would be reported to an appropriate body of the Convention or Protocol. Technical assistance would be provided under the auspices of the Convention or Protocol to facilitate the creation and implementation of national programmes based on sound scientific evidence and best practice.

Preamble

The Parties to this Protocol,

Recalling that the objective of the framework convention on tobacco control includes the reduction of tobacco use,

Recalling that tobacco dependence is classified as a disease under the International Classification of Diseases (ICD-10), and that nicotine addiction is classified as a disease under the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV),

Recognizing that treatment of tobacco dependence reduces mortality and morbidity,

Recognizing also that treatment of tobacco dependence is effective across a wide range of settings,

Acknowledging that treatment of tobacco dependence is a cost-effective intervention,

Recognizing that in order to reduce mortality in the short term current smokers need to be encouraged to quit,

Aware that a high percentage of tobacco users wish to quit,

Confirming that cessation programmes must be gender sensitive,

209. Second Meeting of the Working Group, *supra* note 52. This prospective protocol, first developed by WHO's Tobacco Free Initiative, forms the basis of the author's call for a FCTC protocol to address smoking cessation, to be drafted by the Permanent Secretariat of the FCTC and adopted by the Conference of the Parties.

Concerned that tobacco dependence is a form of addiction and that current success rates of attempts to quit without any form of support are low,

Recognizing the important roles of medical doctors, nurses, pharmacists, social workers, community workers, and other professional groups in the treatment of tobacco dependence,

Have agreed as follows:

Definitions

(Explanatory note: Definitions are usually added late in the negotiation process, when it is apparent, in light of the rest of the text, what terms need to be defined. Moreover, some definitions may be included in the framework convention and be applicable to protocols. Therefore, no proposed draft text is suggested, except for a possible technical definition of “tobacco dependence treatment.”)

Tobacco dependence treatment – includes (singly or in combination) behavioural and pharmacological interventions such as education, brief counselling and advice, intensive support, administration of pharmaceuticals or other interventions that contribute to reducing and overcoming tobacco dependence in individuals and in the population as a whole.

Objective

1. The objective of the Protocol is to reduce and overcome individual dependence on tobacco by ensuring that tobacco users have access to appropriate [and affordable][costeffective] treatment for tobacco dependence, and thereby mitigating the health, welfare, economic and development burdens on individuals, families, communities and governments created by tobacco use.

Section I – Basic obligations

2. Each Party shall take all practical, effective and cost-effective measures to treat tobacco dependence and to promote cessation of tobacco use, taking into account local circumstances and priorities.

3. Each Party shall develop a national programme for the delivery and assessment of measures taken under [paragraph 2].

4. Taking into account local circumstances, each Party shall undertake the following measures:

(1) treatment of tobacco dependence within the national health care and social welfare systems;

(2) routine advice on and support for tobacco cessation by health professionals, including medical doctors, health practitioners, nurses, pharmacists, community workers and social workers based in primary care;

(3) development, implementation and promotion of the use of specialized services such as clinics, pharmacies, community-based support, telephone help lines, or Internet support;

(4) provision of pre- and postqualification education, training and information for health practitioners, community workers and social workers;

(5) promotional and education campaigns aimed at encouraging tobacco cessation;

(6) improved access to proven treatment interventions and products through both the private and public sector;

(7) removal [where appropriate or justified][when feasible] of economic barriers to treatment;

(8) removal of regulatory barriers in order to improve access to products for tobacco dependence treatment consistent with the protection of public health and sound science;

(9) fast-track approval of new proven products for tobacco dependence treatment consistent with protection of public health and sound science;

(10) public funding of proven behavioural and pharmacological treatments of tobacco dependence;

(11) integration of tobacco cessation treatments into reproductive health programmes such as the “safe motherhood” programme.

5. The Conference of the Parties shall take into account the particular needs of developing countries and assist in improving their national capacities and capabilities to participate in the measures referred to in [paragraphs 2 and 3] above.

Section II – Exchange of information and provision of technical support

6. Each Party shall cooperate in exchange of information and skills relevant to meeting the objectives of the Protocol. Each Party in a position to do so shall include in its national programme measures to be taken, if any, to assist other Parties in meeting the objectives of this Protocol either bilaterally or under the auspices of the Convention or Protocol.

7. The Conference of the Parties, at its first meeting, shall consider the establishment of a technical body, *inter alia*, to assist the Parties in undertaking effective cooperation and exchange of information and skills, and to determine guidelines for common statistical approaches to facilitate comparability of data gathered, taking existing surveillance systems into account.

Section III – National reports

8. Each Party shall communicate its national programme and report of

measures taken to implement the present Protocol to the [Conference of the Parties][Secretariat] of the [Convention][Protocol] within [...] months of the entry into force of this Protocol and [...] months before each meeting of the Conference of the Parties to the [Convention][Protocol].

9. Each Party shall undertake progressively, as a part of an integrated national surveillance system, to gather basic statistical data on tobacco cessation treatment services and products; to collect data on the availability of, access to and usage of tobacco dependence treatments; to gather data on their costs and effectiveness; and to include all these data in the planning for its national programme [paragraph 3].

Section IV – Research and development

10. Each Party shall support and further develop, as appropriate, national and international programmes and networks or organizations aimed at defining, conducting, assessing and financing research and data collection, taking into account the need to minimize duplication of effort.

11. Each Party shall, in accordance with its capabilities and the means at its disposal, initiate and cooperate in, directly or through competent international bodies, the further development of effective and necessary means for the treatment of tobacco dependence and ensure that such means for treating tobacco dependence are widely available and affordable, particularly in developing countries. Such research should be linked to improving access to pharmaceutical treatments for tobacco dependence as an important component of elaborating a sustainable national health sector strategy.

12. Each Party shall encourage and support research, development and demonstration activities related to:

- improving the effectiveness of tobacco dependence treatments;
- improving the cost-effectiveness of tobacco dependence treatment;
- improving the access to tobacco dependence treatment; appropriate frameworks and settings for delivery of tobacco dependence treatments;
- effective partnerships between public, private and nongovernmental bodies involved in tobacco dependence treatment;
- appropriate regulatory approaches for tobacco dependence treatments.

Pharmaceutical Arbitrage: Balancing Access and Innovation in International Prescription Drug Markets

Kevin Outterson, J.D., LL.M.*

INTRODUCTION

The price of prescription drugs lies at the heart of two major public health issues: distributing antiretroviral medicines for use against the global AIDS epidemic and purchasing medications from Canada by U.S. consumers using the Internet. Both situations highlight the need to reduce financial barriers to access to medications, while maintaining incentives to promote pharmaceutical innovation.

For better or worse, the World Trade Organization's (WTO) Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS)¹ is a global nexus for drug access issues.² In TRIPS-related

* Associate Professor of Law, West Virginia University College of Law. Special thanks to the American Society of Law, Medicine & Ethics September 2003 Health Law Young Scholars Workshop and the West Virginia University College of Law Faculty Workshop for helpful comments and suggestions. This research was supported by a Hodges Research Grant from the West Virginia College of Law. An early draft of this Article was prepared during my tenure as a Visiting Fellow at the Lauterpacht Research Centre for International Law at the University of Cambridge. Helpful comments were received from many people, including Frances Miller, Ben Moulton, Charity Scott, Ana Iltis, Timothy D. McBride, Thomas Greaney, Jesse Goldner, Sidney Watson, Sandra Johnson, Nicolas Terry, Aidan Hollis, and Andrew Farlow, as well as from participants at the New Economic Windows 2004 conference in Salerno, the symposium at the University of Connecticut School of Law (October 2004), The Canadian Conference on International Health (October 2004), and the Global Forum for Health Research Forum 8 Conference in Mexico City (November 2004).

1. See Agreement on Trade-Related Aspects of Intellectual Property Rights, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, art. 27.1, LEGAL INSTRUMENTS—RESULTS OF THE URUGUAY ROUND vol. 31, 33 I.L.M. 81 (1994) [hereinafter TRIPS or TRIPS Agreement]. The United States implemented the WTO agreements in the Uruguay Round Agreements Act, Pub. L. No. 103-465, 108 Stat. 4809 (1994).

2. The story of how the WTO TRIPS Agreement became the de facto forum for these

discussions, two sets of arguments are usually forwarded. Some argue that pharmaceutical prices are necessarily high because innovation is expensive.³ They posit that the research and development (R&D) enterprise must be nurtured by high prices to yield the next generation of breakthrough therapies.⁴ Others counter that much of the profits going to pharmaceutical companies⁵ are used for marketing and other expenses

issues has been told by many authors. Among the best accounts are books by Peter Drahos and John Braithwaite, Susan Sell, and an article by Laurence Helfer. PETER DRAHOS WITH JOHN BRAITHWAITE, *INFORMATION FEUDALISM: WHO OWNS THE KNOWLEDGE ECONOMY?* (2002); SUSAN K. SELL, *POWER AND IDEAS: NORTH-SOUTH POLITICS OF INTELLECTUAL PROPERTY AND ANTITRUST* (1998); SUSAN K. SELL, *PRIVATE POWER, PUBLIC LAW: THE GLOBALIZATION OF INTELLECTUAL PROPERTY RIGHTS* (2003); Laurence R. Helfer, *Regime Shifting: The TRIPS Agreement and New Dynamics of International Intellectual Property Lawmaking*, 29 YALE J. INT'L L. 1 (2004).

3. See, e.g., Harvey E. Bale Jr., *Patents, Patients and Developing Countries: Access, Innovation and the Political Dimensions of Trade Policy*, in THE ECONOMICS OF ESSENTIAL MEDICINES 100, 102-04 (Brigitte Granville ed., 2002) [hereinafter ECONOMICS OF ESSENTIAL MEDICINES]. Dr. Bale is the Director-General of the International Federation of Pharmaceutical Manufacturers Association.

4. ANDY SCHNEIDER, TAXPAYERS AGAINST FRAUD EDUCATION FUND, REDUCING MEDICARE AND MEDICAID FRAUD BY DRUG MANUFACTURERS: THE ROLE OF THE FALSE CLAIMS ACT 47 (2003) ("Pharmaceutical manufacturers have long maintained that government price controls will thwart the development of vital new drugs with the potential to cure diseases and relieve human suffering. The desired alternative, they argue, is a vigorous free market, with prices set through negotiations between buyers and sellers. For this market to work effectively, manufacturers contend, they must retain the right to keep their prices confidential from competitors.").

5. In this Article, the terms "pharmaceutical companies" and "PhRMA companies" refer to the research-based pharmaceutical companies that are members of the PhRMA trade association. Pharmaceutical companies have traditionally been categorized as either research companies (e.g., Pfizer, Merck) or generic companies without significant research programs (e.g., Mylan Labs, Cipla Ltd.). The United States trade association of research pharmaceutical companies is the Pharmaceutical Research and Manufacturers of America (PhRMA). See Pharmaceutical Research and Manufacturers of America, at <http://www.phrma.org> (last visited Oct. 20, 2004). The international trade association of PhRMA company groups is the International Federation of Pharmaceutical Manufacturers Associations (IFPMA). See International Federation of Pharmaceutical Manufacturers, at <http://www.ifpma.org> (last visited Oct. 20, 2004). Generic drug companies have their own trade associations, such as GPhA, at <http://www.gphaonline.com> (last visited Oct. 20, 2004). In recent years, these distinctions have blurred as research companies have invested in generic subsidiaries and as generic companies have begun substantial research programs. It may sometimes be more accurate to describe research or generic lines of business, rather than companies per se.

rather than for R&D⁶ and that without affordable access, innovation is a cruel taunt.⁷ New wonder drugs will not improve health unless patients are actually able to receive them. A pill you cannot afford is neither safe nor effective. Medicines, according to this argument, are not normal market goods to be distributed primarily to the wealthy.⁸

Nowhere are the arguments for the equitable distribution of medicines made with more force than in the AIDS treatment crisis. Differential pricing is one response to the tension between innovation and access with regard to AIDS medications: It permits antiretroviral drugs to be sold cheaply or donated in low income countries, while maintaining high prices in markets like the United States.⁹ In theory, high prices in high income countries can support innovation, while lower prices in low income countries improve access. However, differences in pricing—and thus opportunities for arbitrage—do not always reflect direct or voluntary efforts to facilitate access in developing countries;¹⁰ they may also result from diverse systems of government regulation and intervention and corporate efforts to maximize profit. The price of drugs is affected by domestic intellectual property (IP) laws, pharmaceutical reimbursement systems, and other legal systems specific to each country. As a result, for example, patented pills in Australia are often cheaper than their equivalents in Canada, which are in turn often cheaper than those in the United States. These pricing gaps create the demand for cross-border pharmaceutical parallel trade, or pharmaceutical arbitrage. It is alleged that if such trade is left unchecked, it will significantly reduce the financial gains reaped in certain countries, most prominently the United States,

6. See, e.g., Marcia Angell, *The Truth About Drug Companies*, N.Y. REV. BOOKS, July 15, 2004, at 51.

7. See, e.g., Ellen 't Hoen, *TRIPS, Pharmaceutical Patents, and Access to Essential Medicines: A Long Way from Seattle to Doha*, 3 CHI. J. INT'L L. 27, 29-30 (2002).

8. MÉDECINS SANS FRONTIÈRES, MSF CAMPAIGN BROCHURE 5 (2004), <http://www.access-med-msf.org/documents/campaignbrochure2004.pdf> ("Medicines aren't just any consumer goods.").

9. A joint workshop was conducted in 2001 on differential pricing for essential drugs by the World Health Organization and the World Trade Organization, with substantial participation from global drug companies. The final report is published as WORLD HEALTH ORG., REPORT OF THE WORKSHOP ON DIFFERENTIAL PRICING AND FINANCING OF ESSENTIAL DRUGS: A WHO/WTO SECRETARIAT WORKSHOP (2001), <http://whqlibdoc.who.int/hq/2001/a73725.pdf>. A more concise summary of the final report was published in 2002. WHO-WTO, *Differential Pricing and the Financing of Essential Drugs*, in ECONOMICS OF ESSENTIAL MEDICINES, *supra* note 3, at 209-31.

10. This is sometimes referred to in this Article as voluntary differential pricing.

which provide financial support for global R&D innovation¹¹ and, thus, may undermine voluntary differential pricing schemes (e.g., AIDS initiatives) that benefit low income countries.¹² So long as R&D costs continue to be partially funded by sales revenues,¹³ the conventional wisdom holds that pharmaceutical arbitrage is a major threat to both differential pricing and innovation. Preventing pharmaceutical arbitrage from low income markets into high income markets is generally viewed as the linchpin of this analysis.

Thus, in the context of differential pricing, pharmaceutical arbitrage is becoming an increasingly prominent subject of debate; in particular, fear of arbitrage is being used to justify expanding pharmaceutical IP rights and related powers of appropriation.¹⁴ This Article explores key functions of pharmaceutical arbitrage, including its impact on access and innovation and its implications for the implementation of the TRIPS Agreement and other government interventions affecting pharmaceutical prices and distribution.

Part I of the Article establishes a theoretical framework for understanding pharmaceutical markets and innovation, using the heuristic device of *optimal pharmaceutical rents* to explore pharmaceutical arbitrage.¹⁵

11. Tom Blackwell, *Canada's Drug Pricing Unfair, U.S. Alleges: Pharma Companies Back Plan To Restrict Cross-Border Sales*, NAT'L POST (TORONTO), May 3, 2004, at A6; Gardiner Harris, *Cheap Drugs from Canada: Another Political Hot Potato*, N.Y. TIMES, Oct. 23, 2003, at C1. *But see* ALAN SAGER & DEBORAH SOCOLAR, DO DRUG MAKERS LOSE MONEY ON CANADIAN IMPORTS? (Boston Univ. Sch. of Pub. Health, Data Brief No. 6, 2004), <http://www.healthreformprogram.org>.

12. PATRICIA M. DANZON & ADRIAN TOWSE, DIFFERENTIAL PRICING FOR PHARMACEUTICALS: RECONCILING ACCESS, R&D AND PATENTS 28-29 (AEI-Brookings Joint Ctr. for Regulatory Studies, Working Paper No. 03-7, 2003).

13. A prominent alternative formulation would be the Hubbard-Love R&D Treaty, which does not rely on retail sales to recover R&D costs. Tim Hubbard, *Alternatives to the Price System*, Presentation at Columbia University (Dec. 4, 2003), at http://www.earthinstitute.columbia.edu/cgsd/accesstomedicines_papers.html; James Love, *A New Trade Framework for Global Healthcare R&D*, Presentation at Columbia University (Dec. 4, 2003), at http://www.earthinstitute.columbia.edu/cgsd/accesstomedicines_papers.html.

14. Powers of appropriation are those mechanisms, including legal rights and entitlements, that allow individuals or entities to control the distribution of (and thus to capture) the value created. *See, e.g.*, David Ellerman, *Introduction to Property Theory* (Apr. 2004) (unpublished manuscript, on file with the *Yale Journal of Health Policy, Law, and Ethics*), <http://www.economics.ucr.edu/seminars/spring04/Intro-to-Prop-Theory.pdf>.

15. In this Article, the term *rents* is generally used in lieu of *patent rents* because in pharmaceutical markets, many legal tools are utilized to make returns on investment appropriable to the innovator in addition to patent law. *See infra* Section I.C. The phrase

In the absence of definitive data on pharmaceutical R&D, the heuristic can offer a guide to policymakers attempting to balance access and innovation. Part II of the Article applies this framework to two situations: the global pricing of antiretroviral drugs and the issue of Canadian-U.S. cross-border arbitrage.

The primary conclusions of this Article fall into two clusters. First, the heuristic indicates that several forms of pharmaceutical arbitrage are beneficial, delivering lower prices to consumers without harming innovation. Arbitrage within and between high income markets, such as the Canadian Internet sales to the United States, will not harm innovation if pharmaceutical rents remain supra-optimal. Pharmaceutical industry claims of sub-optimality must be backed with full transparency to allow for public evaluation of pricing, production cost, and profitability data throughout the world.

More broadly, the heuristic indicates that optimal economic incentives for innovation can be maintained while providing low income populations with greatly expanded access to patented medicines. Unlike physical property, pharmaceutical innovation is generally nonrival.¹⁶ Therefore, in markets which are unlikely to contribute importantly to global pharmaceutical rents, the shackles of intellectual property law and other forms of appropriation are both unnecessary and dangerous; such laws should be set aside in these circumstances, permitting the broadest possible dissemination of pharmaceutical innovation. Practical experience suggests that voluntary differential pricing is unlikely to deliver needed medications at the lowest possible marginal cost.¹⁷ Low transaction cost compulsory licenses are preferable and are consistent with the needs for innovation.¹⁸

Furthermore, while much of the current debate is focused on AIDS (and to a lesser extent on tuberculosis and malaria), the analysis in this Article is not limited to these conditions. Consistent with global optimal pharmaceutical rents, access can be expanded to all categories of global diseases, including cancer and heart disease, without damaging innovation.

pharmaceutical rents is thus meant to capture all of the various ways, including regulation-based market exclusivity, in which pharmaceutical innovators appropriate rents. When the term *patent rents* is used herein, the narrower meaning is intended.

16. Nonrival goods can be utilized simultaneously by multiple users without risk of exhaustion. See JOHN B. TAYLOR & IMAD MOOSA, *MACROECONOMICS* (2d ed. 2002). For further discussion of this term, see *infra* text accompanying note 24.

17. See *infra* note 151 and accompanying text.

18. The need to improve compulsory licensing procedures is discussed in Part II.

In the second cluster of conclusions, I determine that the threat of pharmaceutical arbitrage is overstated and rarely observed empirically. This Article describes the legal and commercial frameworks which generally obstruct arbitrage, and argues that the most dangerous threat to innovation and public health comes from counterfeit medications, not from arbitrage. Resources now being expended to limit diversion in donor programs and differential pricing schemes could be more profitably reallocated to anti-counterfeiting initiatives within high income markets. A prime example of a misdirected anti-arbitrage effort is the initiative within the President's Emergency Plan for HIV/AIDS Relief (PEPFAR) to establish its own supply chain and procurement policies.

I. THE THEORY OF PHARMACEUTICAL ARBITRAGE

A. *The Innovation Theory of IP Law*

From ancient times, law and social conventions have supported the right to exclude—a fundamental component of the concept of personal and real property.¹⁹ Persons investing in the production of goods are able to reap a reward for their efforts because, in part, the law creates a property right in the goods produced. This property right is somewhat exclusive, meaning that other persons cannot take the property without consent, due process, or some important public policy.²⁰ In the language of economics, goods and services are “appropriable.”²¹ At common law, knowledge was not considered personal property,²² perhaps because the

19. See, e.g., *Exodus* 20:15 (NRSV) (“You shall not steal.”). The right to exclude others from an individual's or group's real property developed much later and is not yet fully ascendant in some communities.

20. In physical property, the right to exclude is subject to many exceptions and conditions; critiques of analogies to intellectual property are yielding some interesting research. See, e.g., MARK A. LEMLEY, PROPERTY, INTELLECTUAL PROPERTY, AND FREE RIDING 3-17 (John M. Olin Program in Law & Econ., Working Paper No. 291, 2004); STEWART E. STERK, WHAT'S IN A NAME? THE TROUBLESOME ANALOGIES BETWEEN REAL AND INTELLECTUAL PROPERTY 1-3 (Jacob Burns Inst. for Advanced Legal Studies, Benjamin N. Cardozo Sch. of Law, Working Paper No. 88, 2004).

21. Use of the term “appropriable” can be found in an array of works. See, e.g., James J. Anton & Dennis A. Yao, *Expropriation and Inventions: Appropriable Rents in the Absence of Property Rights*, 84 AM. ECON. REV. 190 (1994).

22. See, e.g., *Wheaton v. Peters*, 33 U.S. (8 Pet.) 591, 657 (1834). The first English copyright statute was the Statute of Anne, 8 Ann., c. 19 (1710), and the first English “patent” statute was the Statute of Monopolies, 21 Jac. 1, c. 3 (1624). See also Carle Hesse,

use of information is subject to at least two peculiar characteristics. First, knowledge is generally inappropriable or nonexcludible: It is typically more difficult to exclude other persons from using knowledge than physical property.²³ Second, knowledge is nonrival: While physical goods like corn or wheat are exhausted when used, knowledge may be used without exhaustion.²⁴

The nonrival nature of knowledge permits its widest possible dissemination without creating shortages, a potential boon for humanity.²⁵ But, the fly in the ointment is appropriation. If *homo econimus* understands that the fruits of research will not be appropriable, then the market offers no financial incentive to innovate. Others will gladly use the innovation without compensating the innovator. The economic model predicts that when the innovator cannot capture the positive externality (or consumer surplus), the incentive to innovate is undermined.

However, this model is overly pessimistic. Inventive knowledge grew in the centuries prior to the adoption of patent law; important books were

The Rise of Intellectual Property, 700 B.C. – A.D. 2000: An Idea in the Balance, DAEDALUS, Spring 2002, at 26-45 (tracing the epistemological foundations of intellectual property). The innovation theory is not the sole justification for patent law, but it is the dominant one in Anglo-American jurisprudence. Another possible ground for patent law is the contract or disclosure theory, which posits that patents are socially preferable over trade secrets due to the socially useful disclosure function. See Vincenzo Denicolo & Luigi Alberto Franzoni, *The Contract Theory of Patents*, 23 INT'L REV. L. & ECON. 365, 366-68 (2004). In pharmaceuticals, the marketing approval process requires disclosure in any event, making the contract theory less applicable.

23. This Article uses the terms inappropriable and nonexcludible interchangeably.

24. While knowledge is not destroyed through use, it may lose value because it is inappropriable. For example, market-moving financial information loses its value quickly, particularly as market participants act on the information. This is a function of inappropriability, rather than exhaustion or rivalry. From a societal perspective, knowledge does not lose value through use, but adds to the public domain.

25. The point is occasionally overlooked. In his critique of the consequences of the TRIPS Agreement, for example, Alan Sykes underemphasizes the nonrival nature of pharmaceutical patents by analogizing compulsory licensure to physical expropriation. Alan O. Sykes, *TRIPS, Pharmaceuticals, Developing Countries, and the Doha "Solution,"* 3 CHI. J. INT'L L. 47, 56 (2002); see also William M. Landes & Richard A. Posner, *Indefinitely Renewable Copyright*, 70 U. CHI. L. REV. 471, 484-86 (2003) (arguing that some forms of IP are rival, particularly trademarks and personal likenesses). Trademarks and personal likenesses indicate origin rather than being knowledge per se. Pharmaceutical knowledge is nonrival in the classic sense, although nonrival use will certainly undercut monopoly pricing and affect *ex ante* innovation incentives.

written before the Statute of Anne.²⁶ This can at least be partially explained by non-economic motives for research, such as curiosity or personal achievement.²⁷ In most industries, patents play a relatively minor role in promoting innovation.²⁸

Nevertheless, pharmaceutical research companies strongly embrace this neo-classical innovation model.²⁹ They argue that without IP laws first

26. The British Statute of Anne is considered the first copyright law. Statute of Anne, 1710, 8 Ann., c. 19 (Eng.). Today's industrialized countries are relatively recent converts to the cause of strong IP laws. DRAHOS WITH BRAITHWAITE, *supra* note 2, at 29-38.

27. The open source movement in science is built upon such factors, as articulated by several leading scientists. Tim Hubbard & James Love, *Medicines Without Barriers: From the Human Genome Project to Open Development Models for Medical R&D*, NEW SCIENTIST, June 14, 2003, at 29; Stephen M. Maurer et al., *Finding Cures for Tropical Diseases: Is Open Source an Answer?*, in BIOTECHNOLOGY: ESSAYS FROM ITS HEARTLAND 33-37 (Lynn Yarris ed., 2004), <http://www.salilab.org/publications>; Sir John Sulston, Open and Collaborative Movements in Science, Presentation at the Trans-Atlantic Consumer Dialogue Future of WIPO Workshop, Geneva (Sept. 13, 2004).

28. For most industries, it appears that patents play a relatively modest role in making invention non-appropriable by free riders. See, e.g., ASHISH ARORA ET AL., R&D AND THE PATENT PREMIUM 4, 34-35 (Nat'l Bureau of Econ. Research, Working Paper No. 9431, 2003) ("Empirical work also suggests that the inducement provided by patents for innovation is small."); WESLEY M. COHEN ET AL., PROTECTING THEIR INTELLECTUAL ASSETS: APPROPRIABILITY CONDITIONS AND WHY U.S. MANUFACTURING FIRMS PATENT (OR NOT) 2, 24-25 (Nat'l Bureau of Econ. Research, Working Paper No. W7552, 2000) (finding that forty years of empirical data demonstrates that patents do not improve innovation, with exceptions in pharmaceuticals, and concluding that patents are not the most significant mechanisms for appropriating returns to innovation in most industries, with secrecy, lead time, and complimentary capabilities leading); Richard C. Levin et al., *Appropriating the Returns from Industrial Research and Development*, in 3 BROOKINGS PAPERS ON ECONOMIC ACTIVITY 783 (Martin Neil Baily & Clifford Winston eds., 1987); Richard C. Levin, *A New Look at the Patent System*, 76 AM. ECON. REV. 199, 200-01 (1986); Edwin J. Mansfield, *Patents and Innovation: An Empirical Study*, 32 MGMT. SCI. 173 (1986). In pharmaceuticals, secrecy is not an option with the public drug application process, and the evidence strongly suggests a link between patents and innovation. ARORA ET AL., *supra*, at 4-5, 35. Arora's study found a significant patent premium (i.e., a positive return on investment), particularly in biotechnology, medical instruments, and drugs. *Id.* at 30, 34-35.

29. One prominent source on R&D expenditures by PhRMA companies is Joseph A. DiMasi et al., *The Price of Innovation: New Estimates of Drug Development Costs*, 22 J. HEALTH ECON. 151 (2003). These claims are defended vigorously by PhRMA and its members. See, e.g., ERNST & YOUNG LLP, PHARMACEUTICAL INDUSTRY R&D COSTS: KEY FINDINGS ABOUT THE PUBLIC CITIZEN REPORT (2001), <http://www.phrma.org/mediaroom/press/release//2001-08-11.277.pdf>; Sidney Taurel, *Hands Off My Industry*, WALL. ST. J., Nov. 3, 2003, at A14. Sidney Taurel is the president, chairman, and chief executive officer of Eli Lilly.

movers would incur all research costs (including failed programs), while free riders (subsequent movers such as generic drug companies) would benefit from significantly lower cost structures.

IP law offers an allegedly second-best solution³⁰ to this impasse—the Constitution’s favorite monopolies “promote the progress of science and useful arts, by securing for limited times, to authors and inventors the exclusive right to their respective writings and discoveries.”³¹ Currently, under U.S. federal law and the TRIPS Agreement, the patent period is not less than twenty years after filing.³²

The social costs of making pharmaceutical knowledge appropriable are generally three-fold. First, the cumulative effect of these laws allows the

30. See, e.g., TOMAS J. PHILIPSON & STÉPHANE MECHOULAN, INTELLECTUAL PROPERTY & EXTERNAL CONSUMPTION EFFECTS: GENERALIZATIONS FROM PHARMACEUTICAL MARKETS 3 (Nat’l Bureau of Econ. Research, Working Paper No. 9598, 2003) (“In the private case, it is well-understood that efficient competition ex-post leads to insufficient R&D incentives ex-ante, which is of course the common second-best rationale for patents.”); *id.* at 8, 14-15. For a timely recognition that a bare patent does not equal the clear right to exclude, see Mark A. Lemley & Carl Shapiro, *Probabilistic Patents*, J. ECON. PERSP. (forthcoming 2004) (manuscript at 19, on file with author). Lemley and Shapiro’s analysis is not specific to pharmaceuticals, where multiple patents and other appropriation strategies heighten the degree of exclusion. See *infra* Part I.

31. U.S. CONST. art. I, § 8, cl. 8. Of course, a bare patent does not grant market power if the invention is unimportant or easily substitutable. Kenneth W. Dam, *The Economic Underpinnings of Patent Law*, 23 J. LEGAL STUD. 247-51 (1994). Pharmaceutical patents of blockbuster drugs are a strong case of patents creating market power and may be more appropriately denominated as a monopoly. The pharmaceutical industry eschews the monopoly label, but nevertheless defends the patent system as essential to encourage R&D. One cannot have it both ways.

32. 35 U.S.C. § 154 (2000); TRIPS, *supra* note 1, art. 33. TRIPS permitted many developing countries to implement on a delayed basis. TRIPS, *supra* note 1, arts. 65-66. After extensions, most developing countries must implement the TRIPS Agreement by January 1, 2005, but the thirty “least developed countries” may defer full implementation for pharmaceutical products until 2016. *Declaration on the TRIPS Agreement and Public Health*, Doha WTO Ministerial 2001, WT/MIN(01)/DEC/2, ¶ 7 (Nov. 20, 2001) [hereinafter *Doha Declaration on TRIPS*]. Despite these concessions, all but three of Africa’s Least Developed Countries (LDCs) have already adopted patent laws for pharmaceuticals. PHIL THORPE, STUDY ON THE IMPLEMENTATION OF THE TRIPS AGREEMENT BY DEVELOPING COUNTRIES 1 (Comm. on Intellectual Prop. Rights, Study Paper 7 (circa 2004)). TRIPS merely sets minimum periods of IP protection; the United States can still unilaterally extend patent protection, and has done so with copyright. WTO Members are also free to negotiate so-called “TRIPS-plus” agreements with additional provisions requiring protections in excess of the TRIPS Agreement’s minimum standards.

innovator to charge a higher price under monopolistic conditions. James Love, Director of the Center for Consumer Project on Technology, estimates the deadweight cost at \$400 billion per year.³³ Second, these higher prices hinder medical access, directly impacting the health of many low income people globally.³⁴ Finally and most generally, appropriation, by necessity, delays the entry of knowledge into the public domain and thus may hinder cumulative innovation.³⁵

The perceived tension between the development and dissemination of knowledge permeates the most compelling issues in pharmaceutical IP policy. Patent doctrines such as scope,³⁶ experimental use,³⁷ and fair use³⁸ are also battlegrounds in the struggle between innovation and the public domain.³⁹ This Article locates additional laws in the policy battleground as well. If too many laws support appropriation (i.e., excessive IP rights and other excessive restrictions on nonrival use), the system needlessly raises costs and restricts access to important pharmaceuticals.⁴⁰ Too few might throttle the R&D enterprise, and society might forgo valuable qualitative improvements. It is far from clear that current policy strikes an appropriate balance. At the celebration of the tenth anniversary of the TRIPS

33. James Love, Statement of Essential Inventions, Inc. to the Commission on Intellectual Property Rights, Innovation and Public Health (Apr. 5, 2004).

34. See *infra* Subsection II.A.1.

35. See *infra* notes 130-131 and accompanying text.

36. Robert P. Merges & Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM. L. REV. 839 (1990) (examining the potential role of patent breadth in fine tuning the efficiency of the patent system). Many economic studies examine elements of this question. See, e.g., WILLIAM D. NORDHAUS, *INVENTION, GROWTH, AND WELFARE: A THEORETICAL TREATMENT OF TECHNOLOGICAL CHANGE* 70-90 (1969) [hereinafter NORDHAUS, *INVENTION, GROWTH & WELFARE*]; William D. Nordhaus, *The Optimum Life of a Patent: Reply*, 62 AM. ECON. REV. 428 (1972) [hereinafter Nordhaus, *The Optimum Life of a Patent*]; F.M. Scherer, *Nordhaus' Theory of Optimal Patent Life: A Geometric Reinterpretation*, 62 AM. ECON. REV. 422-27 (1972) [hereinafter Scherer, *Optimal Patent Life*]. For a recent example, see PHILIPSON & MECHOULAN, *supra* note 30, at 8-13.

37. Rebecca Eisenberg, *Patents and the Progress of Science: Exclusive Rights and Experimental Use*, 56 U. CHI. L. REV. 1017 (1989); Rebecca Eisenberg, *Proprietary Rights and the Norms of Science in Biotechnology Research*, 97 YALE L.J. 177 (1987).

38. Maureen A. O'Rourke, *Toward a Doctrine of Fair Use in Patent Law*, 100 COLUM. L. REV. 1177 (2000).

39. Dam, *supra* note 31, at 261-68.

40. This point assumes that increased consumption of patented pharmaceuticals creates net positive externalities, i.e. that society benefits from increased access and consumption of the drug. PHILIPSON & MECHOULAN, *supra* note 30, at 9.

Agreement, Pascal Lamy, Director of DG-Trade, noted:

IPRs are justified by their societal purpose: they constitute a public policy tool to encourage innovation and creativity. These are the ends, and the patents and copyrights granted to innovators and creators are the means to achieve it. But the hierarchy of ends and means does not end here. Indeed, the encouragement of innovation and creativity is itself serving higher purposes: economic, social and cultural development that should benefit all.

So, international intellectual property policy is a question of striking the right balance between private interests, their public policy objective (access to knowledge) and other public goods. Should this public/private bargain be struck in the same way in all WTO Members? Not necessarily. Here the level of development and the national public policy objectives come into play.⁴¹

B. Differential Pricing and Pharmaceutical Arbitrage

1. Differential Pricing

In simple economic models, goods are sold at a single market-clearing price. In reality, clever selling firms realize that some customers will pay more than the market-clearing price. The selling firm increases its profit by selling each item at the highest price each particular buyer will pay. The economic literature identifies this process as price discrimination, which is synonymous with differential pricing for our purposes.⁴²

Indeed, differential pricing is common: The same product is frequently sold at different net prices to various buyers.⁴³ The seller charges

41. Pascal Lamy, The TRIPs Agreement 10 Years On, Speech to the International Conference on the 10th Anniversary of the WTO TRIPs Agreement (June 23, 2004), http://europa.eu.int/comm/commissioners/lamy/speeches_articles/spla233_en.htm.

42. Price discrimination is the term generally utilized in the economic literature but should not be confused with price discrimination under the Robinson-Patman Act, 15 U.S.C. §§ 13-13b, 21a (2000). This Article follows the usage most common in the essential medicines literature: differential pricing. Tiered pricing, equity pricing, and price segmentation are other terms occasionally used for pharmaceutical differential pricing. *See, e.g.*, DG TRADE, EUROPEAN UNION, TIERED PRICING FOR MEDICINES EXPORTED TO DEVELOPING COUNTRIES, MEASURES TO PREVENT THEIR RE-IMPORTATION INTO THE EC MARKET AND TARIFFS IN DEVELOPING COUNTRIES (EU Working Document, 2002).

43. This particular definition is found in LOUIS PHILIPS, THE ECONOMICS OF PRICE

what each market segment will bear.⁴⁴ A selling firm might attempt to differentiate its prices on an individual sale basis, a pure form of differential pricing which Pigou labeled first-degree price discrimination.⁴⁵ First-degree price discrimination is also known as perfect price discrimination, since it fully extracts all consumer surplus for the benefit of the producer.⁴⁶ In the case of pharmaceuticals, this would provide cash flow for innovation but would impair access through higher consumer cost. In reality, transaction costs almost always make first-degree differential pricing untenable: The seller's marginal costs of collecting and understanding all of the relevant factors for each buyer usually outweigh the gains in marginal revenue.⁴⁷ If the number of market segments is kept relatively small, however, the marginal revenue may exceed the marginal cost, resulting in second- or third-degree price discrimination.⁴⁸ In second-degree price discrimination, purchasers segment themselves into price levels. For example, railroad passengers choose either first, second, or third class seats and coupon clippers segment themselves into distinct markets. In third-degree price discrimination, the producer segments the market, generally using monopolistic power to distinguish the different prices customers are willing to pay. Global sales of patented pharmaceuticals offer examples of both second- and third-degree price discrimination.⁴⁹

DISCRIMINATION 6, 17 (1983).

44. The airline industry provides an oft-cited example. On almost every flight, passengers will have paid many different prices for the same service. The market has been segmented into multiple buyer groups, including business travelers, vacation travelers, frequent flyers, and last minute purchasers. See, e.g., ERNST R. BERNDT, AM. ENTERPRISE INST. FOR PUB. POLICY RESEARCH, UNIFORM PHARMACEUTICAL PRICING: AN ECONOMIC ANALYSIS 5-6, 9-10 (1994). However, it is worth noting that some, like Louis Philips, argue that the airline example is not technically an example of price discrimination, concluding that reserving a seat weeks in advance and buying a last minute ticket are different services. PHILIPS, *supra* note 43, at 9.

45. The classic description of first-, second-, and third-degree price discrimination is found in ARTHUR CECIL PIGOU, *THE ECONOMICS OF WELFARE* 321-47 (4th ed. 1920). A helpful summary of Pigouvian price discrimination may be found in PHILIPS, *supra* note 43, at 11-14.

46. It is perfect from the perspective of the selling firm, rather than the consumer. PHILIPS, *supra* note 43, at 158.

47. PIGOU, *supra* note 45, at 280.

48. See PIGOU, *supra* note 45; PHILIPS, *supra* note 43, at 12-13.

49. Examples of second-degree price discrimination include consumer selection of branded or unbranded drugs, the opportunity to apply for patient assistance programs, and monopsonistic price controls. Examples of third-degree price discrimination include voluntary differential pricing programs by manufacturers.

The primary focus of this Article is third-degree price discrimination, although I typically employ the more general term, differential pricing. The term “voluntary differential pricing” in this Article refers specifically to third-degree price discrimination, as distinguished from second-degree price discrimination such as price controls imposed by monopsonistic payor governments.

Differential pricing is endemic to pharmaceutical markets.⁵⁰ Pharmaceutical companies segment markets for differential pricing purposes, generally along efficient boundaries such as political borders or payor classes, with the support of legal institutions. Voluntary differential pricing exists among different countries⁵¹ and among different buyers or payor classes within countries.⁵² Second-degree differential pricing occurs when price controls are imposed.⁵³

2. Pharmaceutical Arbitrage

Pharmaceutical arbitrage is the theoretical nemesis of differential pricing.⁵⁴ While differential pricing assumes that the first purchaser is the

50. This is true, at least, in recent years. See *infra* Part II. But at least one Wall Street Journal editor is calling on PhRMA companies to abandon voluntary price discrimination for a single price in all developed countries. See Holman W. Jenkins Jr., *Two CEOs, Two Trials*, WALL ST. J., July 14, 2004, at A15 (“A better idea would be for Pfizer and fellow drug makers to publish and stick to a single price at which each drug will be sold to customers in the developed countries. Price discrimination may be socially beneficial; [i]t may allow more people to benefit from a new drug than would be possible if each had to pay an equal share of research costs. Politically, however, price discrimination has become an albatross around the industry’s neck, because other developed nations use price controls to force R&D costs back onto American consumers.”).

51. For example, in many African countries pharmaceutical companies charge less for certain drugs in than they do in the United States. See, e.g., John S. James, *Merck, Bristol-Myers Squibb Announce Major Price Reductions in Poorest Countries*, AIDS TREATMENT NEWS, Feb. 26, 2001, <http://www.aids.org/atn/a-361-03.html>.

52. Examples in the United States include Medicare, Medicaid, Veterans Affairs, federal employees, private health plans, and individuals.

53. See, e.g., *infra* note 199 and accompanying text (discussing Australia’s scheme).

54. For a classic account of the interplay between arbitrage and differential pricing, see PHILIPS, *supra* note 43, at 14-16. A recent study from the London School of Economics does not find any evidence of the predicted price convergence in pharmaceutical parallel trading markets in Europe. PANOS KANAVOS ET AL., *THE ECONOMIC IMPACT OF PHARMACEUTICAL PARALLEL TRADE IN EUROPEAN UNION MEMBER STATES: A STAKEHOLDER ANALYSIS* 15-16 (London Sch. of Econ. & Political Sci., Special Research Paper, 2004), <http://www.lse.ac.uk/collections/LSEHealthAndSocialCare/documents/otherpaperseries>.

ultimate user, arbitrage occurs when buyers in a lower-priced market re-sell the product to consumers in a higher-priced market. Pharmaceuticals sold for five dollars in India may be identical to products sold for one hundred dollars in the United States, creating the opportunity for arbitrage. When arbitrage involves IP and crosses an international border, it is called parallel trade.⁵⁵ Absent other constraints, neo-classical economic theory predicts that arbitrage will erode price-differentiated markets, moving all sales towards an equilibrium price. As a result, arbitrage redirects consumer surplus away from the producer and into the hands of the consumer,⁵⁶ improving access through lower cost. Arbitrage is in fact a normal function of a competitive capitalistic economy, a key component of the invisible hand. Arbitrage loses favor when it threatens innovation by hindering appropriation by pharmaceutical companies. As will be seen later, the empirical reality of pharmaceutical arbitrage departs from the neo-classical model in significant ways.⁵⁷ This Article recognizes that pharmaceutical arbitrage may be either helpful or dysfunctional to consumer welfare.

C. Laws Affecting Pharmaceutical Arbitrage

Successful pharmaceutical price discrimination requires market segmentation and must minimize arbitrage by customers and intermediaries. Several tools may be employed, including contract, product differentiation supported by trademarks, and regulatory structures.⁵⁸ Each affects the degree of appropriation in pharmaceutical markets, and thus, the balance between access and innovation.

htm.

55. Parallel trade, "also called grey-market trade, is the act of taking goods placed into circulation in one market, where they are protected by a trademark, patent or copyright, and shipping them to a second market without the authorization of the local owner of the intellectual property right." Keith E. Maskus & Mattias Ganslandt, *Parallel Trade in Pharmaceutical Products: Implications for Procuring Medicines for Poor Countries*, in *ECONOMICS OF ESSENTIAL MEDICINES*, *supra* note 3, at 57. The practice is not necessarily illegal, depending upon the country's laws concerning exhaustion of IP rights. *See supra* Subsection I.A.1.

56. PHILIPS, *supra* note 43, at 18.

57. *See infra* Subsection II.A.1; *see also supra* note 54 (citing study finding no empirical evidence of price convergence in EU pharmaceutical arbitrage).

58. *See* Jonathan M. Barnett, *Private Protection of Patentable Goods*, 25 *CARDOZO L. REV.* 1251 (2004).

1. Contract

Private ordering may support differential pricing: The contract between a buyer and seller may expressly or implicitly forbid arbitrage.⁵⁹ If the customer breaches the agreement, the seller can pursue contractual remedies. The effectiveness of contractual remedies will in many cases depend upon whether the seller has privity with every arbitrageur and on the monitoring costs required to ensure compliance. In pharmaceutical markets, manufacturers are likely to lack privity with the multiple layers of pharmaceutical distributors and retailers, and contracts of adhesion in the style of shrink-wrap licensing are impractical since pharmaceutical goods are sold rather than licensed. Contractual restrictions on subsequent trade may run afoul of competition law. The European Court of Justice, for example, is generally skeptical of contractual provisions preventing intra-European arbitrage.⁶⁰ Any relaxation of these competition law principles, or a novel expansion of licensing-style restrictions on subsequent transfer, would decrease the potential for arbitrage and expand the appropriation powers of pharmaceutical innovators.

2. Product Differentiation

Successful arbitrage requires that the lower priced product be the same as, or easily substituted for, the more expensive product. When the product is fungible and easily transferable, consumers can cross the price discriminating market segments by choosing the lowest price.⁶¹ However, producers rarely concede strict fungibility; marketing efforts are deployed to influence consumers and reduce their willingness to make substitutions, thus supporting differential pricing.⁶² This process generally occurs

59. Airlines, for example, forbid the transfer of tickets. Some firms refuse to sell equipment, but only lease it with sub-leasing forbidden. The famous example of leased Xerox equipment is described in PHILIPS, *supra* note 43, at 151-53. A more recent example is the software industry's widespread use of non-transferable licenses. These are most often clickwrap or shrink-wrap licenses—contracts of adhesion. See J.H. Reichman & Jonathan A. Franklin, *Privately Legislated Intellectual Property Rights: Reconciling Freedom of Contract with Public Good Uses of Information*, 147 U. PA. L. REV. 875 (1999). Firms may also contractually prohibit parallel trade of their products.

60. Case C-306/96, *Javico Int'l v. Yves Saint Laurent Parfums SA*, 1998 E.C.R. I-1983, [1998] 5 C.M.L.R. 172 (1998).

61. BERNDT, *supra* note 44, at 8-10; PHILIPS, *supra* note 43, at 27.

62. Aspirin might be considered a fungible commodity. The active ingredient is well known and unprotected by patents. Yet, the aspirin market is filled with differentiated

between similar products from competing companies, but parallel traders force companies to confront movements of differentially priced products between geographic markets. Trademarks and laws constraining parallel trade support product differentiation. Granting patents for modest variations in dosage and formulations also supports product differentiation.

Laws regulating pharmaceutical marketing also affect the potential for arbitrage. Drug companies target both consumers and physicians⁶³ with their marketing efforts: Overall, U.S. promotional spending on prescription drugs in 2000 totaled \$15.7 billion.⁶⁴ Even after generic entry, these marketing efforts are remarkably effective in retaining market share.⁶⁵ Finally, transaction costs also influence the ease of substitution. If laws raise arbitrage transaction costs, product differentiation is supported, and arbitrage is hindered.

products. Some aspirins are marketed with brand names as proxies for safety and reliability. Others are compounded with other ingredients such as caffeine or buffering agents. Aspirin may be purchased in particular sizes, shapes, and delivery methods, such as pills, capsules, or gel caps. Despite this product differentiation, at some level all aspirins are subject to substitution. If the preferred brand or form of aspirin is unavailable, or priced too high, some consumers will substitute another form of aspirin, or may even substitute with another class of analgesic, such as ibuprofen or acetaminophen.

63. Companies spend billions of dollars to employ product representatives, who meet with doctors in various venues. In 2000, the industry employed 83,000 drug representatives at a cost of \$4 billion. NAT'L INST. FOR HEALTH CARE MGMT. RESEARCH & EDUC. FOUND., PRESCRIPTION DRUGS AND MASS MEDIA ADVERTISING 5 (2001) [hereinafter NIHCM]. Free samples valued at \$7.9 billion were given to doctors in 2000, and \$1.9 billion was spent on educational conferences for doctors. *Id.* These efforts encourage particular prescribing habits and shift demand between drugs through substitution. *Id.* at 7; *see also* SCHNEIDER, *supra* note 4, at 26-36 (fraud cases); Compliance Program Guidance for Pharmaceutical Manufacturers, 68 Fed. Reg. 3,731, 23,735-38 (May 5, 2003). The industry has also taken steps to suppress negative research. *See* Angell, *supra* note 6, at 62; NAT'L INSTS. OF HEALTH, REPORT OF THE NATIONAL INSTITUTES OF HEALTH BLUE RIBBON PANEL ON CONFLICT OF INTEREST POLICIES 1-5 (2004), http://www.nih.gov/about/ethics_COI_panelreport.pdf [hereinafter NIH, CONFLICT OF INTEREST].

64. NIHCM, *supra* note 63, at 4. Approximately one-third related to one-on-one meetings with doctors, visits to hospitals, or conferences, and only a portion of that could be considered educational. The largest marketing expense is for free drug samples (\$7.9 billion in 2000). *Id.* at 4. In 2000, U.S. unit sales of the fifty most heavily advertised drugs rose at six times the rate of other drugs. *Id.* at 7.

65. CONG. BUDGET OFFICE, HOW INCREASED COMPETITION FROM GENERIC DRUGS HAS AFFECTED PRICES AND RETURNS IN THE PHARMACEUTICAL INDUSTRY, at xii-xiii (1998) [hereinafter CBO, INCREASED COMPETITION].

3. Government Regulation of Pharmaceuticals

Pharmaceutical regulation influences substitution, transaction costs, and arbitrage. Two major legal categories are particularly relevant to pharmaceutical arbitrage: IP laws and national drug regulatory agencies (NDRAs).

i. Intellectual Property (IP) Laws

IP laws facilitate pharmaceutical differential pricing by creating legally enforceable rights, which in turn support the appropriation of rents. Pharmaceutical patents prevent substitution by identical compounds during the patent period. Trademarks support brand identification and differentiation of products to consumers, hindering consumer confusion or unintended substitution.⁶⁶

In many countries, the first sale of a patented product exhausts the public law rights of the patent holder for that item.⁶⁷ This exhaustion rule is a necessary condition⁶⁸ to legal domestic arbitrage, as it permits domestic resale by the purchaser without the permission of the patent holder.⁶⁹ Exhaustion may be applied on a domestic or an international basis. The domestic exhaustion rule renders parallel imports illegal while the international exhaustion rule removes patent law barriers to international

66. TIMOTHY H. HIEBERT, *PARALLEL IMPORTATION IN U.S. TRADEMARK LAW* 151-57 (1994) (discussing the consumer confusion theory underlying the exclusion of parallel imports under trademark law); WARWICK A. ROTHNIE, *PARALLEL IMPORTS* 101-05 (1993) (discussing the role of distinct domestic goodwill to successfully exclude parallel goods under trademark law).

67. ROTHNIE, *supra* note 66, at 125-42 (Anglo-Commonwealth patent law); *id.* at 143-50 (U.S. patent law).

68. It is necessary, but not sufficient: Significant price differentials and relatively low transaction costs are also required. The power of other factors is demonstrated by the persistence of pharmaceutical pricing differentials within the EU, despite a strong internal exhaustion rule and EU firms specializing in pharmaceutical arbitrage. ROTHNIE, *supra* note 66, at 477, 494-97. *See generally* DG TRADE, *supra* note 42, § 3.

69. Domestic parallel trade in pharmaceuticals is legal within the EU and the United States. *See, e.g.*, Case 187/80, *Merck v. Stephar*, 1981 E.C.R. 2063, [1981] 3 C.M.L.R. 463 (holding that parallel drug trade is legal in the EU); DONALD S. CHISUM, *CHISUM ON PATENTS* § 16.03[2] (2003) (explaining the U.S. domestic exhaustion rule). *But see* Case T-41/96, *Bayer A.G. v. E.C. Commission*, [2001] 4 C.M.L.R. 4 (holding that unilateral acts by pharmaceutical company to choke off supply of drugs to parallel exporters is not actionable under EU law); *Glaxo Group Ltd. v. Dowelhurst Ltd.* [2004] EWCA Civ. 129 (requiring repackaging in some cases to protect the trademark).

parallel trade.⁷⁰ United States law only recently rejected the international patent exhaustion rule,⁷¹ and the extent of the rejection may not yet be clear,⁷² although the recent Free Trade Agreement with Australia commits

70. DG TRADE, *supra* note 42, § 3.1 (“A country providing for international exhaustion effectively makes parallel imports legal, while a country (or regional group) that provides for national (or regional) exhaustion enables rightholders to act against such imports.”). TRIPS does not commit to a position on exhaustion, specifically reserving the issue to domestic law. TRIPS, *supra* note 1, art. 6. Some commentators writing on the economics of essential medicines mention in passing that U.S. patent law rejects the international exhaustion rule. *See, e.g.*, JOHN H. BARTON, DIFFERENTIATED PRICING OF PATENTED PRODUCTS (WHO, Comm’n on Macroeconomics & Health, Working Paper No. 2, 2001); JEAN O. LANJOUW, INTELLECTUAL PROPERTY AND THE AVAILABILITY OF PHARMACEUTICALS IN POOR COUNTRIES 19 n.29 (Ctr. for Global Dev., Working Paper No. 5, 2002), *reprinted in* 3 INNOVATION POLICY AND THE ECONOMY (2002) [hereinafter LANJOUW, INTELLECTUAL PROPERTY].

71. One distinguished commentator states, without discussion, that the 1994 amendments reject international exhaustion for U.S. patents, which might imply that the Uruguay Round required this result. CHISUM, *supra* note 69, § 16.05[3]. The amendment was included as part of the Uruguay Round Agreements Act by which the United States joined the WTO. Uruguay Round Agreements Act, Pub. L. No. 103-465, 108 Stat. 4809 (1994). Section 533 of the Uruguay Round Agreements Act amended 35 U.S.C. § 271(a) to expand the definition of infringement to include importation into the United States of a patented product. The legislative history of this provision is obscure. The House Reports on the Uruguay Round Agreements Act do not include an analysis of Section 533, and the only mention in the summary description is: “amends the definition of infringing activity to include offers for sale and importation of a patented good.” H.R. REP. NO. 103-826(I), at 8 (1994). The unofficial summary by the Congressional Research Service merely states: “(Sec. 533) Deems offering to sell or import a patented invention into the United States to be patent infringement.” Cong. Research Serv., Bill Summary & Status, H.R. 5110 (Pub. L. No. 103-465), 103d Cong. (Sept. 27, 1994).

72. Four points are important. First, prior to the 1994 amendments, U.S. patent law was leaning in favor of the international exhaustion rule, a trend which resulted in the 1995 U.S. Supreme Court case, *K Mart Corp. v. Cartier, Inc.*, 486 U.S. 281 (1988). *See also* CHISUM, *supra* note 69, § 16.05[3]; ROTHNIE, *supra* note 66, at 183. Second, it is not clear at all that Congress intended to overturn the international exhaustion exception by the enactment of § 533. One may declare importation an act of infringement and yet retain the narrower exception for authorized sales abroad being imported legally under international exhaustion. *But see* CHISUM, *supra* note 69, § 16.05[3]. Third, the provision, enacted as part of the Uruguay Round Agreements, was not required, as WTO Members retain domestic flexibility to choose any exhaustion rule. Finally, the heuristic of optimality, *see infra* Section I.D, suggests that any provision which strengthens drug patent rights will enhance beneficial innovation only if rents are sub-optimal. This issue was not demonstrated to Congress in the legislative history to the 1994 amendment.

both parties to the domestic exhaustion rule.⁷³

If the United States does follow the domestic exhaustion rule for pharmaceutical patents, drugs sold in the United States, exported to Canada, and then re-imported back into the United States arguably qualify for domestic exhaustion.⁷⁴ However, the Prescription Drug Marketing Act of 1987 blocks reimportation by anyone other than the manufacturer, forbidding this form of arbitrage.⁷⁵

ii. National Drug Regulatory Agencies

The TRIPS Agreement generally leaves the drug approval process to individual countries.⁷⁶ The global diversity of regulatory actors creates the possibility that each country will have a unique drug regulatory environment, with different approaches to issues such as generic substitution, drug approval, reimbursement, parallel trade, advertising, and pharmaceutical arbitrage. In addition, each country's market may differ due to other significant factors such as economic development and demand elasticity. The net result is that law assists in the creation of unique market characteristics in each country, which may result in differentiated prices either by facilitating voluntary differentiation and impeding conditions necessary for arbitrage, or by taking actions, such as price controls, that essentially demand a differentiated scheme.

To begin, a country's regulatory conditions may uniquely affect the potential for product differentiation—an impediment to arbitrage—by allowing or disallowing certain marketing efforts or dictating transactions costs. For example, in 1997, the United States's national drug regulatory

73. In 2004, the U.S.-Australia Free Trade Agreement committed both parties to the domestic exhaustion rule for patents. Free Trade Agreement, May 18, 2004, U.S.-Austl. § 17.9.4, http://www.ustr.gov/Trade_Agreements/Bilateral/Australia_FTA/Final_Text/Section_Index.html.

74. See Rebecca S. Eisenberg, *The Shifting Functional Balance of Patents and Drug Regulation*, 19 HEALTH AFF. 119, 129-32 (2001). Re-imported patented drugs are produced in the United States under proper authority, legally exported to a second country (such as Canada) and then re-imported by a third party, arguably exhausting U.S. patent rights over the pills themselves. There is no evidence that the 1994 modifications to 35 U.S.C. § 271(a) were intended to waive the domestic exhaustion rule on re-imported goods. See *supra* notes 71-72. As discussed *infra* Section II.B, Canadian pharmaceutical arbitrage has recently exploded despite this restriction.

75. Prescription Drug Marketing Act of 1987, 21 U.S.C. §§ 331(t), 381(d) (2000).

76. TRIPS Agreement, *supra* note 1, art. 1, § 1.

agency, the Food and Drug Administration (FDA),⁷⁷ modified its regulations to permit direct to consumer (DTC) advertising for pharmaceutical drugs.⁷⁸ Virtually no other countries permit the practice.⁷⁹ The creation of the DTC rule by the FDA modifies information costs related to substitution.⁸⁰

Other government regulations also influence pharmaceutical marketing. For example, federal law prohibits the sale of a drug sample⁸¹ or the domestic resale of deeply-discounted drugs sold to certain hospitals,⁸² hindering arbitrage of these products and thus supporting their provision at differential prices. The U.S. Department of Health and Human Services applies Medicare fraud and abuse laws to the practices of drug representatives, forbidding remuneration to encourage particular prescribing practices within federal programs.⁸³

Regulatory postures can alter manufacturing costs of potential competitors. The current *de facto* global standard for quality pharmaceutical manufacturing is the Standard of Good Manufacturing

77. 21 U.S.C. § 355 (2000).

78. The regulations are now found at 21 C.F.R. § 202.1 (2004).

79. In recent years, only the United States and New Zealand permit DTC broadcast ads. NIHCM, *supra* note 63, at 16. In 2002, Canada permitted restricted DTC advertising and is affected by spillover from American media. CAN. INST. FOR HEALTH INFO., DRUG EXPENDITURE IN CANADA: 1985-2002, at 41 (2003). In the 2004 U.S.-Australia Free Trade Agreement, web-based DTC advertising is permitted. Free Trade Agreement, *supra* note 73, Annex 2-C, ¶ 5.

80. PHILIPS, *supra* note 43, ch. 12. DTC campaigns build consumer demand, encouraging the patient to ask for a prescription by name. Advertising shifts the demand curve for prescription drugs to the right. NIHCM, *supra* note 63, at 2 (noting that DTC advertising increases consumer sales of patented pharmaceuticals); CBO, INCREASED COMPETITION, *supra* note 65, at 20. Spending for DTC advertising grew at an annual rate of 44.9% from 1995 to 2000 and is now growing at an annual rate of 9.4%. Stephen Heffler et al., *Health Spending Projections for 2002-2012*, HEALTH AFF., Feb. 7, 2003 (Web Exclusive), at <http://content.healthaffairs.org/cgi/content/full/hlthaff.w3.54v1/DC1>. Product shift, increased unit prices, and increased volumes each account for about a third of the growth in prescription drug spending. C. Daniel Mullins et al., *The Impact of Pipeline Drugs on Drug Spending Growth*, 20 HEALTH AFF. 210, 213 (2001). In 2000, the most heavily advertised drugs accounted for 47.8% of the \$20.8 billion increase in U.S. retail spending on prescription drugs. NIHCM, *supra* note 63, at 2.

81. 21 U.S.C. §§ 331(t), 353(d) (2000).

82. *Id.* § 353(c)(3).

83. SCHNEIDER, *supra* note 4, at 26-36 (reviewing False Claim Act litigation against drug companies, particularly involving marketing related fraud); Compliance Program Guidance for Pharmaceutical Manufacturers, 68 Fed. Reg. 23,731, 23,733-39 (May 5, 2003).

Practice (GMP). PhRMA companies are now cooperating with the United States, the EU, and Japan to develop a higher global standard, known as the International Conference on Harmonization (ICH).⁸⁴ Imposition of ICH would discourage substitution of drugs manufactured by less-expensive non-OECD⁸⁵ pharmaceutical companies. This effort could be viewed as rent-seeking behavior through technical standards. Likewise, donor agencies often face substitution choices during the procurement process, which may be subject to regulation or political intervention.⁸⁶

Furthermore, international arbitration may simply be proscribed by NDRA's. Under the Food, Drug and Cosmetics Act, drugs cannot be imported unless approved by the FDA,⁸⁷ creating a non-tariff barrier to international trade. Some drugs are produced in the United States and exported to countries with price controls such as Canada.⁸⁸ Since the drugs are produced in the United States, they arguably comply with FDA rules, and could be re-imported back into the United States by arbitrageurs. However, the U.S. Prescription Drug Marketing Act of 1987 prohibits the reimportation of a prescription drug by anyone other than the manufacturer.⁸⁹ The law was ostensibly intended to address safety concerns for the U.S. pharmaceutical supply chain,⁹⁰ but its effect is to prevent

84. GRAHAM DUKES, UN MILLENNIUM PROJECT, INTERIM REPORT OF TASK FORCE 5 WORKING GROUP ON ACCESS TO ESSENTIAL MEDICINES 32 (2004).

85. The OECD is the Organisation for Economic Co-operation and Development; its membership consists almost exclusively of high income countries. See OECD, Information by Country, at http://www.oecd.org/infobycountry/0,2646,en_2649_201185_1_1_1_2_1,00.html; World Bank, Data & Statistics: Country Group, at <http://www.worldbank.org/data/countryclass/classgroups.htm> (last visited Nov. 22, 2004). Non-OECD pharmaceutical companies are essentially those based outside of Japan, North America, and Europe, such as India's Cipla and Ranbaxy. These companies are typically best known for their production of generic products. See, e.g., Donald G. McNeil, Jr., *Selling Cheap 'Generic' Drugs, India's Copycats Irk Industry*, N.Y. Times, Dec. 1, 2000, at A1.

86. The United States's unilateral effort on AIDS (PEPFAR) has chosen to ignore the WHO prequalification process, as well as all recipient country drug regulatory agencies, and now imposes a supplementary FDA approval process for AIDS drug procurement. Sarah Lueck, *White House Aims To Answer Critics of Its AIDS Fight*, WALL ST. J., Apr. 29, 2004, at A9; Sarah Lueck, *White House Gets Pressure on AIDS Plan*, WALL ST. J., Mar. 25, 2004, at A4. This decision, ostensibly made on quality grounds, also supports the product line of PhRMA companies by imposing additional regulatory requirements on their generic competitors located in India, South Africa, Thailand, and Brazil.

87. 21 U.S.C. §§ 360(i), 381(a) (2000).

88. See *infra* Subsection II.A.2.

89. 21 U.S.C. §§ 331(t), 381(d) (2000).

90. H.R. REP. NO. 100-76, at 7 (1987).

international pharmaceutical arbitrage or parallel trade.⁹¹

Finally, PhRMA companies generally do not enjoy unconstrained monopoly power to set prices on patented drugs. In high income countries, regulatory systems, as well as payor monopsony, will likely yield countervailing pricing power. In some countries, the government sets pharmaceutical prices by regulatory process, including reference pricing⁹² and rate setting.⁹³ In others, price regulation occurs when the government enters the market as a purchaser and acts with monopsony power.⁹⁴ Private payors (health plans or their agents such as pharmacy benefit managers) may either mimic the government prices, or utilize their own market power to negotiate prices.⁹⁵ Moreover, most third-party payors have pharmaceutical substitution agendas of their own which are subject to government regulation. Many health plans now require prescriptions to be filled with generic equivalents whenever medically appropriate. In the United States, state and federal laws generally support these efforts.⁹⁶

91. The government also has the power to seize counterfeit or improperly diverted drugs. For an interesting story on the diversion of Serostim within the United States, see Christopher Windham, *Cracking Down on Illicit Use of AIDS Drugs*, WALL ST. J., Jan. 19, 2004, at B1.

92. PATRICIA DANZON & JOHATHAN D. KETCHAM, REFERENCE PRICING OF PHARMACEUTICALS FOR MEDICARE: EVIDENCE FROM GERMANY, THE NETHERLANDS AND NEW ZEALAND (Nat'l Bureau of Econ. Research, Working Paper No. W10007, 2003) (discussing reference price systems in Germany, The Netherlands, and New Zealand).

93. House of Commons Select Comm. on Health, Minutes of Evidence (Jan. 23, 2002), <http://www.publications.parliament.uk/pa/cm200102/cmselect/cmhealth/515/2012321.htm> (examination of Dr. John Patterson, President-elect, Association of the British Pharmaceutical Industry) ("Prices almost never go up on medicines in this country [England], as you saw from the report to Parliament in December. In brief, the PPRS is a scheme which caps profits and profitability in our industry at a level equivalent to the average return on capital of the FT 100."). The United States effectively sets rates for government purchase of services from physicians and hospitals, but generally not for pharmaceuticals.

94. In the United States, the recently-enacted Medicare Act disabled federal monopsony power in the purchase of outpatient prescription drugs under Medicare. Medicare Prescription Drug Improvement and Modernization Act of 2003, Pub. L. No. 108-173, § 301 (to be codified at 42 U.S.C. § 1395).

95. CBO, INCREASED COMPETITION, *supra* note 65, at xi.

96. See, e.g., W. VA. CODE § 30-5-12 (2004) (allowing pharmacists to substitute generic medicines for brand name medicines without approval from the prescriber); *id.* § 23-4-3 (requiring generic substitution within the Workers' Compensation program). But see DANZON & KETCHAM, *supra* note 92, at 7 (noting that Germany restricts generic substitution).

iii. *The Hatch-Waxman Act*

Traditionally, IP law regulates the economic incentives of innovation while NDRA regulations and related laws control drug efficacy and safety. However, the patent system is not the only source of exclusive, or monopoly, rights. Under the Hatch-Waxman Act⁹⁷ and other legislation such as the Orphan Drug Act,⁹⁸ the FDA may grant additional exclusive marketing periods under an array of circumstances—for example, rewarding first-mover generic drugs,⁹⁹ certain drugs for uncommon conditions (so-called orphan drugs),¹⁰⁰ or compliance with social goals such as testing drugs for efficacy and safety on children.¹⁰¹ Indeed, when examining the incentives for pharmaceutical innovation, it is not the length of the patent period that matters most but the duration of this exclusive marketing period.¹⁰² PhRMA companies are maximizing their opportunities under these provisions.¹⁰³

97. Drug Price Competition and Patent Term Restoration Act of 1984, Pub. L. No. 98-417, 98 Stat. 1585 (codified as amended in scattered sections of 15, 21, 28 and 35 U.S.C.) [hereinafter Hatch-Waxman Act]. Under the Hatch-Waxman Act, the FDA also influences the patent process, since Hatch-Waxman extends the patent for half of the period that a drug is undergoing clinical trials, plus the full amount of time spent in the FDA approval process. 35 U.S.C. §§ 155, 155A, 156 (2000).

98. 21 U.S.C. §§ 360aa-360ee (2000).

99. *Id.* § 355(j).

100. *Id.* §§ 360aa-360ee.

101. *Id.* § 355a.

102. The term “exclusive marketing period” means the actual period during which a pharmaceutical company sells an FDA-approved drug in the United States without direct competition. The legal sources of this period include patent law, non-patent “exclusive marketing” rights granted by the FDA under Hatch-Waxman, the use of litigation and agreements to forestall competitive entry, and the evergreening of patents through filings for new uses and formulations.

103. For example, the number of putative orphan drugs qualifying for tax credits and extended exclusive marketing periods has soared as companies have narrowly defined markets to remain under the 200,000-person threshold. Steven R. Salbu, *AIDS and Drug Policy: In Search of a Policy*, 71 WASH. U. L.Q. 691, 692, 704-06 (1993) (noting that the FDA designated AZT as an orphan drug in 1987 and half of AIDS drugs as of August 1991 were designated as orphans); John J. Flynn, *The Orphan Drug Act: An Unconstitutional Exercise of the Patent Power*, 1992 UTAH L. REV. 389 (noting that the FDA designated early AIDS drugs such as AZT, and other best-selling drugs such as EPO and Taxol as orphan drugs). The tax expenditure on the Orphan Drug Act is now \$200 million per year, not including the cost of the grant of market exclusivity. JOINT COMM. ON TAXATION, ESTIMATES OF FEDERAL TAX EXPENDITURES FOR FYS 2004-2008 (Joint Committee Print 2003). Public Citizen notes the

The 1984 Hatch-Waxman Act was the first major piece of legislation to link patent law and FDA regulations in this way. The Act regulates patent expiry and generic entry following patent expiration, directly addressing the balance between innovation and access.¹⁰⁴ The United States is now exporting portions of the Hatch-Waxman Act to other countries through bilateral free trade agreements.¹⁰⁵

After a patent or exclusive marketing period expires,¹⁰⁶ competition by generic drugs is not automatic. Generic drugs must receive FDA approval as well, albeit under an abbreviated process. The generic entry process can take some time, particularly if existing data on safety and efficacy cannot be used or if the manufacturing processes are complex. PhRMA companies have resorted to strategic litigation and collusive agreements to lengthen effective exclusive marketing periods.¹⁰⁷ These abuses prompted amendments to Hatch-Waxman in 2003.¹⁰⁸ PhRMA companies are already responding with new tactics to keep generic drugs off the market by denying the generic companies an adequate financial return for the expensive generic approval process.¹⁰⁹

inefficiency of the incentive mechanism: Pediatric tests cost only \$3.9 million per drug on average, but the six-month patent extension can result in huge financial rewards exceeding \$1 billion. PUBLIC CITIZEN'S CONG. WATCH, PUBLIC CITIZEN, THE OTHER DRUG WAR II: DRUG COMPANIES USE AN ARMY OF 623 LOBBYISTS TO KEEP PROFITS UP 4 (2002). The FDA estimates the total cost of the pediatric testing initiative from 2001 to 2021 to be \$14 billion, approximately equal to the proposed five year AIDS program. FDA, THE PEDIATRIC EXCLUSIVITY PROVISION: STATUS REPORT TO CONGRESS (2001).

104. See, e.g., Elizabeth Stotland Weiswasser & Scott D. Danzis, *The Hatch-Waxman Act: History, Structure, and Legacy*, 71 ANTITRUST 585 (2003). Philipson and Mechoulan describe this balance in the language of economics: "Appropriate policy must *simultaneously* solve the externality problem ex-post and the R&D problem ex-ante." PHILIPSON & MECOULAN, *supra* note 30, at 12 (emphasis in original).

105. See, e.g., Free Trade Agreement, *supra* note 73, §17.9.6.

106. By the late 1990s, the U.S. pharmaceutical exclusive marketing period was approximately fourteen years. CBO, INCREASED COMPETITION, *supra* note 65, at 45-48. If someone undertakes to update this figure, care should be taken to account for all of the factors affecting effective exclusive rights.

107. FTC, GENERIC DRUG ENTRY PRIOR TO PATENT EXPIRATION 13-23 (2002).

108. Pub. L. No. 108-173, 117 Stat. 2066 (2003), tit. IX (codified at 21 U.S.C. § 355(j) and uncodified). The Congressional Research Service prepared a summary of the Act which provides some guidance on Congress's intent in amending Hatch-Waxman. See Cong. Research Serv., Prescription Drug and Medicare Improvement Act of 2003, Bill Summary and Status, S.1, 108th Cong. (June 13, 2003).

109. Leila Abboud, *Drug Makers Use New Tactic To Ding Generics*, WALL ST. J., Jan. 27, 2004, at B1.

*D. The Heuristic of Globally Optimal Pharmaceutical Rents*¹¹⁰

1. Nonrival Access to Pharmaceutical Knowledge

The goal of IP laws should be to maximize nonrival access to pharmaceutical knowledge, with just enough legal support for the appropriation of rents to protect socially optimal R&D. Since pharmaceutical knowledge is nonrival, it should be disseminated in the widest possible fashion at the lowest possible cost for the greatest possible benefit to global public health. This Article describes this condition as “nonrival access.”

The pharmaceutical industry has borrowed language from the world of physical property to attack nonrival access. They call nonrival access “theft” or “piracy.” At best, nonrival users are characterized as “free riders.”¹¹¹ These terms are inappropriate since nonrival use of pharmaceutical knowledge does not cause anything to be lost,¹¹² so long as the socially optimal level of appropriation for R&D is still achieved. In a world of excessive rents, we should call it theft (or, in some cases, genocide) to deny nonrival access to low income populations.

For the pharmaceutical industry, the globally optimal level of appropriation through rents¹¹³ must be sufficient to fund the socially

110. Once again, the broader term *rents* is used here in lieu of *patent rents* in order to encompass the various mechanisms beyond patent law which facilitate appropriation, as described *supra* Section I.C. The use of the term *patent rents* is meant to signify only the narrow meaning of patent-based appropriation.

111. DRAHOS WITH BRAITHWAITE, *supra* note 2, at 19-29 (piracy); LEMLEY, *supra* note 20, at 3-16 (property and free riding); STERK, *supra* note 20, at 24-25 (analogies to tangible property).

112. Cf. Selling Life-Saving Drugs to Poorer Countries: At What Cost?, Research at Penn (Nov. 6, 2002), at <http://www.upenn.edu/researchatpenn/article.php?504&hlt>.

113. The economic analysis of socially optimal patents has been undertaken by Nordhaus and Scherer. NORDHAUS, INVENTION, GROWTH & WELFARE, *supra* note 36, at 70-92; Nordhaus, *The Optimum Life of a Patent*, *supra* note 36, at 428; Scherer, *Optimal Patent Life*, *supra* note 36, at 422. Scherer argues that shortening patent life will reduce R&D only for the most marginal inventions, particularly in industries with nonpatent barriers to entry and post innovation pricing discipline. Scherer, *Optimal Patent Life*, *supra* note 36, at 426. The pharmaceutical research industry contains both conditions. Nordhaus concluded that a fixed patent life was not optimal, but given that requirement, the length of the life should err to a longer rather than a shorter period. Nordhaus, *The Optimum Life of a Patent*, *supra* note 36, at 428. Philipson and Mechoulan cover the same territory when they argue that “[a]ppropriate policy must *simultaneously* solve the externality problem ex-post and the R&D

optimal level of R&D. Optimization must balance concerns of cost, quality, and access, looking for the greatest net gain to global public welfare. Excessive rents harm human health without advancing socially optimal R&D. Society must decide when the best level of appropriation has been reached.¹¹⁴

Maximizing R&D at all costs should not be our objective. Resources devoted to R&D are not available for other uses.¹¹⁵ Uwe Reinhardt puts it this way: "Year after year, the last dollar spent on drug research and development (R&D) should yield society as much benefit as it would have yielded if it had been spent to produce other goods or services."¹¹⁶

We should also avoid the assumption that all R&D targets are equally valuable. Some innovations are more valuable than others. Companies allocate research funds in response to price signals from commercial pharmaceutical markets. As a result, Americans now have a third drug for erectile dysfunction,¹¹⁷ and funds for neglected disease innovation are literally going to the dogs,¹¹⁸ but malaria and AIDS vaccines are not

problem ex-ante." PHILIPSON & MECHOULAN, *supra* note 30, at 12, 12-15. Recently, Christopher Yoo undertook a nuanced review of copyright law which covers some of the same terrain as my approach, but with assumptions of copyright market entry and substitutability which do not apply to pharmaceutical patents. See Christopher S. Yoo, *Copyright and Product Differentiation*, 79 N.Y.U. L. REV. 212 (2004).

114. Philipson and Mechoulan make a similar point in the language of economics: "Under external effects in consumption, rewards to innovation should not be guided by potential *consumer* surplus, as under private goods, but the entire *social* surplus that includes benefits to non-consumers as well as consumers . . ." PHILIPSON & MECHOULAN, *supra* note 30, at 2.

115. Currently the United States spends more than fifteen percent of its GDP on health care. Robert Pear, *Health Spending Rises to Record 15 % of Economy*, N.Y. TIMES, Jan. 9, 2004, at A16. Perhaps we can agree that increasing pharmaceutical R&D to twenty percent or fifty percent of GDP would be excessive.

116. Uwe E. Reinhardt, *An Information Infrastructure for the Pharmaceutical Market*, 23 HEALTH AFF. 107 (2004).

117. Viagra (sildenafil) was approved by the FDA in 1998. *First Oral Therapy for Erectile Dysfunction*, 28 FDA MEDICAL BULL. 1 (1998), <http://www.fda.gov/medbull/summer98/erectile.html>. Levitra (vardenafil) was approved in August 2003. FDA, Talk Paper, FDA Approves New Drug for Treatment of Erectile Dysfunction in Men (Aug. 19, 2003), <http://www.fda.gov/bbs/topics/ANSWERS/2003/ANS01249.html>. Cialis (tadalafil) was approved in November 2003. FDA, Talk Paper, FDA Approves Third Drug To Treat Erectile Dysfunction (Nov. 21, 2003), <http://www.fda.gov/bbs/topics/ANSWERS/2003/ANS01265.html>.

118. In 1999, the FDA approved two drugs to treat canine Cognitive Dysfunction Syndrome, also known as separation anxiety in dogs. FDA, Talk Paper, FDA Approves First

available.¹¹⁹

You get the sense that ships are passing in the night on this issue. James Love estimates the static global deadweight loss on pharmaceutical patents at over \$400 billion per year,¹²⁰ and Larry Lessig implores us not to allow IP law to be perverted while a holocaust devastates millions in the developing world.¹²¹ Meanwhile Joseph DiMasi and Henry Grabowski suggest that the “dynamic benefits created by patents on pharmaceuticals can, and almost surely do, swamp in significance their short-run inefficiencies.”¹²² Yet, in a major study, the Congressional Budget Office conceded that no one knows whether current levels of pharmaceutical R&D are optimal.¹²³ This is the pressing question.

2. Globally Sub-Optimal Pharmaceutical Rents

Globally sub-optimal pharmaceutical rents would stifle the production of pharmaceutical knowledge, creating a generational equity issue. The present group of patients might benefit from sub-optimal pharmaceutical rents because such rents result in innovative treatments that are cheaper and thus more accessible, but future quality would be compromised. Pharmaceutical companies would invest less when creating inappropriable knowledge. This is the nightmare scenario portrayed by PhRMA companies when they argue that constraints on their ability to appropriate rents will squelch innovation.

Behavioral Drugs for Dogs (Jan. 5, 1999), <http://www.fda.gov/bbs/topics/answers/ans00934.html>. Perhaps soon a drug will be developed for erectile dysfunction in dogs.

119. For an introduction to donor efforts led by the Bill & Melinda Gates Foundation to stimulate development of a malaria vaccine, see Malaria Vaccine Initiative, at <http://www.maliavaccine.org>.

120. James Love, *supra* note 33, at 2.

121. Lawrence Lessig, *The International Information Society*, 24 LOY. L.A. ENT. L. REV. 33, 36-37 (2004).

122. JOSEPH A. DIMASI & HENRY G. GRABOWSKI, PATENTS AND R&D INCENTIVES: COMMENTS ON THE HUBBARD AND LOVE TRADE FRAMEWORK FOR FINANCING PHARMACEUTICAL R&D 2 (2004) (citation omitted), <http://www.who.int/intellectualproperty/news/en/Submission3.pdf>. DiMasi and Grabowski cite the 2003 study by Philipson and Mehoulan, but that study assumes sub-optimality rather than proves it. See PHILIPSON & MECHOULAN, *supra* note 30.

123. The 1998 study by the Congressional Budget Office states: “No one knows whether that amount of investment in R&D is over or under the optimal level.” CBO, INCREASED COMPETITION, *supra* note 65, at 48.

3. Globally Supra-Optimal Pharmaceutical Rents

Globally supra-optimal pharmaceutical rents are rarely recognized as a potential problem by PhRMA companies. By definition, supra-optimal pharmaceutical rents are not necessary to fund R&D; they simply harm consumers by raising prices and restricting access without providing the counterbalancing benefits of future innovation.

i. Are Supra-Optimal Pharmaceutical Rents Possible?

One economist reviewer of an earlier draft of this Article suggested that pharmaceutical rents cannot be supra-optimal because PhRMA companies have not fully appropriated all consumer surplus associated with their products.¹²⁴ This is another way of saying that PhRMA companies have not yet achieved first-degree differential pricing (or Ramsey Optimal Pricing). While Ramsey Optimal Pricing would maximize the sales and profits of PhRMA companies, it does not respond to the distributional balance between innovation and access. Nor does it address the quality of research undertaken with the surplus so completely extracted from consumers. In a market beset with profound agency problems and information disparities, it is absurd to assume that consumers will purchase pharmaceuticals at the cost-effective price. Given what we know about pharmaceutical markets, it is at least equally likely that PhRMA companies will stimulate demand which varies from optimal therapeutic need, while neglecting less lucrative markets.

This critique also fails to account for important negative externalities. PhRMA companies have failed to get the right pills to the right people at the right price. If another regime would result in greater global welfare (improved therapeutic outcomes) without damage to dynamic innovation incentives, then it should be preferred even if it reduces pharmaceutical rents slightly.

Consider the vast global gains in welfare which would result if nonrival access could be accomplished without diminishing the incentive to innovate. The opportunity cost of failing to do so is staggering. The net gains to global social welfare would be very significant, even if nonrival access came at the cost of a modest slice of innovation. It is in this sense that pharmaceutical rents may be supra-optimal.

124. I thank Aidan Hollis for this comment. My criticisms are not directed at him.

ii. Are Pharmaceutical Rents Supra-Optimal?

Some empirical evidence suggests that PhRMA companies earn well above market rates of return, one possible indicator of supra-optimal pharmaceutical rents.¹²⁵ The industry's long-term profits are four times the rate of the Fortune 500.¹²⁶ Analysis of IRS data from 1990 to 1996 demonstrates that the drug industry's after-tax profits are more than triple the rate for all industries.¹²⁷

Calculating optimal pharmaceutical rents must account for other sources of public funding for R&D, such as government grants, direct government expenditures, foundation donors, and tax incentives. The industry receives substantial tax incentives, resulting in an effective U.S. federal income tax rate of 16.2%, compared with 27.3% generally.¹²⁸

The ways in which PhRMA companies currently opt to expend their cash flows may also indicate supra-optimality. The pharmaceutical industry currently spends more on sales and marketing than on R&D.¹²⁹ Large marketing expenses are not proof that pharmaceutical rents are supra-optimal, but merely indicate that the industry believes the return on investment in marketing is greater than alternative investments such as R&D. If the industry holds a relatively low view of the value of an additional

125. The barriers to this calculation are both empirical and theoretical. On the empirical front, internal company data are not generally available to researchers. Studies by DiMasi, Hansen, and Grabowski rely on self-reported PhRMA company data rather than a truly objective data set. DiMasi et al., *supra* note 29. IRS data shows extraordinary profits and low taxation but is protected against public disclosure by the Internal Revenue Code. See GARY GUENTHER, CONG. RESEARCH SERV., *FEDERAL TAXATION OF THE DRUG INDUSTRY FROM 1990 TO 1996* (1999). Accurate pricing data is unavailable outside of the companies. CBO, *INCREASED COMPETITION*, *supra* note 65, at 20. On the theoretical front, useful questions are posed by Reinhardt, *supra* note 116; and William S. Comanor, *Political Economy of the Pharmaceutical Industry*, 24 J. ECON. LIT. 1178, 1182-86 (1986).

126. DAVID H. KRELING ET AL., THE KAISER FAMILY FOUND., *PRESCRIPTION DRUG TRENDS: A CHARTBOOK UPDATE* exhibit 32 (2001). The judgment of the equity markets is significant, even under a weak form of the efficient capital markets hypothesis.

127. GUENTHER, *supra* note 125.

128. *Id.*

129. KRELING ET AL., *supra* note 126, exhibit 30 (noting that the top ten major pharmaceutical manufacturers in 2000 spent 34.4% of their revenues on "marketing, general and administrative" expenses and 13.7% on "research and development."). But see Uwe E. Reinhardt, *Perspectives on the Pharmaceutical Industry*, 20 HEALTH AFF. 136 (2001) (suggesting that not all SG&A expenses are truly marketing). With deference to Reinhardt, the differential is large enough to suggest that R&D receives less than marketing, absent more specific and verifiable data.

dollar of R&D investment, then perhaps society would be better served with that additional dollar being used to provide life-saving access to medicines.

Some scholars, including proponents of the anti-commons movement,¹³⁰ suggest that the neo-classical link between patents and innovation is overstated, particularly for industries marked by cumulative innovation¹³¹ such as genetics.¹³² If so, optimal rents may be lower than previously expected.

The most important data required to resolve this question are in the hands of the pharmaceutical industry and are not available in a reliable form to independent researchers.¹³³ This fact alone is a compelling reason to demand transparency. It certainly seems plausible to presume that supra-optimal rents are currently being collected. The burden of coming forward with contrary evidence should be placed on the parties controlling the relevant information: the PhRMA companies.

4. Implications of Global Optimality

Pending the resolution of the empirical issue, the concept of globally optimal pharmaceutical rents is useful as a heuristic tool. The following Subsections outline several implications which follow from applying this tool to pharmaceutical markets.

130. The leading article is Michael A. Heller & Rebecca S. Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 SCIENCE 698 (1998). For a recent study on the importance of maintaining a scientific commons, see J.H. Reichman & Paul F. Uhler, *A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment*, 66 LAW & CONTEMP. PROBS. 315 (2003).

131. Oren Bar-Gill & Gideon Parchomovsky, *The Value of Giving Away Secrets*, 89 VA. L. REV. 1857 (2003). While Bar-Gill and Parchomovsky list “pharmacology” as one such industry, they do not make that case convincingly in the article. If PhRMA companies are eager to publish and forego patents, it is a nascent trend.

132. The work of Tim Hubbard and James Love is particularly interesting in this regard. Hubbard & Love, *supra* note 27.

133. See *supra* note 125. Pharmaceutical pricing and profitability data are notoriously opaque and misleading. SCHNEIDER, *supra* note 4; Gardiner Harris, *Drug Companies Settle 7 Suits for \$1.6 Billion*, N.Y. TIMES, Nov. 6, 2003, at 8 (“Drug companies have paid a total of \$1.6 billion since 2001 to settle seven suits brought by whistle-blowers that accused them of marketing fraud and overbilling Medicare and Medicaid . . .”). Some researchers suggest that increased pricing opacity is necessary to sustain differential pricing for low income countries. DANZON & TOWSE, *supra* note 12, at 16-20. I suggest that transparency will better serve global public health.

i. Nonrival Access to Pharmaceutical Innovation

Patented pharmaceuticals can be delivered at marginal cost of production to low income populations without harming innovation. The majority of AIDS patients in low income countries are quite poor and are not part of the global market for patented drugs. Supplying their needs is a humanitarian response, and pharmaceutical companies do not actually lose viable commercial markets as a result.¹³⁴ These non-market patients could receive unlicensed or royalty-free drugs without impacting the cash flow of PhRMA companies.¹³⁵

Even if global pharmaceutical rents are currently sub-optimal, unlicensed or royalty-free production should still be allowed so long as it does not replace any commercial market, and thus does no financial harm to the patent owner.¹³⁶ Certainly if global pharmaceutical rents are now supra-optimal, PhRMA companies could bear the expenses of monitoring and enforcing differential pricing without harming innovation. Supra-optimality also permits expansion of differential pricing programs to middle income markets, even with some displacement of commercial markets. The magnitude of expense and market loss that could be tolerated would depend on the amount by which pharmaceutical rents are supra-optimal.

a. The Need for a Credible Threat of Compulsory Licensing

OECD members with monopsonistic public sector purchasing of pharmaceuticals can negotiate or impose domestic second-degree differential pricing to meet local needs. For developing countries, which often lack a significant publicly financed pharmaceutical sector,

134. See John H. Barton, *TRIPS and the Global Pharmaceutical Market*, 23 HEALTH AFF. 146, 148 (2004).

135. Frederic M. Scherer recently made a similar point when he argued for allowing free riding by developing countries on pharmaceutical patents. F.M. Scherer, *A Note on Global Welfare in Pharmaceutical Patenting*, 27 WORLD ECON. 1127, 1141 (2004) [hereinafter Scherer, *Global Welfare*].

136. Philipson and Mechoulam criticize this position, but their stance is undermined if global pharmaceutical rents are supra-optimal. PHILIPSON & MECHOULAN, *supra* note 30, at 19-20. Even if one assumes sub-optimality, differential pricing for ARVs does not reduce R&D incentives if cash flows to the innovators are untouched. Philipson and Mechoulam's argument thus collapses to a complaint that differential pricing does not improve upon status quo R&D incentives. If the effect in innovation is positive or neutral, the health gains (positive externalities) from increased access should drive policy.

compulsory licensing, or at least the credible threat thereof, may be required.

At the Fourth WTO Ministerial Conference in Doha, WTO members agreed to the Doha Declaration as an interpretation of TRIPS.¹³⁷ The Doha Declaration allows WTO Members to take measures to “protect public health and, in particular, to promote access to medicines for all.”¹³⁸ Specifically, WTO Members may compel licensure to protect public health, without limitation to AIDS or any particular disease.¹³⁹

Sovereign threats of such compulsory licenses, public pressure from NGOs, and actual competition from generic¹⁴⁰ companies persuaded PhRMA companies and the United States to embrace differential pricing of antiretroviral (ARV) medications for a number of poor countries combating HIV/AIDS. Médecins sans Frontières (MSF) and others consider the threat and use of compulsory licenses to have been essential in convincing companies to establish meaningful differential pricing

137. *Doha Declaration on TRIPS*, *supra* note 32. The legal status of the Doha Declaration is discussed in James Thujo Gathii, *The Legal Status of the Doha Declaration on TRIPS and Public Health Under the Vienna Convention on the Law of Treaties*, 15 HARV. J.L. & TECH. 291 (2002); and in CARLOS M. CORREA, IMPLICATIONS OF THE DOHA DECLARATION ON THE TRIPS AGREEMENT AND PUBLIC HEALTH 5 (WHO, Health Econ. and Drugs, EDM Series No. 12, 2002), <http://www.iprsonline.org/resources/health.htm> [hereinafter CORREA, IMPLICATIONS OF DOHA]. The legal status of the Cancun General Council Decision is a joint commitment by WTO Members to abide by its terms in good faith. Press Release, European Comm’n, EU Strongly Welcomes WTO Deal on Generic Medicines, IP/03/1189 (Sept. 1, 2003) [hereinafter EU, Cancun]. (The EU uses the phrase “Perez Motta text” to describe the Cancun General Council Decision.) Practically speaking, it would be virtually impossible to prevail at DSB on a provision contrary to the Cancun General Council Decision. The legal status of both Doha and Cancun are expected to be clarified in a planned 2004 amendment to TRIPS. General Council, World Trade Org., *Implementation of Paragraph 6 of the Doha Declaration on the TRIPS Agreement and Public Health: Decision of the General Council*, WT/L/540, ¶ 11 (Aug. 30, 2003) [hereinafter *Cancun General Council Decision*]; see also *Doha Declaration on TRIPS*, *supra* note 32, ¶ 7.

138. TRIPS Agreement, *supra* note 1, art. 31(f); *Doha Declaration on TRIPS*, *supra* note 32, ¶ 4.

139. *Doha Declaration on TRIPS*, *supra* note 32, ¶ 5; ‘t Hoen, *supra* note 7, at 40-41.

140. This Article is generally focused on generics of controversial legal status, sometimes referred to as “unlicensed” generics (i.e., a copy of a patented pill made by a manufacturer that has not been authorized by the originator company). This terminology can be confusing in light of the role of compulsory licenses and the questionable need for licensing in some situations. It is simply important to keep in mind that we are not speaking simply of generics for off-patent products, but the more complex market for generics of drugs that may be subject to patents in the United States or elsewhere.

programs.¹⁴¹ PhRMA companies strongly resisted both significant price reductions as well as generic ARV drugs, citing both TRIPS and domestic IP legislation.¹⁴²

Several examples illustrate the effectiveness of the credible threat of generic production. Brazil's threat to issue a compulsory license, coupled with its non-recognition of pharmaceutical patents prior to the adoption of TRIPS, permitted the distribution of free ARVs within Brazil.¹⁴³ In January 2001, the United States requested a WTO panel against Brazil to prevent Brazilian "local manufacture" of ARVs.¹⁴⁴ Under international pressure, the United States withdrew the panel request on June 25, 2001, in the months leading up to the Fourth WTO Ministerial Conference in Doha.¹⁴⁵

141. MÉDECINS SANS FRONTIÈRES, SURMOUNTING CHALLENGES: PROCUREMENT OF ANTIRETROVIRAL MEDICINES IN LOW- AND MIDDLE-INCOME COUNTRIES 7, 9, 42 (2003), <http://www.accessmed-msf.org/documents/procurementreport.pdf> [hereinafter MSF, SURMOUNTING CHALLENGES] (report prepared by MSF at the request of the WHO); Marleen Boelaert et al., Letter to the Editor, 287 JAMA 840, 840 (2002) ("This impressive discount offered by the companies to developing countries was not merely due to public outcry, but mostly as a response to competition by generic drugs.").

142. MSF, SURMOUNTING CHALLENGES, *supra* note 141; Judy Rein, *International Governance Through Trade Agreements: Patent Protection for Essential Medicines*, 21 NW. J. INT'L L. & BUS. 379, 394-404 (2001); Susan K. Sell, *TRIPS and the Access to Medicines Campaign*, 20 WIS. INT'L L.J. 481, 491-96 (2002) [hereinafter Sell, *TRIPS*]; 't Hoen, *supra* note 7, at 30-33.

143. U.K. COMM'N ON INTELLECTUAL PROPERTY RIGHTS, INTEGRATING INTELLECTUAL PROPERTY RIGHTS AND DEVELOPMENT POLICY 43 (2002); JEROME H. REICHMAN WITH CATHERINE HASENZAH, UNCTAD, NON-VOLUNTARY LICENSING OF PATENTED INVENTIONS: HISTORICAL PERSPECTIVE, LEGAL FRAMEWORK UNDER TRIPS, AND AN OVERVIEW OF THE PRACTICE IN CANADA AND THE USA 2 (Project on IPRs and Sustainable Development, Issue Paper No. 5, 2003); Jorge Bermudez, *Expanding Access to Essential Medicines in Brazil: Recent Economic Regulation, Policy-Making and Lessons Learnt*, in *ECONOMICS OF ESSENTIAL MEDICINES*, 178, 193 (2002).

144. Permanent Mission of the United States, *Brazil Measures Affecting Patent Protection, Request for the Establishment of a Panel by the United States*, WT/DS199/3 (Jan. 9, 2001), <http://www.cptech.org/ip/health/c/brazil/Req4EstabPanel.html>. Executive Order 13155 had specifically reserved to the United States the right to seek such a panel. Exec. Order No. 13155, 65 Fed. Reg. 30521, 30522 (May 10, 2000) ("This order does not prohibit the United States Government from invoking the dispute settlement procedures of the World Trade Organization to examine whether any such law or policy is consistent with [TRIPS]."). For an overview of the Brazilian and South African situations by the Congressional Research Service, see JOHN R. THOMAS, CONG. RESEARCH SERV., HIV/AIDS DRUGS, PATENTS AND THE TRIPS AGREEMENT: ISSUES AND OPTIONS 13-17 (2001) [hereinafter THOMAS, CRS REPORT].

145. CORREA, IMPLICATIONS OF DOHA, *supra* note 137, at 2 & n.6; THOMAS, CRS REPORT, *supra* note 144, at 15; 't Hoen, *supra* note 7, at 38-47.

Indeed, even the United States has resorted to this tactic in recent years: During the anthrax scare, threats of compulsory licensing of ciprofloxacin were instrumental in securing a lower price from Bayer,¹⁴⁶ and compulsory licensing remains an important remedy in litigation.¹⁴⁷

Finally, voluntary no-royalty licenses, such as Merck's recent grant to South African-Indian company Thembalami Pharmaceuticals,¹⁴⁸ must be viewed in the context of South Africa's compulsory licensing law. That is, such licenses can be seen as responses to the looming threat of compulsory licensing.¹⁴⁹

146. Jill Carroll & Ron Winslow, *Bayer Agrees To Slash Prices for Cipro Drug*, WALL ST. J., Oct. 25, 2001, at A3 ("The agreement comes after a high-stakes threat by Tommy Thompson, HHS secretary, to break Bayer's patent for Cipro if he didn't get the price he wanted."). The U.S. compulsory license statutes are 7 U.S.C. § 2404 (2000) (patents necessary for the nation's food supply); 17 U.S.C. § 115 (2000) (copyrights to certain musical works); 28 U.S.C. § 1498 (2000) (patents); 35 U.S.C. § 203 (2000) (patents developed through the use of government research funding under the Bayh-Dole Act); and 42 U.S.C. § 2183 (2000) (atomic energy). The U.S. compulsory license statutes do not contain the restrictions required by Article 31 of TRIPS. See TRIPS Agreement, *supra* note 1, art. 31. For an authoritative review of United States and Canadian experience with compulsory licensure, see REICHMAN WITH HASENZAH, *supra* note 143, at 19-22.

147. Makan Delrahim, Deputy Assistant Attorney General, Antitrust Division, U.S. Department of Justice, *Forcing Firms To Share the Sandbox: Compulsory Licensing of Intellectual Property Rights and Antitrust*, Presentation at the British Institute of International and Comparative Law (May 10, 2004).

148. Press Release, Merck & Co., Inc., Merck & Co., Inc. Grants License for HIV/AIDS Drug Efavirenz to South African Company, Thembalami Pharmaceuticals in Effort To Accelerate Access to Life-Saving Treatment (July 14, 2004), <http://www.pressmethod.com/releasestorage/5003645.htm>.

149. South Africa passed a compulsory licensing law in 1997, Medicines and Related Substances Control Amendment Act No. 90 of 1997 (Republic of South Africa). The government was promptly sued by PhRMA companies. The U.S. government suspended bilateral economic assistance to South Africa as punishment for defending the suit. Omnibus Consolidated and Emergency Supplemental Appropriations Act, 1999, Pub. L. No. 105-277, 112 Stat. 2681-153 (1999) (suspending appropriation of all bilateral economic assistance to South Africa, including AIDS/HIV programs, until steps are taken to repeal section 15(c) of South Africa's Medicines and Related Substances Control Amendment Act No. 90 of 1997). Many commentators have written about the case and the U.S. trade pressure exerted upon South Africa. See, e.g., Lissett Ferreira, *Access to Affordable HIV/AIDS Drugs: The Human Rights Obligations of Multinational Pharmaceutical Corporations*, 71 FORDHAM L. REV. 1133, 1155 (2002); Rein, *supra* note 142, at 400-02; 't Hoen, *supra* note 7, at 30-31. Doha paragraph 4 discourages Members from exerting bilateral pressure which hinders the exercise of TRIPS and Doha rights. CORREA, IMPLICATIONS OF DOHA, *supra* note 137, at 12. The U.S. government and PhRMA companies relented under great pressure in April 2001,

Although threats of compulsory licensing may lead to differential pricing, it is worth noting that reliance on voluntary price discrimination to achieve marginal-cost distribution to low income populations has proven very disappointing. Over the past five years there have been many announcements of dramatic price cuts or voluntary programs, yet these announcements have not resulted in much actual treatment in 2004.¹⁵⁰ Each PhRMA company creates idiosyncratic policies specifying which countries qualify for differential pricing on any particular drug. Many of these policies are limited to sub-Saharan Africa or specific low income countries, thereby excluding AIDS crises in Asia, the former Soviet states, Latin America, or most of the Caribbean. Transaction costs are high when essential access discounts are negotiated on a case-by-case basis. Company policies vary by the status of the purchaser (e.g., NGO, IGO, government, private buyer).

Voluntary programs of differential pricing also fail to achieve differential pricing at the marginal cost of production, which is absolutely necessary for nonrival access. Voluntary negotiations kept ARV prices unnecessarily high for years and delayed effective treatment for millions of dying people. The Médecins sans Frontières pricing guide confirms that most voluntary differential pricing programs continue to price significantly above generic levels,¹⁵¹ a practice generally followed in the United States

shortly before Doha. Editorial, *South Africa's Moral Victory*, 357 THE LANCET 1303 (2001); THOMAS, CRS REPORT, *supra* note 144, at 16.

150. See, for example, the correspondence concerning access to Pfizer's Diflucan Donation Program, announced with great fanfare several years ago, but apparently still unavailable on the ground in the Dominican Republic. E-mail from Eugene Schiff, Agua Buena, to Joseph Saba, Axios (Sept. 20, 2004) (on file with author). A five company group negotiated with five UN agencies for a year in 2000 and 2001 without tangible success. Each company ended up negotiating access deals with each individual country. Paul Blustein & Barton Gellman, *HIV Drug Prices Cut for Poorer Countries; Other Firms May Follow Merck's Lead*, WASH. POST, Mar. 8, 2001, at A1.

151. MÉDECINS SANS FRONTIÈRES, UNTANGLING THE WEB OF PRICE REDUCTIONS: A PRICING GUIDE FOR THE PURCHASE OF ARVs FOR DEVELOPING COUNTRIES (6th ed. 2004), <http://www.accessmed-msf.org> [hereinafter MSF, UNTANGLING THE WEB]. Merck makes Stocrin (efavirenz, EFV) 600 mg available in Columbia for US\$767 per year. *Id.* at 9. The lowest cost generic provider is Hetero of India at US\$347 per year. *Id.* at 12, 22. Merck matches the generic price only in Low Human Development Index (HDI) countries and Medium HDI countries with adult HIV prevalence of one percent or greater. *Id.* at 12. The distinction is lost on very poor persons living with HIV/AIDS in a Medium HDI country with prevalence under one percent, such as Columbia.

after generic market entry.¹⁵²

Voluntary differential pricing should be extended to target populations in a larger group of countries. If pharmaceutical rents are supra-optimal, loss of some elite markets will not harm innovation. Even if pharmaceutical rents are sub-optimal, additional countries can receive differential pricing if they undertake serious measures to segment and protect the local elite market.¹⁵³ As the AIDS epidemic expands in Eastern Europe and Central Asia, access must be expanded to regions beyond sub-Saharan Africa.

Given the apparent limitations of the efficacy of voluntary pricing, it is important that compulsory licensing be more than a threat—that it be a viable way for countries to introduce generic competition. However, lack of manufacturing capacity in the lowest income countries limits the practicability of domestic production of generic pharmaceuticals. The TRIPS Agreement seemingly restricts compulsory licenses predominantly to domestic use, effectively preventing exports.¹⁵⁴ Since many countries do not have domestic pharmaceutical production capacity, the no-export rule prevents many countries from delivering low-cost ARVs to HIV/AIDS patients.¹⁵⁵ For example, compulsory licenses are arguably not useful to Malawi absent the opportunity to import from other countries, such as Brazil, India, or South Africa. The ensuing debate was energetic, leading up to the Cancun WTO meeting in 2003.

Immediately prior to the Cancun meeting, on August 30, 2003, the United States conceded the point. Under the Cancun General Council Decision, the WTO now permits exports of compulsory licensed drugs to the poorest countries—an important development if compulsory licensing is to be a meaningful option for countries without manufacturing capacity.¹⁵⁶ The Cancun General Council Decision established a WTO notification process for cross-border compulsory licenses: The TRIPS Council must be notified, but WTO approval is not required.¹⁵⁷ In May 2004, Canada amended the Canadian Patent Law to permit compulsory

152. CBO, INCREASED COMPETITION, *supra* note 65, at xiii.

153. *See infra* Part II.

154. TRIPS Agreement, *supra* note 1, art. 31(f).

155. *See Doha Declaration on TRIPS*, *supra* note 32, ¶ 6.

156. *Cancun General Council Decision*, *supra* note 137; EU, Cancun, *supra* note 137. While the Cancun General Council Decision has the potential to positively impact access, it has not yet had an effect on drug availability. *See infra* note 293 and accompanying text.

157. *Cancun General Council Decision*, *supra* note 137, ¶ 2; *see also* EU, Cancun, *supra* note 137 (noting that WTO approval is not required).

licenses for certain drug exports to needy nations.¹⁵⁸ As of September 2004, no WTO Member has notified the TRIPS Council.¹⁵⁹

Phil Thorpe's study on TRIPS implementation recently found that most developing countries have not taken advantage of the flexibilities and exceptions permitted under TRIPS.¹⁶⁰ He does not explore the reasons behind this failure, but two are likely. First, many countries may lack the impartial technical assistance needed to implement these provisions, including restrictions on "new use" patents, Bolar provisions, and international exhaustion rules. When the World Intellectual Property Association has provided assistance, developing countries have found WIPO's agenda to be IP maximalist rather than aimed at taking full advantage of TRIPS flexibilities.¹⁶¹ Second, the TRIPS-plus¹⁶² offensive of the U.S. Trade Representative (USTR) and the "Special 301" reports from that same office are frequently used to bluster countries into modifying

158. The Jean Chretien Pledge to Africa Act, House of Commons, 3d Sess., 37th Parliament, 52-53 Eliz. II, 2004 (Bill C-9) (received Royal Assent on 14 May 2004) [hereinafter Canadian Bill C-9]. The law created a positive list of drugs eligible for compulsory licensure, a procedural hurdle not required by the WTO. *Id.* Sched. 1. France and Norway have recently followed suit. Law. N. 2004-800, Aug. 6, 2004, J.O, Aug. 7, 2004, p. 18 (Fr); Regulations Amending the Patent Regulations (In Accordance With the Decision of the WTO General Council of 30 August 2003, Paragraphs 1(b) and 2(a)), Royal Decree of 14 May, 2004 (Nor.), <http://www.cptech.org/ip/health> [hereinafter Norwegian Compulsory License Regulation]. Norway does not have a significant pharmaceutical sector, so the impact of the regulation is modest. *Id.* ¶ 7 (official explanation of the regulation). Canada is more likely to actually export, but the Canadian law is more restrictive than the Norwegian. See CANADIAN HIV/AIDS LEGAL NETWORK, GLOBAL ACCESS TO TREATMENT: CANADA'S BILL C-9 AND THE COMPULSORY LICENSING OF PHARMACEUTICALS FOR EXPORT TO COUNTRIES IN NEED (2004), <http://www.aidslaw.ca/Maincontent/issues/cts/patent-amend/billC-9flyer300604.pdf>. The Norwegian regulations are far less restrictive.

159. The WTO has established a web page to announce notifications under Doha and Cancun, http://www.wto.org/english/tratop_e/trips_e/public_health_e.htm. None are posted as of September 28, 2004.

160. THORPE, *supra* note 32, at 1.

161. Public comments by the official delegations from several non-OECD countries at the Trans Atlantic Consumer Dialogue Future of WIPO Workshop, Geneva (Sept. 13, 2004) (the author was present); see also Geneva Declaration on the Future of the World Intellectual Property Organization, <http://www.cptech.org/ip/wipo/futureofwipo/declaration.pdf> (last visited Sept. 21, 2004) ("The WIPO technical assistance programs must be fundamentally reformed.").

162. "TRIPS Plus" refers to provisions which exceed the floors established under the TRIPS Agreement.

domestic law to the liking of U.S. owners of IP. WTO Members should have a realistic opportunity to implement the flexibilities bargained for in TRIPS, including compulsory licensure, unhindered by unilateral U.S. interests.

b. Compulsory Licensing Need Not Harm Optimal Innovation

Assuming that production for compulsory licensure is limited to non-commercial markets, production by a third party does not add any marginal cost to the innovator and, thus, will not impede innovation.¹⁶³ If global pharmaceutical rents are supra-optimal, then compulsory licenses without royalties can be utilized without loss of innovation incentives. The burden of proof of sub-optimality should be on the innovator companies seeking a royalty, and the royalty rate in conditions of sub-optimality should balance innovation and access goals. In all cases, such nonrival use by low income populations should be viewed as an opportunity rather than a problem.¹⁶⁴

A free rider problem may emerge if compulsory licensure decisions are evaluated solely at the domestic level. Each country may rationally choose to shirk its share of R&D costs, the same free rider problem afflicting innovation generally. Some form of global coordination may be required to address the negative externality.¹⁶⁵ Second-degree price discrimination such as price controls or other domestic rules affecting the ability of companies to appropriate rents also raise global coordination issues, which are now being negotiated in U.S. bilateral free trade agreements.¹⁶⁶

163. This result holds without regard for whether rents are currently supra- or sub-optimal. Critiques of compulsory licenses by Merges and others are not applicable here because the goal is not the initiation of efficient bargaining around a rule, but the provision of essential medicines at marginal cost without harming innovation. See Robert P. Merges, *Contracting into Liability Rules: Intellectual Property Rights and Collective Rights Organizations*, 84 CAL. L. REV. 1293 (1996) (arguing that compulsory licenses in digital media are less efficient than private contractual efforts).

164. Scherer, *Global Welfare*, *supra* note 135, at 1141.

165. This is particularly true amongst the OECD, where free riding has the greatest potential to affect global rents. See *infra* Section II.B.

166. BUDDHIMA LOKUGE & THOMAS FAUNCE, TRADE DISPUTES AND THE PHARMACEUTICAL BENEFITS SCHEME: CONSTRUCTIVE AMBIGUITIES, NON-VIOLATION NULLIFICATION DISPUTES AND THE AUSTRALIA US FREE TRADE AGREEMENT 8-9 (Austl. Nat'l Univ., Informal Working Paper, Sept. 2004) (on file with author).

ii. Dysfunctional Pharmaceutical Arbitrage

The second implication of global optimality concerns dysfunctional pharmaceutical arbitrage. The form of pharmaceutical arbitrage which is most likely to reduce rents is diversion from charitable non-commercial markets into high income markets.¹⁶⁷ If global pharmaceutical rents are sub-optimal (or made sub-optimal thereby), this arbitrage may be labeled dysfunctional. The EU recognizes that its attempts to support differential pricing for essential medicines depend in part upon blocking arbitrage into high income markets.¹⁶⁸

It is important to note the limited scope of the case against dysfunctional pharmaceutical arbitrage. It does not apply to generic drugs because protecting the generic company's profits will not create incentives for innovative R&D, and thus arbitrage restrictions on generic drugs are not supportable on innovation grounds.¹⁶⁹

Restrictions are also inappropriate between and to low income markets, so long as commercial markets are not replaced. Arbitrage restrictions could be lifted on sales to and within low and medium income countries. Outside of high income markets, the international exhaustion rule should always apply, as there is no proven innovation-based warrant for denying nonrival access.

Some level of arbitrage to recent immigrants to high income countries might be tolerable. Very little money is at stake for PhRMA companies and the likely high income country consumers of smuggled African drugs might well be at the margins of the country's health care system. Recent immigrants may not be full market participants either, despite their physical location in a high income country. The well-publicized confiscation of thirty-six thousand packages of African AIDS medications in the Netherlands in October 2002 might fit this profile.¹⁷⁰ Even if the

167. Parallel trade from poor countries to rich countries is incompatible with differential pricing of essential medicines. See DANZON & TOWSE, *supra* note 12 (noting that parallel trade defeats the objectives of differential pricing); David A. Malueg & Marius Schwartz, *Parallel Imports, Demand Dispersion, and International Price Discrimination*, 37 J. INT'L ECON. 167, 193 (1994).

168. DG TRADE, *supra* note 42, at 2.

169. Restrictions might be appropriate on other grounds, such as safety. If a generic drug has not been approved in a market, importing it would not be arbitrage. For unpatented or generic products, no innovation-based case for banning parallel trade can be offered.

170. DUKES, *supra* note 84, at 50 n.1. For surprising details on this case, see *infra* notes 313-323 and accompanying text.

patients are market participants, receiving familiar medications from home, in their native language, might well be the best medical practice. In the United States, the uniform use of English labels in a multicultural society is not a culturally competent practice for recent immigrants lacking good English skills.

Arbitrage controls may be unnecessary between and within high income markets if pharmaceutical rents are supra-optimal. Put another way, parallel trade in patented pharmaceuticals within high income markets may be permitted.¹⁷¹ If rents are sub-optimal, the domestic exhaustion rule should apply in high income markets, forbidding parallel imports into such countries and raising pharmaceutical rents. Otherwise, the international exhaustion rule should apply to sales between high income markets on free trade principles since consumers will benefit while innovation incentives remain intact.

iii. Domestic Pharmaceutical Arbitrage

The current TRIPS approach is tied to state sovereignty, affecting legal regimes along national political boundaries. TRIPS aggregates customers into country-level markets, reflecting both transaction costs and the political realities of sovereignty. This state-centric system is not surprising, given that only states are WTO Members, but the process suffers from both over-inclusion and under-inclusion.

Over-inclusion occurs when an entire country is granted an exception, extension, or flexibility under TRIPS, even though some people within these low or middle income countries can afford to pay high income market prices for drugs. Even in the poorest countries, an elite cadre of individuals control enough wealth to afford these drugs. In middle income countries such as India, Brazil, Chile, Mexico, South Africa, China, and Argentina, these markets are significant and growing.¹⁷² The elites in low

171. Pharmaceutical arbitrage within high income markets is the subject of Section II.B on Canadian-U.S. pharmaceutical arbitrage.

172. In its 2001 submission to the United States Trade Representative, PhRMA claimed that \$260 million was lost annually due to unlicensed generic drug products in Argentina. Sell, TRIPS, *supra* note 142, at 496 n.55 (citing PHARM. RESEARCH MFRS. OF AM., NATIONAL TRADE ESTIMATE REPORT ON FOREIGN TRADE BARRIERS (2001)). In 2003, the pharmaceutical industry's estimate ballooned to \$600 million and was included in the 2003 National Trade Estimate Report without any apparent verification from outside of the industry. U.S. OFFICE OF TRADE REPRESENTATIVE, 2003 NATIONAL TRADE ESTIMATE REPORT ON FOREIGN TRADE BARRIERS, ARGENTINA 6 (2003), http://www.ustr.gov/assets/Document_Library/Reports_Publications/2003/2003_NTE_Report/asset_upload_file997_6178.pdf. The estimate was

and middle income countries are actually part of the high income market and should be expected to participate in this market on normal commercial terms.¹⁷³

Theory suggests that providing low-cost AIDS drugs to impoverished South Africans might make it more difficult to charge full price to wealthy or middle class South Africans, but apparently PhRMA companies effectively segment these markets,¹⁷⁴ much as they do in the United

dropped from the 2004 National Trade Estimate Report. U.S. OFFICE OF TRADE REPRESENTATIVE, 2004 NATIONAL TRADE ESTIMATE REPORT ON FOREIGN TRADE BARRIERS, ARGENTINA (2004), http://www.ustr.gov/assets/Document_Library/Reports_Publications/2004/2004_National_Trade_Estimate/2004_NTE_Report/asset_upload_file568_4735.pdf.

173. Pharmaceutical companies may currently prefer to keep the small full-priced elite market in developing countries rather than risk arbitrage. FREDERICK M. SCHERER & JAYASHREE WATAL, WHO COMM'N ON MACROECONOMICS & HEALTH, POST-TRIPS OPTIONS FOR ACCESS TO PATENTED MEDICINES FOR DEVELOPING COUNTRIES (2001) [hereinafter SCHERER & WATAL, POST-TRIPS OPTIONS]; Oxfam, *Fatal Side Effects: Medicine Patents Under the Microscope*, in ECONOMICS OF ESSENTIAL MEDICINES, *supra* note 3, 81, 93 (suggesting drug companies profit from elite households in Argentina, Brazil, India, and China); W. Duncan Reekie, *The Development Trilemma and the South African Response*, in THE ECONOMICS OF ESSENTIAL MEDICINES, *supra* note 3, at 167-68 (showing that the top twenty percent of South Africans enjoy a per capita GNP of \$27,699, comparable to OECD levels and are therefore a significant market for drug companies); World Health Organization-World Trade Organization, *Differential Pricing and the Financing of Essential Drugs*, in ECONOMICS OF ESSENTIAL MEDICINES, *supra* note 3, at 213 (recognizing elite drug markets in developing nations); Patricia Danzon & Michael Furukawa, Prices and Availability of Pharmaceuticals: Evidence from Nine Countries exhibit 8 (undated presentation), at <http://hc.wharton.upenn.edu/danzon/index.htm> (showing that prices normalized by national income in Chile and Mexico are at 528% and 529% of the U.S. prices, which I interpret to mean that drug purchasers in Chile and Mexico must have personal incomes far in excess of the national average). In their public filings with the U.S. Securities and Exchange Commission, PhRMA companies acknowledge the growing middle class markets in the developing world. Merck & Co, Inc., SEC Form 10-k, at 14 (filed Mar. 10, 2004) [hereinafter Merck, SEC Form 10-k]. PhRMA companies have recognized the potential of these markets for some time. *Foreign Trade Practices (Part 2): Hearing Before the Subcomm. on Oversight and Investigations, and the House Comm. on Energy and Commerce*, 99th Cong. 196 (1985) (statement of Gerald Mossinghoff, PhRMA President).

174. In South Africa, the NGO and public sector price for a triple therapy regime (ZDV/3TC+NVP) was US\$400 per person year while the private sector price in South Africa was US\$2007. MSF, SURMOUNTING CHALLENGES, *supra* note 141, at 37. A recent WHO survey found significant variations in prices of essential medications within most countries surveyed. Jeanne Madden, *Basic Results That the WHO/HAI Survey Offers Country-Level Investigators*, 33 ESSENTIAL DRUG MONITOR 15 (2003). Significant domestic price variations indicate that various legal and market-based segmentation approaches were apparently

States.¹⁷⁵ The persistence of domestic differential pricing, even in the face of extensive donor programs, is a testament to the effectiveness of market segmentation by PhRMA companies and the apparent weakness of actual pharmaceutical arbitrage pressure. Possible mechanisms are brand campaigns with trademarks, differential pricing by payor, and domestic legal restrictions on arbitrage.¹⁷⁶

Under-inclusion occurs when a middle income country does not qualify for exceptions or flexibilities, or is discouraged from taking advantage of them, despite the needs of some desperately poor citizens therein.¹⁷⁷ The state-centric system lays responsibility for low income patients on the middle and high income countries in which they reside. Here we see a weakness of any system of defining market segments by state political borders rather than actual health needs or ability to afford medicines. It also illustrates the arbitrary categories of development and the difficulties a country might face when it 'graduates' to a higher category.

Accommodations (such as nonrival access to low income populations) may be offered to middle income countries without damaging innovation, so long as domestic price discrimination legal structures are successfully maintained.¹⁷⁸ Again, if global pharmaceutical rents are supra-optimal, PhRMA companies could bear the loss of some elite markets without harming innovation.

A simple estimate in the case of HIV drugs may be useful: PhRMA would likely not suffer significant lost profit if all sales of HIV products in

functioning.

175. See W. VA. PHARM. COST MGMT. COUNCIL, REFERENCE PRICING SUBCOMMITTEE 2-3, app. A-1, A-2 (2004) (demonstrating significant price discrimination within West Virginia between prescription drug prices under Medicaid, private payors, the Public Health Service's 340b program, and the Federal Supply Schedule, as well as Canadian and Australian prices) [hereinafter WEST VIRGINIA REPORT], <http://www.wvc.state.wv.us/got/pharmacycouncil/default.cfm>.

176. Within the U.S. market, internal diversion is illegal in many cases. See Heather Won Tesoriero & Gary Fields, *FBI, FDA Investigates Big Drug Wholesaler*, WALL ST. J., Sept. 19, 2003, at B1 (reporting alleged diversion from discounted hospital markets to higher-priced secondary markets).

177. See, e.g., Letter to Jong-Wook Lee, Director General, World Health Organization, and Peter Piot, Executive Director, Joint United Nations Programme on HIV/AIDS (Apr. 5, 2004), <http://www.aidsinfonyc.org/tag/activism/UNltrOnPriceReductions.html> (discussing the plight of lower middle income countries such as Egypt, Ukraine, Costa Rica, El Salvador, and Panama where ARVs are priced at unaffordable levels).

178. For a discussion of these legal structures, see *supra* Section I.C.

every low and middle income country dropped to zero. GlaxoSmithKline, the largest participant in the market for HIV drugs, reports sales in three geographic regions: the United States, Europe, and “International.” This latter category includes high income countries such as Japan, Canada and Australia, as well as low and middle income countries in Latin America, Asia, Africa and the Middle East. Even so, total international HIV drug sales in 2003 were only £155 million,¹⁷⁹ in a year in which gross profit was £17.2 billion and selling, general, and administrative (SG&A) expenses were £7.5 billion. Actual profits from ARV sales in both low and middle income markets are likely to be negligible to GSK’s global profits and R&D, particularly if elite markets in these countries remain commercial.

iv. Optimizing Subsidies

Another form of optimization creates subsidies to achieve particular goals. Push subsidies include tax credits for R&D, general research grants such as those distributed by the United States’s National Institutes of Health (NIH), and the orphan drug tax credit. Pull subsidies directly address the issue of the appropriation of rents; such mechanisms include the patent system, exclusive marketing periods for orphan and pediatric drugs, and donor purchase commitments for development of a specific pharmaceutical, such as an AIDS or malaria vaccine¹⁸⁰ or antidotes to bioterrorism.¹⁸¹

The heuristic suggests three implications. First, for drugs or conditions with sub-optimal pharmaceutical rents, government intervention should increase pharmaceutical rents toward optimal levels. For example, subsidies are essential for neglected diseases, where the target population cannot afford any commercial price for therapy. Second, subsidies can be limited to drugs with sub-optimal pharmaceutical rents without harming innovation. Scarce subsidies should not be directed to drugs with strong commercial potential, but should be reserved for neglected diseases.

179. GLAXOSMITHKLINE PLC, 2003 ANNUAL REPORT, Form 20-F, at 61-63.

180. Michael Kremer, *Pharmaceuticals and the Developing World*, 16 J. ECON. PERSP. 67, 82-85 (2002). For a recent example, see Press Release, Inst. for OneWorld Health, Institute for OneWorld Health Receives Gates Foundation Grant to Fund Development of Malaria Vaccine (July 13, 2004), <http://www.oneworldhealth.org/media/details.php?prID=76>.

181. The Congressional Research Service indicates that “guaranteeing a market through contract authority” is an aspect of President Bush’s Project BioShield to develop bioterror countermeasures. FRANK GOTTRON, PROJECT BIOSHIELD 1 (Cong. Research Serv. Report for Congress, RS21507, 2003). The proposed size of the pull subsidy for bioterror countermeasures is \$5.593 billion through FY 2013. *Id.* at 3.

Finally, for patented drugs with supra-optimal pharmaceutical rents, the government may intervene to achieve other goals, such as improved nonrival access, without undermining R&D innovation.

Applying these implications to recent policy proposals is instructive. Frederic M. Scherer and Jayashree Watal have proposed expanding U.S. tax incentives for donating pharmaceuticals to poor countries,¹⁸² but this additional push subsidy is warranted only if pharmaceutical rents are sub-optimal. Likewise, the U.S. Congress on September 23, 2004 authorized \$7.6 billion to extend the expiring R&D tax credit for another eighteen months without targeting specific disease conditions.¹⁸³ Proposals for indiscriminate tax credits are unsupported absent evidence of sub-optimality of global pharmaceutical rents.

v. National Drug Regulation and WHO Prequalification

National regimes for testing the safety and efficacy of patented drugs are inefficient, duplicating scientific work and wasting resources unnecessarily. Each New Chemical Entity (NCE) requires clearance by the FDA in the United States and parallel regulatory authorities in every country where the drug will be sold. Prior to the establishment of the EMEA,¹⁸⁴ some estimates put the cost of duplicative NDRA processes within the EU at £500 million per year.¹⁸⁵ NDRA rules also delay the launch of innovative drugs in many countries.¹⁸⁶ A "reference" approval process would reduce duplicative costs and speed market entry of pharmaceuticals.¹⁸⁷

182. SCHERER & WATAL, POST-TRIPS OPTIONS, *supra* note 173.

183. Working Families Tax Relief Act of 2004, Pub. L. No. 108-311, §301(1), 118 Stat. 1166 (2004) (conference report approved by House and Senate, Sept. 23, 2004); Rob Wells & Maya Jackson Randall, *Tax-Cut Bill Aids Firms; Research Credit Extended*, WALL ST. J., Sept. 27, 2004, at A5.

184. Council Regulation 2309/93, O.J. (L 214), as amended by Commission Regulation 649/98 O.J. (L 88) 7.

185. ROTHNIE, *supra* note 66, at 493-94 (citing various sources).

186. See PATRICIA M. DANZON ET AL., IMPACT OF PRICE REGULATION ON THE LAUNCH DELAY OF NEW DRUGS: EVIDENCE FROM TWENTY-FIVE MAJOR MARKETS IN THE 1990S (Nat'l Bureau of Econ. Research, Working Paper No. 9874, 2003). This study collects data on launch delay and concludes that in addition to difficulties with the drug approval process, many companies delay applications to enter some smaller markets due to fears of pharmaceutical arbitrage. If global rents are supra-optimal, this industry practice is reprehensible, as it voluntarily withholds important drugs from patients.

187. Many NDRAs practice a form of reference approval when they require, as a

A reference approval system requires at least four provisions. First, safety and efficacy testing would be referenced against approval in certain benchmark countries. For example, if a compound was approved as safe and efficacious by either the U.S. FDA or the EU's EMEA, then it could automatically be deemed to meet standards in the target country. Second, WHO prequalification (or a similar process) would be deemed to satisfy other domestic NDRA requirements, such as bioequivalence for generic entry and good manufacturing practices. Third, IP rights and drug marketing approvals should also be de-linked. IP rights would still be enforceable under domestic law and TRIPS, but NDRA approval should proceed apace. Finally, in categories of strong local collective preference (such as RU-486), the NDRA may retain a veto.

The United States opposes the first three of these elements, without an innovation warrant. Expansion of the WHO prequalification process is a clear example. WHO Prequalification is clearly useful in many regions, with many different companies producing generic ARVs under unknown conditions.¹⁸⁸ In the 2004 World Health Assembly, the United States pushed to remove the word "strengthening" from the WHO HIV/AIDS Resolution concerning prequalification.¹⁸⁹ The word was retained in the final document,¹⁹⁰ but the United States continues to marginalize the prequalification process in PEPFAR.¹⁹¹ The United States also implicitly opposed reference approvals in various free trade agreements, on the ground that the rights of data exclusivity must be protected.¹⁹² The recent

condition of application for marketing approval, prior marketing approval in either the United States., the EU, or Japan. My suggestion is that NDRAs could consider extending the practice for all of the biological aspects of the marketing approval process, retaining only the right to veto based on a collective preference, as well as approval of the labeling.

188. AM. FOUND. FOR AIDS RESEARCH, TREAT ASIA SPECIAL REPORT: EXPANDED AVAILABILITY OF HIV/AIDS DRUGS IN ASIA CREATES URGENT NEED FOR TRAINED DOCTORS 4 (2004), <http://www.amfAR.org/treatment/news/TADoc7.pdf> [hereinafter AmfAR].

189. *Compare* World Health Org., A57/A/Conf.Paper No. 3 Rev. 1 (May 20, 2004), with Rev.2 (May 21, 2004).

190. *Scaling up Treatment and Care Within a Coordinated and Comprehensive Response to HIV/AIDS*, World Health Assembly, 57th Ass., Agenda Item 12.1, at 3, WHO Doc. WHA57.14 (May 22, 2004) [hereinafter World Health Assembly, *Scaling up Treatment*].

191. Marilyn Chase, *Generic AIDS Pill Gets Acceptance*, WALL ST. J., July 2, 2004, at B3.

192. MÉDECINS SANS FRONTIÈRES, MSF BRIEFING NOTE, ACCESS TO MEDICINES AT RISK ACROSS THE GLOBE: WHAT TO WATCH OUT FOR IN FREE TRADE AGREEMENTS WITH THE UNITED STATES 4-6 (2004), <http://www.accessmed-msf.org/documents/ftabriefingenglish.pdf> [hereinafter MSF, FREE TRADE AGREEMENTS]; DAVID VIVAS-EUGUI, QUAKER U.N. OFFICE, REGIONAL AND BILATERAL AGREEMENTS AND A TRIPS-PLUS WORLD: THE FREE TRADE AREA OF THE AMERICAS

Free Trade Agreement (FTA) with Australia requires linkage between drug approval and patent status for the first time, exporting a portion of Hatch-Waxman to Australia.¹⁹³

Resources are also wasted in the generic entry process. NDRA's should not require generic applicants to repeat any clinical studies without a clear benefit to public health.¹⁹⁴ Generic companies also expend resources to reverse-engineer patented drugs. Reverse-engineering in this case is a wasteful effort and needlessly delays launch in low income countries by several years.¹⁹⁵ The United States's TRIPS-plus proposals to extend data exclusivity to five or ten years¹⁹⁶ will further increase costs and delay generic entry. If pharmaceutical rents are already supra-optimal, all of this is a social loss. Taking unnecessary costs out of the NDRA system makes R&D more efficient, lowers the threshold for cost-effective innovation, and delivers innovative drugs to patients more quickly.

vi. Price Controls

This Article is agnostic on the question of the desirability of pharmaceutical price controls generally. The purpose of this Section is to describe what form price controls should (or should not) take if policy makers choose to adopt them.

The heuristic suggests five conclusions about pharmaceutical price controls. It confirms three relatively uncontroversial points: (1) price

(FTAA) 16-18 (2003) (on file with author) [hereinafter VIVAS-EUGUI, QUAKER UN OFFICE].

193. M. Kevin Outterson, *Free Trade in Pharmaceuticals*, 181 MED. J. AUSTL. 260-61 (2004); Ken J. Harvey et al., *Will the Australia-United States Free Trade Agreement Undermine the Pharmaceutical Benefits Scheme?*, 181 MED. J. AUSTL. 256-59 (2004).

194. PhRMA companies withhold much of this data as trade secrets or seek "data exclusivity" to block generic entry, but when a patent is set to expire there is no innovation warrant to delay generic entry, unless all generic entry is premature.

195. Watal reports a lag of about two years for generic pharmaceuticals reverse-engineered in India. Jayashree Watal, *Pharmaceutical Patents, Prices and Welfare Losses: Policy Options for India Under the WTO TRIPS Agreement*, 23 WORLD ECON. 733-52 (2000).

196. MSF, FREE TRADE AGREEMENTS, *supra* note 192, at 4-6; OXFAM INT'L, UNDERMINING ACCESS TO MEDICINES: COMPARISONS OF FIVE US FTA'S 13-15 (Oxfam Briefing Note, 2004); VIVAS-EUGUI, QUAKER UN OFFICE, *supra* note 192, at 16-18. The United States is currently pressuring Israel to accept five to ten years of data exclusivity, a threat targeting Teva Pharmaceuticals, one of the world's largest producers of generic medicines, based in Israel. *Teva Opposes 10-Year Data Exclusivity Provision for Israel*, 21 GENERIC LINE, May 5, 2004, 2004 WL 65711471; Hadas Manor, *US to Israel: Grant 5-Year Exclusivity for Ethical Drugs*, GLOBES (Israel), (July 1, 2004), <http://www.globes.co.il/serveen/globes/docview.asp?did=810543&fid=942>.

controls should exclude generic products; (2) developing country differential prices should not be used in high income country external reference pricing systems; and (3) price controls should be stable over long periods of time. The last two conclusions are likely to meet more controversy: (4) optimization of rents is preferable to price-fixing and reference pricing; and (5) PhRMA company data should be more transparent on a global basis.

First, generic pharmaceutical products must be excluded from price controls. The special case for government intervention in pharmaceutical prices derives from the monopoly market power granted by the state to patented drugs. Generic products do not generate patent rents, and thus should be exempt.¹⁹⁷

Second, virtual forms of dysfunctional arbitrage must be blocked. High income markets should not utilize developing country differential prices as an external reference price within these countries.¹⁹⁸ At present, this is not a problem, as it appears that no high income country uses donor prices in its reference pricing system.

Third, price controls must be stable over long periods of time. Pharmaceutical research requires long lead times before marketing. Companies should receive accurate ex ante pricing signals that are reliable. Otherwise, companies will discount the current price signals for the political risk of more onerous price controls.

Fourth, the heuristic prefers optimization over price-fixing and

197. Internal reference pricing systems may refer to generic prices within the therapeutic class, but generics themselves should not be reimbursed under an internal reference pricing system. Inclusion is not warranted, and may actually keep the generic prices artificially high. No pro-innovation goal is served by artificially high generic prices, other than a very indirect and inefficient subsidy of the innovator companies.

198. F.M. Scherer & Jayashree Watal, *The Economics of TRIPS Options for Access to Medicines*, in *ECONOMICS OF ESSENTIAL MEDICINES*, *supra* note 3, at 32, 48-49 (arguing for a ban on external reference pricing which uses prices in low income nations). Just like physical arbitrage, this practice should be restricted only when it flows from poor to rich nations. External reference pricing within high income countries, or within low and middle income countries does not undermine differential pricing for the poor. *But see* Scherer & Watal, *supra*, at 49 (suggesting preventing parallel exports from any price-controlled country). Danzon and Towse address the external reference pricing problem by suggesting increased pricing obscurity and opacity so that the rock-bottom prices are not “directly observable.” DANZON & TOWSE, *supra* note 12, at 6, 16-17. Their solution is vigorously rejected by Médecins sans Frontières, which has been very active in negotiating price discounts and distributing ARVs in sub-Saharan Africa. MSF, UNTANGLING THE WEB, *supra* note 151; MSF, SURMOUNTING CHALLENGES, *supra* note 141, at 7.

reference pricing. In this context, the policy goal should be to take the widest possible advantage of nonrival access, limited only by setting the minimum level of rents necessary to ensure optimal R&D. Any modifications to the strength of the power to appropriate rents must be evaluated in this light, whether it falls in the domain of IP law, contract, market regulation, national drug regulation, or trade agreements.

By contrast, price-fixing implies a price level without considering these other issues. Reference pricing schemes also may proceed automatically. By contrast, the reimbursement systems in Australia and the United Kingdom illustrate two different optimization approaches which support innovation.

In Australia's Pharmaceutical Benefits Scheme (PBS), each new drug must be approved under an economic evaluation process if governmental reimbursement is desired. The company must submit a dossier to the Pharmaceutical Benefits Advisory Committee (PBAC) proposing a price for the drug and supporting the economic efficiency of that price, given the drug's clinical advantages over existing therapies. In other words, Australia pays for value: Highly innovative drugs receive a much higher price; me-too drugs are priced with the lowest-cost equivalent. The incentives are obvious.¹⁹⁹

The United Kingdom's National Institute of Clinical Effectiveness (NICE) also performs economic evaluation of drugs, but targets a drug company's UK return on investment for its drug portfolio to the FTSE 100 London stock market index. One can argue about transfer pricing games and whether the FTSE 100 is an appropriate target, but the overall structure of the program is designed to support a reasonable return on R&D investment.²⁰⁰

Finally, greater transparency is warranted. Although biological data from clinical trials is generally applicable worldwide, many NDRAs accept confidentiality restrictions on data submitted for marketing approval and

199. LOKUGE & FAUNCE, *supra* note 166, at 7-8; SANJOY ROY, W. VA. PHARM. COST MGMT. COUNCIL, AN OVERVIEW OF PHARMACEUTICAL REFERENCE PRICING AND REIMBURSEMENT: ANALYSIS OF THE AUSTRALIAN PHARMACEUTICAL BENEFITS SCHEME 12-23 (2004), <http://www.wvc.state.wv.us/got/pharmacycouncil/default.cfm>.

200. JOAN-RAMON BORRELL, PRICES OF MEDICINES: A CASE-STUDY OF THE IMPACT OF THE RATE-OF-RETURN REGULATION IN THE UNITED KINGDOM 22 (Univ. of Barcelona Working Paper, 1997) (finding the regulation largely ineffective in controlling UK medicine prices); SARAH F. JAGGAR, GEN. ACCOUNTING OFFICE, GAO/HEHS-94-30, PRESCRIPTION DRUGS: SPENDING CONTROLS IN FOUR EUROPEAN COUNTRIES 69-77 (1994) (France, Germany, Sweden, and the United Kingdom).

reimbursement, needlessly reinventing the wheel each time. These unnecessary costs raise rents without social benefit. The economic evaluation studies submitted to the Australian PBAC would be very helpful in formulary and reimbursement decisions worldwide.²⁰¹ Further, if certain forms of price controls are adopted, optimizing pharmaceutical rents will require accurate global data on pharmaceutical pricing, profitability, and innovation. This information is not currently available to independent researchers, forcing policy makers to rely on the DiMasi study of secret and unverified industry data.²⁰² It strains credulity to base important pharmaceutical policy decisions on secret industry data, unavailable for study by other researchers.

vii. Free Riders

The heuristic has additional implications for the free rider problem in pharmaceutical innovation. If the free rider is a low income country (or low income person), we can consider the situation either a gift or harmless nonrival use.²⁰³ Free riding by high income countries is a more complicated problem.

Most high income countries have created direct or indirect governmental reimbursement of prescription drugs. One cannot expect governments to passively accept third degree differential pricing dictated by the drug companies. Nor do governments accept Ramsey Optimal Pricing based upon the government's ability to pay. Governmental resources are too scarce to completely resist the monopsony power, with the possible (temporary) exception of the United States.²⁰⁴

Acting solely in the national interest, governments may negotiate for the lowest possible prices, unconcerned about the possible negative global

201. Outtersson, *supra* note 193, at 260-61.

202. DiMasi et al., *supra* note 29. The R&D expenditure data source for this study was a "confidential survey" returned from ten PhRMA companies, *id.* at 152, 156, as well as unverified PhRMA aggregate data, *id.* at 179.

203. Scherer, *Global Welfare*, *supra* note 135, at 1141.

204. The federal government is prohibited from exercising monopsony powers in the new Medicare Part D drug benefit. Medicare Prescription Drug Improvement and Modernization Act of 2003, Pub. L. No. 108-173, § 301, 117 Stat 2066 (to be codified at § 1808(c)(1)(C) of the Social Security Act, 42 U.S.C. 139b-9). However, the United States is not entirely immune to rate-setting inclinations in health care. Almost every other major health care good or service purchased by Medicare or Medicaid is subject to rate-setting, including the services of physicians, hospitals, ambulatory surgical centers, and home health agencies.

effects on innovation. PhRMA companies may respond by raising prices in uncontrolled markets. The United States is the largest such market. Put bluntly, high income countries with price controls are said to be free riders on American innovation.²⁰⁵

Whether the free rider thesis is true empirically is an open question.²⁰⁶ Perhaps the crusade against the scourge of low-priced drugs is misplaced. Perhaps American prices are supra-optimal, and Canadian prices are optimal.²⁰⁷ Other countries may make up for their lower prices with higher volumes, eliminating the free rider problem. In many EU countries, drug prices are lower but account for a higher percentage of health expenditure than in the United States.²⁰⁸ It may be unfair to label such countries as free riders. Empirical doubts are also raised when the United States tolerates significant domestic free riders without apparent harm. Canadian prices are similar to the Federal Supply Schedule. Some Medicaid rebates and the U.S. Public Health Service's 340b program get better deals than Australia or Canada.²⁰⁹ Before one picks up stones to cast, check the glazing at home. PhRMA companies act as if the empirical question is beyond doubt, proceeding apace to the solution phase. Answering these questions properly requires transparent access to confidential company data. In any event, free riding is an innovation problem only if global pharmaceutical rents are sub-optimal.

The current PhRMA company solution is to use U.S. free trade agreements to raise drug prices outside of the United States.²¹⁰ To this end,

205. U.S. Editorial, *Lower US Prices Through Higher International Prices*, 9 PHARMA PRICING & REIMBURSEMENT 222 (2004).

206. Kevin Outterson, *Free Trade Against Free Riders?*, 9 PHARMA PRICING & REIMBURSEMENT 254-55 (2004).

207. I thank Professor Jim Friedberg for this suggestion. The free rider hypothesis assumes a joint sunk cost, but another possibility is that lower-priced countries such as Canada are efficiently avoiding waste. One empirical study suggests that PhRMA companies still make sufficient profits on Canadian sales, undercutting the free rider hypothesis. SAGER & SOCOLAR, *supra* note 11, at 1.

208. ORG. FOR ECON. COOPERATION AND DEV., OECD HEALTH DATA 2004, tbl.14 (2004), http://www.oecd.org/document/16/0,2340,en_2825_495642_2085200_1_1_1_1,00.html (noting that Canada, France, Germany, Switzerland, and the UK have higher total expenditures on pharmaceuticals as a percentage of total expenditure on health as compared to the United States).

209. WEST VIRGINIA REPORT, *supra* note 175, app. A-2, A-2 (comparing Medicaid, private payor, 340b, Canadian, FSS, and Australian data).

210. Peter Drahos & David Henry, *The Free Trade Agreement Between Australia and the United States: Undermines Australian Public Health and Protects U.S. Interests in Pharmaceuticals*, 328

USTR recently created the post of Assistant United States Trade Representative for Pharmaceutical Policy. Bilateral treaties are an awkward response to this global coordination problem. USTR may succeed in raising drug prices in the least appropriate places. The greatest success will be found in the poorest countries, or other smaller countries desperately seeking preferential access to the U.S. market.²¹¹ Small, poor countries offered a free trade deal with the United States may well agree to provisions which undermine health in order to serve commercial interests. But these small and generally poor markets can make very little contribution to the global fight against pharmaceutical free riders. The U.S. stance should be the opposite: Low income markets are the best targets for the enlightened policy of nonrival access.

If the USTR's solution is to be significant for innovation, it must involve the EU and Japan, but the USTR will find them better positioned to resist bilateral U.S. pressure to modify sensitive domestic health policy. Nor is there any guarantee that increased prices abroad will result in lower prices in the United States. A strategy which depends upon offending America's best trading partners should be preceded by proof that innovation and access will be improved. The ultimate free riders are counterfeiters, not governments, and any strategy to increase global pharmaceutical prices will increase the opportunity for counterfeits.²¹²

BRIT. MED. J. 1271-72 (2004); Elizabeth Becker, *Drug Industry Seeks To Sway Prices Overseas*, N.Y. TIMES, Nov. 27, 2003, at A5; Elizabeth Becker & Robert Pear, *Trade Pact May Undercut Importing of Inexpensive Drugs*, N.Y. TIMES, July 12, 2004, at A1; Marilyn Werber Serafini, *Drug Prices: A New Tack*, NAT'L J., Apr. 17, 2004, at 1177 ("So [House Speaker] Hastert and [Senator] Kyl championed the novel idea that the key to lowering U.S. prescription drug prices is to persuade foreign governments to raise their prices. . . . The idea of trying to level the international playing field on prescription drug pricing originated with the U.S. pharmaceutical industry. But Hastert and Kyl played significant roles last fall in persuading the Bush administration to embrace this strategy. . . . The result was the United States' first free-trade agreement that included modest concessions on pharmaceutical price controls."); Marilyn Werber Serafini, *The Other Drug War*, NAT'L J., Mar. 20, 2004, at 871-72; Mark B. McClellan, Speech Before the First International Colloquium on Generic Medicine (Sept. 25, 2003), <http://www.fda.gov/oc/speeches/2003/genericdrug0925.html>.

McClellan's speech was widely reported. See, e.g., Christopher Bowe & Geoff Dyer, *Americans Lured by Lower Prices*, FIN. TIMES, May 5, 2004, at 17 ("The rhetoric intensified in September when Mark McClellan, then head of the FDA, attacked European drug price controls and said other rich nations should pay more of the development cost for drugs.").

211. Witness the TRIPS-plus provisions in negotiated or pending FTAs with Morocco, Singapore, Jordan, Israel, Central America (CAFTA), and the Western Hemisphere (FTAA). See *supra* note 196 and accompanying text.

212. Aidan Hollis may well be the first to make this connection to counterfeiting explicit.

Other forms of global coordination should be considered, such as James Love and Tim Hubbard's Global R&D Treaty.²¹³ The R&D Treaty would serve as a global coordination mechanism amongst the high income countries, while permitting prices to decline to marginal manufacturing costs since R&D would no longer be recovered through the price mechanism. At lower price levels, access is greatly improved and the opportunity for counterfeits diminishes.

viii. Neglected and Global Diseases

a. Neglected Disease Innovation Does Not Require Increased Appropriation of Pharmaceutical Rents in Low Income Countries

Jean Lanjouw and Alan Sykes each support the enactment of IP laws in low income countries to encourage the development of local markets for treating neglected diseases.²¹⁴ Lanjouw cites empirical results from India suggesting that implementation of TRIPS is encouraging the largest Indian pharmaceutical companies to invest in R&D for new chemical entities (NCEs),²¹⁵ but those NCEs are either me-too generics or target global diseases.²¹⁶ Sykes argues that the huge disease burden in the developing

Aidan Hollis, An Efficient Reward System for Pharmaceutical Innovation (July 2, 2004) (discussing prizes based upon therapeutic value) (unpublished manuscript, on file with author) [hereinafter Hollis, Efficient Reward System]. See *infra* Subsection II.A.2 for a description of counterfeits.

213. JAMES LOVE, FROM TRIPS TO RIPS: A BETTER TRADE FRAMEWORK TO SUPPORT INNOVATION IN MEDICAL TECHNOLOGIES (Workshop on Economic Issues Related to Access to HIV/AIDS Care in Developing Countries, 2003); Hubbard, *supra* note 13; Love, *supra* note 13.

214. JEAN O. LANJOUW, A PATENT POLICY PROPOSAL FOR GLOBAL DISEASES 4 (The Brookings Institution, Working paper No. 84, 2001) (on file with author); Sykes, *supra* note 25, at 58-62.

215. JEAN O. LANJOUW, THE INTRODUCTION OF PHARMACEUTICAL PRODUCT PATENTS IN INDIA: 'HEARTLESS EXPLOITATION OF THE POOR AND SUFFERING?' (1998), <http://papers.nber.org/papers/wb366>.

216. Hannah E. Kettler & Rajiv Modi, *Building Local Research and Development Capacity for the Prevention and Cure of Neglected Diseases: The Case of India*, 79 BULL. WORLD HEALTH ORG. 742, 744-45 (2001) (finding that Indian companies are likely to target the largest markets, i.e., for global diseases rather than neglected diseases). A decade after the signing of TRIPS, a leading Indian pharmaceutical company reports that indeed its R&D budgets are growing rapidly, from 2.7% of sales in 2000 to 7.6% in 2003 and a projected 10% in 2004, but the primary output are generic pharmaceuticals. Adam Levitt, Dr. Reddy's Laboratories:

world should stimulate markets if patents were available. He thus looks to use IP laws to extract a greater portion of consumer surplus from the developing poor, in order to strengthen the incentives to innovate.²¹⁷ Surely this burden should be imposed on the world's poorest people only as a last resort. We should not demand the widow's mite in order to fund PhRMA.

Moreover, strong IP laws in low income countries are simply and unfortunately insufficient to create new markets for neglected disease drugs. If most patients in such countries are unable to purchase neglected disease drugs in commercial quantities and prices, the offer of patent protection will not stimulate R&D.²¹⁸ An exclusive offer to sell drugs at a loss is not valuable.²¹⁹ Profit-maximizing Indian drug companies will focus

Driving Growth 17-25 (Bear Stearns Healthcare Conference, Sept. 8, 2003) (on file with author) [hereinafter Levitt, Dr. Reddy's Laboratories]. The primary new drug application filed by the company is amlodipine maleate, which is the salt version of an innovative drug, Norvasc. The NDA is being opposed in federal court by the innovator company. *Id.* at 20. Of the eight NCEs in the company's pipeline, seven will treat global diseases such as diabetes, cancer, metabolic disorders, and cardiovascular disease. The eighth is an anti-infective drug, also for global diseases, but with more applicability in developing countries. *Id.* at 27. These are hardly the type of innovations that Lanjouw hoped for, and in fact this activity could hurt global innovation by reducing expected rents to innovator companies through early generic entry by aggressive Indian companies.

217. Sykes, *supra* note 25, at 61-62. Notably, Sykes has critiqued F.M. Scherer on the question of the net value of IP laws for developing countries.

218. The relative size of the commercial and non-commercial markets is important here. The growth of India and China's middle and upper classes one day will be sufficient to support commercial pricing of innovative drugs for conditions endemic only to the developing world. PhRMA companies do recognize a growing middle class market in these nations. Merck, SEC Form 10-k, *supra* note 173, at 14 ("In recent years, the Company has been expanding its operations in countries located in Latin America, the Middle East, Africa, Eastern Europe and Asia Pacific where changes in government policies and economic conditions are making it possible for the Company to earn fair returns. Business in these developing areas, while sometimes less stable, offers important opportunities for growth over time.").

219. Keith E. Maskus, *Ensuring Access to Essential Medicines: Some Economic Considerations*, 20 WISC. INT'L L.J. 563, 574 (2002) (casting doubt on the efficacy of patents to improve R&D on neglected drugs); see Kettler & Modi, *supra* note 216, at 742 (noting that Indian pharmaceutical companies will still require financial incentives to research and develop drugs for neglected diseases). A recent study of neglected vaccine projects found patent incentives to be completely ineffective. JASON C. HSU & EDUARDO S. SCHWARTZ, A MODEL OF R&D VALUATION AND THE DESIGN OF RESEARCH INCENTIVES 37, 43-45 (Nat'l Bureau of Econ. Research, Working Paper No. 10041, 2003).

on their best economic opportunities;²²⁰ neglected disease drugs will not be at the top of that list.²²¹ The leading Indian drug companies derive most of their profits from sales in the United States and other high income markets.²²² Nor are strong IP laws important to develop indigenous manufacturing capacity. The absence of pharmaceutical patents in India was the proximate cause of India's vibrant generic pharmaceutical sector. Implementation of TRIPS and restrictions on PEPFAR procurement will hinder this path of development.²²³

Developing non-OECD pharmaceutical R&D capacity has the potential to improve the efficiency of global research. Non-OECD PhRMA companies may have significantly lower cost structures, enabling R&D on disease markets with less market potential. Cipla, Ltd. and other Indian pharmaceutical companies pay their India-based chemists and investigators a fraction of the prevailing OECD pharmaceutical company research wages. These companies may also be better poised to understand and respond to the developing market and less likely to discount the actual market size due to unfamiliarity. Network effects and sunk costs are also present in pharmaceutical sales and marketing: While OECD companies have invested in marketing systems in OECD countries, emerging companies may invest in regional markets heretofore overlooked by OECD

220. Kettler & Modi, *supra* note 216, at 745. For the leading Indian pharmaceutical company, in early 2004 only a negligible percentage of sales were of New Chemical Entities (NCEs). Most sales were either active pharmaceutical ingredients (APIs, i.e. intermediate ingredients for drugs) to the United States and Europe or branded (generic) formulations sold in India and other similar markets. Levitt, Dr. Reddy's Laboratories, *supra* note 216, at 9-10.

221. Jean O. Lanjouw & Iain Cockburn, *New Pills for Poor People?: Empirical Evidence After GATT*, 29 WORLD DEV. 265, 265-89 (2001) (finding in 1998 survey of Indian drug firms that only 16% of the firms' R&D targeted developing country markets). In fiscal year 2002-2003, Cipla's major innovative introduction was TIOVA, a long-acting bronchodilator for Chronic Obstructive Pulmonary Disease (COPD), a global disease. Cipla also launched a new generic ARV Fixed Dose Combination (FDC). CIPLA, SIXTY-SEVENTH ANNUAL REPORT 2002-2003, at 5 [hereinafter CIPLA 2002-2003 ANNUAL REPORT].

222. See, e.g., CIPLA 2002-2003 ANNUAL REPORT, *supra* note 221, at 7 ("During the year, Cipla's strategic alliances with leading generic companies in the USA and Europe were expanded to include additional products and projects. Currently, there are nearly 50 such projects in various stages of development in the USA alone."); Rasul Bailay, *Cipla May Find Right Rx for Success: Indian Drug Firm Partners with Peers in U.S. To Crack No. 1 Market for Generics*, WALL ST. J., Oct. 20, 2003, at A15. For Dr. Reddy's Laboratories, the U.S. market accounted for fifty-seven percent of 2003 gross margin. Levitt, Dr. Reddy's Laboratories, *supra* note 216, at 11.

223. On PEPFAR procurement, see *infra* Subsection II.B.1.

companies²²⁴ and invest in process developments to lower production costs.²²⁵

Most neglected disease conditions lack a market not because of the absence of IP rights in low income countries but because of the poverty of the patients. Perhaps the best description of a neglected disease drug is that market-based innovation is unlikely because the target population will require the drug or vaccine to be distributed at or below the lowest possible marginal cost of production. Any such drug will require non-market funding for innovation and distribution, with or without IP regimes.

Michael Kremer's model of a donor purchase commitment is a prominent example of a non-market mechanism,²²⁶ attracting many commentators on the proper design of such a prize.²²⁷ Prize systems and donor purchase commitments do not require IP laws in low income countries. The donor may reference the patent law of some country (such as the United States), without requiring the target populations to have any IP laws at all. The appropriate incentives are in place so long as the donor is bound to a credible commitment to act *as if* they are bound by the IP laws of a reference country such as the United States. This process would create a "reference" or "virtual" IP regime. This is a significant point, not well developed by supporters of TRIPS implementation in low income countries. Virtual IP regimes would achieve all of the claimed advantages of TRIPS implementation in low income countries for prizes, without the blocking effects of local IP laws.

224. India, Russia, China, Brazil, Mexico, Africa, and other markets are major markets for Indian companies such as Dr. Reddy's Laboratories. Levitt, Dr. Reddy's Laboratories, *supra* note 216, at 17; *see also* Kettler & Modi, *supra* note 216, at 743 (describing the Indian pharmaceutical industry).

225. Kettler & Modi, *supra* note 216, at 743-45. But Kettler and Modi do not assume an Indian comparative advantage in cost.

226. Michael Kremer has thoughtfully analyzed and articulated the donor purchase commitment model. Michael Kremer, *Creating Markets for New Vaccines: Part I: Rationale & Part II: Design Issues*, in 1 INNOVATION POLICY AND THE ECONOMY 35-109 (Adam B. Jaffe et al. eds., 2001).

227. *See, e.g.*, Hollis, Efficient Reward System, *supra* note 212; Michael Abramowicz, *Perfecting Patent Prizes*, 56 VAND. L. REV. 115 (2003); Steven Shavell & Tanguy Van Ypersele, *Rewards Versus Intellectual Property Rights*, 44 J.L. & ECON. 525 (2001) (concluding that optional reward systems are superior to IP rights systems).

*b. Global Disease Innovation Does Not Require Increased
Appropriation of Pharmaceutical Rents in Low Income Countries*

The neglected disease debate tends to overlook the fact that the chronic conditions of the high income and low income worlds are converging.²²⁸ Global diseases²²⁹—conditions which affect patients in both rich and poor countries—include cancer and cardiovascular disease,²³⁰ as well as AIDS.²³¹

228. Non-communicable disease accounts for forty-seven percent of the global burden of disease. WORLD HEALTH ORG., WHO GLOBAL STRATEGY ON DIET, PHYSICAL ACTIVITY AND HEALTH (May 22, 2004). Cancer and cardiovascular disease are the second and third largest causes of death in developing countries. WORLD HEALTH ORG., WORLD HEALTH REPORT 2003 (2003). Stephen Leeder et al., A RACE AGAINST TIME: THE CHALLENGE OF CARDIOVASCULAR DISEASE IN DEVELOPING ECONOMIES 12-15 (2004) (“In 1998, non-communicable diseases were responsible for 59% of total global mortality and 43% of the global burden of disease. Importantly, 78% of [non-communicable disease] deaths were borne by low- and middle income countries, as was 85% of the NCD burden of disease . . . nearly 50% of deaths worldwide were due to CVD, diabetes, cancer and chronic lung disease.”). PhRMA agrees with this position when it argues that the current “Western oriented” R&D program actually includes diseases endemic to the entire world, such as cancer and CVD. Response of the Research-Based Pharmaceutical Industry to the Interim Report of the Task Force on Access to Essential Medicines (Feb. 1, 2004), *reprinted in* DUKES, *supra* note 84, app. 2, at 7-8.

229. Herein, the term global disease refers to conditions for which a therapeutic market exists in high income countries, and the condition is also endemic to the low or middle income world. The definition of global disease is not static. Malaria was once a global disease, but is now largely eradicated in high income countries, rendering it neglected. Diseases may also move in the opposite direction. Increased international mobility is likely to further blur the epidemiological effect of political borders, causing neglected diseases to migrate into the global disease category. The eastward expansion of the EU is importing additional infectious disease threats into the EU, requiring enhanced public health responses to tuberculosis and AIDS. Richard J. Coker et al., *Health-care System Frailties and Public Health Control of Communicable Disease on the European Union’s New Eastern Border*, 363 THE LANCET 1389-92 (2004).

230. Carlos Correa interprets the *Doha Declaration* to include global diseases such as asthma and cancer. CORREA, IMPLICATIONS OF DOHA, *supra* note 137, at 5. Others ask why diseases such as cancer and diabetes are not covered by Doha. Julian Fleet, *U.N. Approach to Access to Essential AIDS Medications, Intellectual Property Law and the WTO TRIPS Agreement*, 17 EMORY INT’L L. J. 451, 465 (2003).

231. North America and Western Europe account for less than two million of the thirty-four to forty-six million people living with HIV/AIDS in 2003. UNAIDS/WHO, AIDS EPIDEMIC UPDATE 37 (2003) [hereinafter UNAIDS/WHO, AIDS EPIDEMIC UPDATE]. While AIDS is a global disease, at least two market failures plague public health. First, one strain of AIDS (Type A) is largely confined to the developing world, and thus receives less research

As an example of crossover potential of global diseases, consider the WHO Prequalification Project. The WHO has requested prequalification dossiers on four cancer drugs (vinblastine, etoposide, bleomycin and vincristine)²³² and two have been prequalified.²³³ These drugs are all related to the treatment of AIDS-related cancers and are off-patent in the United States.²³⁴ For the treatment of TB, the WHO prequalified non-licensed generic forms of patented ciprofloxacin from India and Spain.²³⁵ But these drugs may be used to treat conditions other than TB and AIDS related cancers.²³⁶ The Doha Declaration was not limited to these three diseases, despite U.S. efforts to narrow the exception.²³⁷

To the pharmaceutical industry, extending nonrival access to global disease drugs outside of AIDS, malaria, and TB opens Pandora's Box. Roger Bate acknowledges that the United States negotiated to limit flexibilities to twenty two diseases, keeping "lifestyle complaints and major western diseases" off the table.²³⁸ Nonrival access should not be limited to these three diseases, or narrow "on label" uses, but should be extended to any global disease, on the basis of global disease burden and public health need. Furthermore, this extension will not adversely affect innovation. The most important proposition about global diseases is that a robust level of

attention. Second, pediatric AIDS is also primarily a developing country issue, including the debates over the use of Nevirapine and the absence of pediatric formulations of most AIDS drugs. MÉDECINS SANS FRONTIÈRES, UNTANGLING THE WEB OF PRICE REDUCTIONS: A PRICING GUIDE FOR THE PURCHASE OF ARVs FOR DEVELOPING COUNTRIES 5 (4th ed. 2003) ("Children living with HIV/AIDS are one of the most neglected populations: pediatric formulations are lacking and/or formulations do not meet children's and caregivers' needs . . . unpleasant tasting syrup, tablets too big to swallow, need to refrigerate some products, unbreakable tablets, lack of fixed dose combinations, and non-adapted dosages. For example there are currently no combinations for paediatric use.").

232. WORLD HEALTH ORG., ACCESS TO HIV/AIDS DRUGS AND DIAGNOSTICS OF ACCEPTABLE QUALITY, PROCUREMENT QUALITY AND SOURCING PROJECT (15th ed. 2004) [hereinafter WHO HIV/AIDS PREQUALIFICATION, 15th ed.], <http://mednet3.who.int/prequal/>.

233. Those two drugs are vinblastine and vincristine. WHO HIV/AIDS PREQUALIFICATION, 15th ed., *supra* note 232.

234. FDA ORANGE BOOK (Oct. 2004), <http://www.fda.gov/cder/ob/default.htm>.

235. WHO HIV/AIDS PREQUALIFICATION, 15th ed., *supra* note 232.

236. According to the FDA label, vincristine is indicated in acute leukemia, Hodgkin's disease, non-Hodgkin's malignant lymphomas, rhabdomyosarcoma, neuroblastoma, and Wilms' tumor. Ciprofloxacin is a widely used antibiotic.

237. CORREA, IMPLICATIONS OF DOHA, *supra* note 137, at vii, 15-16 (discussing the definition of emergency); 't Hoen, *supra* note 7, at 39-42.

238. Roger Bate, *Entire IP System Could Easily Fall*, FIN. TIMES (London), Aug. 18, 2003, at 10.

innovation is assured by high income markets alone. A few hundred thousand early AIDS cases in the United States (and government funding) were sufficient to encourage successful research programs.²³⁹ Likewise, aggressive research programs are underway to treat most or all of the chronic conditions endemic in the high income countries. Since global disease knowledge is nonrival, it can be offered to low income populations without detriment. With innovation assured, IP law can stand aside and permit nonrival access for the poor.

Together, these implications suggest a new approach to the innovation-access conundrum, calling for a radical re-evaluation of the role of TRIPS and other laws to encourage nonrival access, with substantial potential gains in global public health.

II. THE PRAXIS OF PHARMACEUTICAL ARBITRAGE

In Part II of this Article, the theory of pharmaceutical arbitrage will be placed in two different contexts: the AIDS crisis in sub-Saharan Africa and prescription drug importation from Canada to the United States.

Certain forms of pharmaceutical arbitrage are dysfunctional—for example, diversion of differentially priced ARVs from sub-Saharan countries into high income countries. Other forms of arbitrage benefit consumers without damaging optimal innovation. The desirability of Canadian-U.S. pharmaceutical arbitrage hinges on whether global pharmaceutical rents are supra-optimal or not and whether one credits the safety of Canadian sourced drugs. If global pharmaceutical rents are supra-optimal and safety concerns properly addressed, then U.S. consumers are needlessly overcharged for patented drugs, and many unnecessarily suffer negative health outcomes from restricted access.

A. Pharmaceutical Arbitrage of AIDS Drugs in Sub-Saharan Africa

PhRMA companies have been reluctant to make patented ARV drugs available on a nonrival basis in sub-Saharan Africa. Fear of pharmaceutical arbitrage and undermining IP laws are the purported causes of this reluctance. Delayed treatment has been the proximate cause of the death of millions. Applying the theory of pharmaceutical arbitrage to AIDS may transcend the competing goals of innovation and access, by improving access while supporting optimal R&D.

239. Indeed, many early AIDS-related drugs qualified for orphan drug status in the United States when the expected U.S. market was fewer than 200,000 persons. Salbu, *supra* note 103, at 703-07.

1. Financial Constraints Limit Access to AIDS Drugs in Sub-Saharan Africa

Globally, AIDS is not under control, with approximately forty million persons living with HIV/AIDS worldwide.²⁴⁰ Ninety-five percent live outside of North America and Western Europe. Two-thirds of infected persons, new infections, and deaths are in sub-Saharan Africa.²⁴¹ An estimated 5.5 million people in developing countries need ARV treatment for HIV/AIDS, but only five percent of those currently receive it; in sub-Saharan Africa in 2003, only one percent of the people who need ARV therapy actually receive it.²⁴² Large scale roll-out of ARV therapy in low income countries is now a major global public health goal.²⁴³

Purchasing AIDS drugs at U.S. prices is not an option for the vast majority of these people. The per capita annual cost of a popular first-line ARV in the United States is \$7215,²⁴⁴ and the recently introduced Fuzeon (enfuvirtide) costs \$20,000 per year.²⁴⁵ The annual per capita health expenditures in sub-Saharan Africa averages \$29.30²⁴⁶ and range from \$12

240. UNAIDS/WHO, AIDS EPIDEMIC UPDATE, *supra* note 231, at 2. While much progress has been made, AIDS is not fully under control in high income countries. In 2003, 66,000 to 94,000 persons were newly infected with HIV in North America and Western Europe. *Id.* at 38. But these numbers are quite small when compared to sub-Saharan Africa, and the health and longevity of the U.S. patients have improved. *Id.* at 28-30 (“AIDS mortality continues to drop, thanks to the widespread availability of antiretroviral treatment.”).

241. UNAIDS/WHO, AIDS EPIDEMIC UPDATE, *supra* note 231, at 38; Robert Greener, UNAIDS, *HIV/AIDS and Absorptive Capacity* (Kaiser Family Foundation HealthCast, Jan. 29, 2004) (2003 data) [hereinafter Greener, UNAIDS], http://www.kaisernetwork.org/health_cast/hcast_index.cfm?display=detail&hc=1066.

242. MSF, SURMOUNTING CHALLENGES, *supra* note 141, at 2, 5; *UN To Seek \$6 Billion To Fight AIDS in Third World*, REUTERS, Nov. 6, 2003.

243. INST. OF MED., SCALING UP TREATMENT FOR THE GLOBAL AIDS PANDEMIC: CHALLENGES AND OPPORTUNITIES (2004); World Health Assembly, *Scaling up Treatment*, *supra* note 190.

244. Sixty tablets of Combivir cost \$592.99. Drugstore.com, at <http://www.drugstore.com> (last visited Dec. 4, 2004).

245. Vanessa Fuhrmans, *Medical Dilemma: Costly New Drug for AIDS Means Some Go Without*, WALL ST. J., Jan. 13, 2004, at A1. Fuzeon is the first fusion inhibitor treatment for HIV, developed at Duke University. Ironically, high cost has forced the North Carolina AIDS assistance project to strictly ration the number of residents who can receive the treatment. *Duke University: North Carolina Firm's New AIDS Drug Development On Hold*, U-WIRE, Jan. 22, 2004, 2004 WL 59460572 (“Steve Sherman, director of North Carolina’s ADAP, said the program set a cap for 25 state residents to be eligible for Fuzeon treatment at any one time, creating a system of rationing medical care.”). Other states such as Alabama have decided the cost is too high to cover the drug at all, despite its effectiveness. Fuhrmans, *supra*.

246. WORLD BANK, 2004 WORLD DEVELOPMENT INDICATORS (2004) (citing 2001 data).

(Malawi) to \$253 (South Africa).²⁴⁷ Radically reducing the price of AIDS medications for the poor is thus a necessary condition to extending ARV treatments to millions of afflicted persons worldwide.²⁴⁸ Indeed, for many patients, the drugs must be free. Recognizing the important public health issues, Brazil,²⁴⁹ India,²⁵⁰ South Africa,²⁵¹ and China²⁵² produce generic ARVs for the poor, provoking conflicts between human rights and IP rights.

The European Commission has embraced voluntary “tiered [differential] pricing as the principal means of rendering essential medicines affordable . . . to the poorest populations.”²⁵³ Differential pricing is possible because of relatively low marginal costs of production. Most patented drugs can be produced relatively cheaply, absent R&D cost recovery. The primary variable expenses are direct manufacturing costs, which are a small fraction of the retail prices of patented ARVs. A high ratio of retail prices to direct manufacturing costs enables a company to sell at highly differentiated prices without selling below marginal cost.²⁵⁴

247. *Id.*; see also Markus Haacker, *Providing Health Care to HIV Patients in Southern Africa*, in ECONOMICS OF ESSENTIAL MEDICINES, *supra* note 3, at 242, 244. After adjustments for purchasing power parity, Haacker’s figures rise to \$44.8 (Malawi) and \$552.3 (South Africa).

248. Funds for ARVs and drugs to treat opportunistic infections are scarce. UNAIDS estimates these needs at approximately thirty-seven percent of the total \$10.7 billion which should be spent on HIV/AIDS in 2005 for a comprehensive response. Total unmet financial need in 2005 is projected at approximately five billion dollars. Greener, UNAIDS, *supra* note 241. If these drugs were available at a much lower cost, resources could be redeployed to prevention and other unmet priorities.

249. ‘t Hoen, *supra* note 7, at 32-33.

250. Mark Schoofs, *Clinton Program Would Help Poor Nations Get AIDS Drugs*, WALL ST. J., Oct. 23, 2003, at B1 (Indian and South African drug companies); see also CIPLA 2002-2003 ANNUAL REPORT, *supra* note 221, at 7 (“In HIV/AIDS care, the Company continued its pioneering role in making available a range of antiretroviral drugs including unique combination products. These were made available at reasonable prices not only in India but also in other parts of the world.”).

251. Schoofs, *supra* note 250 (Indian and South African drug companies); ‘t Hoen, *supra* note 7, at 30-31 (describing South Africa’s efforts to provide royalty-free ARVs to its population and the legal and political challenges to those actions by the United States and PhRMA companies).

252. Jim Yardley, *China Begins Giving Free H.I.V./AIDS Drugs to the Poor*, N.Y. TIMES, Nov. 7, 2003, at A3.

253. DG TRADE, *supra* note 42, § 2.2. Low income countries targeted for essential medications by the EU had a per capita income of less than \$765 in 2000.

254. SAGER & SOCOLAR, *supra* note 11, at 7 (roughly estimating marginal U.S. manufacturing and distribution costs for prescription drugs to be 9.9%).

While the public does not know the true marginal manufacturing costs of most patented drugs, differential pricing and generic production provide useful proxies.²⁵⁵ Differential pricing ratios currently exceed 30:1 in ARV drugs, implying marginal costs of production in the range of 3 to 4%. For example, in November 2003, a daily dose of GlaxoSmithKline's best selling combination ARV drug Combivir²⁵⁶ costs about \$19.76 per day or \$7215 per year by mail order in the United States.²⁵⁷ In sub-Saharan Africa in 2003, GlaxoSmithKline sells Combivir to health agencies at ninety cents per day or \$329 per year,²⁵⁸ and has announced a new non-profit price of sixty-five cents per day.²⁵⁹ Even this low price may not reflect GlaxoSmithKline's marginal cost, because Cipla sells a generic form of Combivir to governments and nonprofit agencies at thirty-three cents per day or \$197.10 per year.²⁶⁰ The differential pricing ratio for Combivir is

255. Compulsory licensure enables ex-factory pricing closer to true marginal manufacturing cost, particularly if the tender process is competitive. Generic competition pierces the pricing veil, accelerates differential pricing toward true marginal production costs, and does not rely on public disclosure of confidential financial information from the companies. Given the endemic opacity of all PhRMA data on costs, perhaps the best way to calculate marginal cost is through compulsory licensure. PhRMA simply asserts that "there is no guarantee that generic companies will price at marginal cost." Response of the Research-Based Pharmaceutical Industry to the Interim Report of the Task Force on Access to Essential Medicines, *supra* note 228, *reprinted in* DUKES, *supra* note 84, app. 2, at 27. Absent the patent monopoly, generic companies in a competitive environment will certainly price much closer to marginal cost than PhRMA companies.

256. Combivir is a fixed dose combination (FDC) of 300 mg zidovudine (ZDV or AZT) and 150 mg of lamivudine (3TC). MSF, UNTANGLING THE WEB, *supra* note 151, at 13. The best clinical FDC also adds a NNRTI. Gregory K. Robbins et al., *Comparison of Sequential Three-Drug Regimens as Initial Therapy for HIV-1 Infection*, 349 NEW ENG. J. MED. 2293 (2003); Robert W. Shafer et al., *Comparison of Four-Drug Regimens and Pairs of Sequential Three-Drug Regimens as Initial Therapy for HIV-1 Infection*, 349 NEW ENGLAND J. MED. 2304 (2003).

257. Calculation of the U.S. price comes from Drugstore.com, at <http://www.drugstore.com> (sixty tablets of Combivir for \$592.99, taken twice per day) (last visited July 7, 2004).

257. 't Hoen, *supra* note 7, at 32-33.

258. MSF, UNTANGLING THE WEB, *supra* note 151, at 13.

259. GLAXOSMITHKLINE PLC, 2003 ANNUAL REPORT, Form 20-F, at 4, <http://www.sec.gov/edgar/searchedgar/companysearch.html>.

260. MSF, UNTANGLING THE WEB, *supra* note 151, at 15. GSK itself also issued a voluntary license to Aspen Pharmacare for production of generic Combivir. GLAXOSMITHKLINE PLC, 2003 ANNUAL REPORT, Form 20-F, *supra* note 259, at 29. Aspen's price is not yet listed in the pricing guide. MSF, UNTANGLING THE WEB, *supra* note 151, at 15.

approximately 35:1.²⁶¹ This ratio is likely to increase: MSF aims for an annual per patient cost of fifty to one hundred dollars in the near future.²⁶² The organization notes that achieving the lowest possible price is an urgent necessity: "If you have the option of spending \$200 per person per year or \$600 per person per year, and you're electing to spend \$600, that means you're treating one person when you could be treating three."²⁶³

Triomune is Cipla's brand name for the most important triple-drug therapy fixed dose combination (FDC) for sub-Saharan Africa, containing nevirapine (NVP), stavudine (d4T), and lamivudine (3TC). Triomune is produced as a generic by Cipla Ltd. and sold for sixty-seven cents per day or US\$244 per year.²⁶⁴ As of July 2004, Triomune is not available in a licensed FDC form, a rare inversion in which a generic is a sole-source supplier.²⁶⁵ The patents for nevirapine, stavudine, and lamivudine are held by different companies,²⁶⁶ and they are apparently unable to conclude a cross-licensing agreement. Triomune's components, taken as six separate pills per day, cost about \$936 per month in the United States, a ratio exceeding 46:1.²⁶⁷

High differential pricing ratios are not limited to ARVs. Ciprofloxacin is available in generic form in Africa at \$0.0189 per 500 mg tablet;²⁶⁸ in the U.S. retail market it sells for about five dollars a pill,²⁶⁹ a ratio of 264:1. A high ratio is not necessarily a bad thing; in fact, if nonrival access is truly

261. The numerator is \$7215, and the denominator is \$204.

262. MSF, SURMOUNTING CHALLENGES, *supra* note 141, at 9. As of 2004, the WHO 3 x 5 program estimates the per person per year cost for first-line drugs at US\$304. WORLD HEALTH ORG., THE WORLD HEALTH REPORT 2004: CHANGING HISTORY 30 (2004).

263. Ellen Nakashima & David Brown, *U.S. Rule on AIDS Drugs Criticized: Ban on Using Aid To Buy Foreign Generics Hinders Treatment, Experts Say*, WASH. POST, July 14, 2004, at A12 (quoting Rachel Cohen of MSF).

264. MSF, UNTANGLING THE WEB, *supra* note 151, at 15.

265. UNICEF-UNAIDS-MSF, SOURCES AND PRICES OF SELECTED MEDICINES AND DIAGNOSTICS FOR PEOPLE LIVING WITH HIV/AIDS 18 (2004) (NVP + d4T + 3TC fixed dose combination not available in the US).

266. FDA ORANGE BOOK, *supra* note 234.

267. Epivir (lamivudine) costs about \$9 per day or \$270 per month; Zerit (stavudine) costs about \$10.51 per day or \$316 per month; Viramune (nevirapine) costs about \$11.67 per day or \$350 per month. All data is in U.S. dollars and is taken from <http://www.drugstore.com> (last visited July 8, 2004). The ratio numerator is \$936, and the denominator is \$20.

268. Management Sciences for Health, International Drug Price Indicator Guide, at <http://erc.msh.org/dmpguide/> (last visited Oct. 20, 2004).

269. Drugstore.com, at <http://www.drugstore.com> (last visited July 8, 2004).

provided to the world's poorest communities, one would expect to find very large differential pricing ratios.

2. IP Laws Hinder Delivery of ARVs in sub-Saharan Africa

In a widely-cited 2001 study, Attaran and Gillespie-White demonstrated the relative paucity of ARV patents in many sub-Saharan countries.²⁷⁰ This article has been widely interpreted to claim that patents do not hinder ARV access in sub-Saharan Africa.²⁷¹ Attaran published a follow-on report in *Health Affairs* in 2004, again suggesting that patents have not been the major hindrance to ARV access.²⁷² This conclusion is not warranted from the data.

ARVs were available in the high income countries for many years before the developing world first began to receive treatment.²⁷³ As recently

270. Amir Attaran & Lee Gillespie-White, *Do Patents for Antiretroviral Drugs Constrain Access to AIDS Treatment in Africa?*, 286 JAMA 1186 (2001). After the manuscript was submitted, Merck gave a \$25,000 grant. Several critical letters to the editor were printed in the next volume of the journal. Boelaert et al., *supra* note 141, at 840-41; Eric Goemaere et al., Letter to the Editor, 287 JAMA 841 (2002); Michael J. Selgelid & Udo Schuklenk, Letter to the Editor, 287 JAMA 842 (2002) ("In the world of politics the carefully qualified conclusions of Attaran and Gillespie-White are likely to be misrepresented by pharmaceutical industry lobbyists claiming that 'it has been shown that patents do not matter,' with the aim of blocking proposed TRIPS agreement amendments that weaken pharmaceutical patent protection in developing countries."). In their reply to these letters, Attaran and Gillespie-White do not make the broad claim that patent laws are no barrier to ARVs in sub-Saharan Africa, but merely suggest that where patents exist, other alternatives can be pursued, such as voluntary licensure or switching to another therapy. Where patents do not exist, they call for generic production, ignoring the industrial infrastructure issue described above. Amir Attaran & Lee Gillespie-White, *In Reply*, 287 JAMA 842-43 (2002); *see also* Amir Attaran, *How Do Patents and Economic Policies Affect Access to Essential Medicines in Developing Countries?*, 23 HEALTH AFF. 155 (2004).

271. LANJOUW, INTELLECTUAL PROPERTY, *supra* note 70, at 11-12 ("[I]ndustry uses this fact [the Attaran & Gillespie-White study] to stress that patents in the poorest countries are not impeding access to drugs."); *see also, e.g.*, Harvey E. Bale, Jr., *Patents, Patients and Developing Countries: Access, Innovation and the Political Dimensions of Trade Policy*, in ECONOMICS OF ESSENTIAL MEDICINES, *supra* note 3, at 100, 106 n.10. Bale is the head of the international PhRMA company trade association.

272. Attaran, *supra* note 270, at 156 ("Briefly, I find that patents for essential medicines are uncommon in poor countries and cannot readily explain why access to those medicines is often lacking, suggesting that poverty, not patents, imposes the greater limitation on access.").

273. Combination therapy was available in the United States from December 1995 with

as 2002, no person in the developing world had received ARVs through official donor support from any country or multilateral institution.²⁷⁴ When MSF and Partners In Health independently began ARV treatment in Thailand, South Africa, and Haiti in 2000 and 2001, some were puzzled at their attempts, due to costs per patient exceeding \$10,000 to \$15,000 for patented drugs.²⁷⁵ Access to AIDS medications was discussed at the highest levels at the WHO as early as 1991,²⁷⁶ and at the International AIDS Conference in Durban in 2000.²⁷⁷ Thirteen years later, in 2004, the world is just beginning to scale-up toward universal provision of ARVs, and it is still expected to take a long time to achieve. Precious years were lost because the drugs were too expensive for the developing world, and they were too expensive because of patent protection and fears of arbitrage.²⁷⁸ Millions

the approval of the first protease inhibitors, Invirase (SQV) on December 7, 1995 and Crixivan (IDV) and Norvir (r) in early 1996. Lamivudine was approved for marketing in the United States on November 17, 1995. The U.S. Centers for Disease Control was well aware of the growing HIV/AIDS epidemic in Africa no later than the mid-1980s. RANDY SHILTS, AND THE BAND PLAYED ON: POLITICS, PEOPLE, AND THE AIDS EPIDEMIC 49, 193, 392-93, 460 (1988) ("Equatorial Africans faced death on the scale of the Holocaust.") (citing Dr. Don Francis, CDC AIDS Research, June 1984).

274. *African HIV/AIDS Crisis: Pursuing Both Treatment and Prevention: Hearing Before the Senate Comm. on Foreign Relations, Subcomm. on African Affairs*, 107th Cong. (2002) (statement of Jeffrey D. Sachs).

275. As of December 2000, the World Bank still considered ARV treatment in poor countries to be "cost-ineffective." See Barton Gellman, *An Unequal Calculus of Life and Death; As Millions Perished in Pandemic, Firms Debated Access to Drugs*, WASH. POST, Dec. 27, 2000, at A1 [hereinafter Gellman, *Unequal Calculus*].

276. *Id.*

277. Esther Kaplan, *Time's Up*, THE NATION, July 22, 2004, at 30, 30 ("It's been four years since the International AIDS Conference was first held in the developing world, in Durban, South Africa, where the activist demand for universal treatment access was catapulted onto the world stage. Then, the idea of treating the millions of HIV-infected people worldwide was considered farfetched. . . . The official policy of wealthy nations was to focus on prevention and leave the millions already infected to die.").

278. JOAN-RAMON BORRELL & JAYASHREE WATAL, IMPACT OF PATENTS ON ACCESS TO HIV/AIDS DRUGS IN DEVELOPING COUNTRIES (Center for Int'l Dev., Harvard Univ., CID Working Paper No. 92, 2002) (finding that a significant increase in ARV uptake would have resulted absent patents). This paper provides a static analysis which ignores the innovation question and does not model subsidized ARV markets. Had it done so, it might have demonstrated that patents have a much larger negative impact. Barton Gellman, *A Turning Point That Left Millions Behind; Drug Discounts Benefit Few While Protecting Pharmaceutical Companies' Profits*, WASH. POST, Dec. 28, 2000, at A1 ("For a decade, makers of AIDS medications had rejected the idea of lowering prices in poor countries for fear of eroding

have died, untreated, for the principle of IP law.

Attaran defends his conclusions by identifying many sub-Saharan countries wherein patents had not been filed for some ARVs. This fact is both misleading and irrelevant because the sub-Saharan countries where patents have not been filed did not possess the domestic industrial base to manufacture ARVs.²⁷⁹ Throughout almost all of 2004, only one company was producing generic ARVs in Africa—South Africa's Aspen Pharmacare.²⁸⁰ As one might expect, Attaran finds Aspen's home market, South Africa, to be effectively covered by patent filings.

With South Africa stymied, generic ARVs would have to be imported into sub-Saharan Africa from elsewhere, such as Brazil or India. Brazil was sued to block this practice,²⁸¹ and India has faced a U.S.-requested WTO dispute resolution on its implementation of TRIPS for pharmaceuticals,²⁸² as well as U.S. "Special 301" threats.²⁸³ The USTR frequently used the Special 301 watch list to discipline countries attempting to produce generics, even if legal under domestic law or TRIPS.²⁸⁴

The mere possibility of a patent filing is likely to act as a deterrent to the filing of generic drug applications and the creation of generic manufacturing capacity in sub-Saharan Africa, since the innovator could undercut the generic company's market investment by tying them up in litigation. Unfortunately, the intended lessons of the United States's attacks on Brazil, South Africa, and India were not lost on other developing countries: All but three of Africa's least developed countries have implemented laws for pharmaceutical patents as of 2004, despite the

profits in rich ones.").

279. CORREA, IMPLICATIONS OF DOHA, *supra* note 137, annex 2.

280. Wendell Roelf, *Aids Drugs Are Available—But Are There Enough?*, MAIL & GUARDIAN, May 17, 2004. Thembalami Pharmaceuticals in South Africa is a joint venture with Indian producer Ranbaxy, importing the APIs from India. In July 2004, Archy Pharmaceuticals Limited commissioned the creation of a pharmaceutical plant in Nigeria to produce ARVs. Godwin Haruna, *First HIV/AIDS Drug Plant Takes Off*, THIS DAY (Lagos), July 28, 2004, <http://allafrica.com/stories/200407280351.html>. By contrast, Asia has twenty-seven companies producing generic ARVs in eight countries. AMFAR, *supra* note 188, at 4.

281. See *supra* notes 144-145 and accompanying text.

282. World Trade Org., *India-Patent Protection for Pharmaceutical and Agricultural Chemical Products*, WT/DS50/R (Sept. 5, 1997); World Trade Org., *India-Patent Protection for Pharmaceutical and Agricultural Chemical Products*, WT/DS50/AB/R (Dec. 19, 1997).

283. U.S. OFFICE OF TRADE REPRESENTATIVE, 2004 SPECIAL 301 REPORT (2004), http://www.ustr.gov/assets/Document_Library/Reports_Publications/2004/2004_Special_301/asset_upload_file16_5995.pdf.

284. Sell, *TRIPS*, *supra* note 142, at 492.

flexibility granted by the Doha Declaration to delay implementation until 2016.²⁸⁵

Procurement policies by donors also undercut Attaran's argument. All of the AIDS/HIV drugs on the WHO Prequalification list are produced either in high income countries or in India,²⁸⁶ countries now covered by TRIPS.²⁸⁷ USTR and President's Emergency Plan For HIV/AIDS Relief (PEPFAR) also hinder procurement of generic ARVs by multilateral and official donors.²⁸⁸

The patent thicket effectively covers all sources of ARVs for Africa, forming an effective deterrent to ARV commercialization by generic companies, even in the absence of a formal patent filing in every sub-Saharan country. Even today, treatment with generic ARVs occurs by either complying with TRIPS flexibilities, or by (temporary) forbearance by the United States and PhRMA. Perhaps Attaran and Gillespie-White should say that patent law should not be used to delay access any longer. If so, we are in agreement. But it is historical revisionism of the foulest kind to claim that patents did not matter over the last decade for access to cheap ARV therapy in Africa.

3. Achieving Both Nonrival Access and Optimal Innovation

Next I present my proposals for maximizing public health while optimizing innovation, in light of the urgent problem of access illustrated by this case study. Nonrival access is embraced, whether through voluntary differential pricing or compulsory licensure for low and middle income markets. Dysfunctional pharmaceutical arbitrage from low income markets to high income markets is forbidden, but is not found to be a significant empirical problem. Much more troubling is the threat of counterfeit drugs. All other forms of pharmaceutical arbitrage are encouraged as a means to lower consumer prices. Finally, PEPFAR is critiqued, particularly

285. THORPE, *supra* note 32, at 1.

286. WHO, HIV/AIDS Prequalification, 15th ed., *supra* note 232.

287. See *supra* notes 137-139 and accompanying text. As of January 1, 2005, concessions under TRIPS will be largely limited to the thirty poorest members of the WTO, excluding middle income countries such as Mexico, India, China, and Brazil. Only a few countries have notified the TRIPS Council of their intent to delay full TRIPS implementation until the January 1, 2005 deadline, namely Argentina, Cuba, India, Pakistan, Jordan, Uruguay, Egypt, United Arab Emirates, and Turkey. Indeed, all but three least developed countries in Africa have already adopted pharmaceutical patents, many years prior to the 2016 deadline. THORPE, *supra* note 32, at 1.

288. See *infra* Subsection II.A.3.iv.

for its procurement and supply chain policies which are based upon a fear of pharmaceutical arbitrage.

i. Streamline Compulsory Licensure

Compulsory licenses are difficult to administer under TRIPS. The procedures under Article 31 are time-consuming and expensive. The first two national laws implementing the Cancun General Council Decision require the grant of compulsory licenses in both the importing and exporting countries when patents have been filed.²⁸⁹ The Canadian version requires a good faith effort to negotiate a voluntary license “on reasonable terms and conditions” before applying for a compulsory license, following the general language of TRIPS Article 31(b).²⁹⁰ Requirements like these raise transaction costs and may allow pharmaceutical companies to delay the process for many months or years. This process is wasteful, particularly when duplicated in multiple countries.²⁹¹ Good faith negotiations are not required in two circumstances: public non-commercial use, and “national emergency or other circumstances of extreme urgency.”²⁹² Both exceptions fit the AIDS crisis, and the former is broad enough to encompass nonrival access for global diseases.

Absent the credible threat of compulsory licensure, PhRMA companies have few economic reasons to cooperate with differential pricing, particularly for global diseases outside of the media glare of AIDS. This is not an academic exercise, as annual preventable deaths in low and middle income countries now number in the millions. The process must be simpler and faster for nonrival access to medicines.

At this time, no compulsory license has been issued under the Cancun General Council Decision (i.e., involving import or export), and only a handful of countries have issued any TRIPS-compliant compulsory licenses. Malaysia issued a compulsory license to Cipla Ltd. in February 2004 for importation.²⁹³ As of May 2004, compulsory licenses have also been issued by Cameroon, Mozambique, and the Philippines.²⁹⁴

289. See, e.g., Norwegian Compulsory License Regulation, *supra* note 158, § 108; Canadian Bill C-9, *supra* note 158, § 21.04 (3) (d) (i) (B).

290. Canadian Bill C-9, *supra* note 158, § 21.04 (3) (c) (i).

291. Blustein & Gellman, *supra* note 150.

292. TRIPS Agreement, *supra* note 1, art. 31(b).

293. *Cipla Gets Malaysian Nod for AIDS Drugs: In a Trailblazing Move, Malaysia Has Issued a Compulsory License*, BUS. STANDARD, Feb. 26 2004, at A6.

294. Statement of the Representative of Brazil, speaking on behalf of the GRULAC countries, WHO Executive Board Meeting (May 25, 2004).

For most countries, compulsory licensing is not yet a principal mechanism for introducing generic competition and protecting public health. Many companies are engaged in cross-border sales of generic ARVs, without currently necessitating the Cancun process.²⁹⁵ For example, Brazil, as permitted by its national patent laws, has been producing a number of generic ARVs for both domestic purposes and aid projects to Africa.²⁹⁶ Taking another approach, Thailand simply ruled Bristol-Myers's didanosine patent invalid on public health grounds.²⁹⁷

These various tactics have led to significant generic production of at least first-line ARVs, although much of the market remains in a grey area outside of the compulsory licensing regime established by Doha.²⁹⁸ A more streamlined process for compulsory licensing could bring this market in from the grey. Moreover, it could be particularly helpful with regard to second-line ARVs, where there is still inadequate availability of generics. For example, no FDC containing the second-line drug efavirenz is currently available.²⁹⁹ The medical need for such second-line ARV therapy in the low and middle income countries is significant and growing. The WTO (and the USTR) should make the ARV production and export process more rational, not more difficult. A more streamlined approach to compulsory licensing would provide a well-defined mechanism that might encourage greater participation in these potential markets.

295. Most are not prequalified by the WHO: The amfAR July 2004 report notes the difficulties with twenty-seven companies in eight countries in Asia producing generic ARVs, and only one of them (Cipla Ltd.) operating with WHO Prequalification. AMFAR, *supra* note 188, at 4.

296. Médecins sans Frontières, Brazilian Generic Drugs in South Africa—Background (Jan. 29, 2002), at <http://www.msf.org/countries/page.cfm?articleid=F8557436-9B60-4D00-BC5F0476D8B7A5E1>.

297. Gov't Pharm. Org. v. Bristol-Myers, Thailand Central Intellectual Property Court, Oct. 2002.

298. MSF, UNTANGLING THE WEB, *supra* note 151, at 5-7.

299. *Id.* at 6. This may be because of pressures applied to potential source countries—such as India and Thailand—by the USTR under the Special 301 process. U.S. TRADE REP., 2004 NATIONAL TRADE ESTIMATE REPORT ON FOREIGN TRADE BARRIERS, 217-220 (2004) (India); *id.* at 463-65 (Thailand). The United States and Thailand are negotiating a free trade agreement with TRIPS-plus provisions. See Thai Free Trade Agreement Watch Website, at <http://www.ftawatch.org> (last visited Sept. 2003); see also Marwaan Macan-Markar, *Thailand-U.S.: Freer Trade Weakens Access to HIV/AIDS Drugs*, INTER PRESS SERVICE NEWS AGENCY, May 21, 2004, <http://www.ipsnews.net/africa/interna.asp?idnews=23849>. Fortunately, despite trade-related pressures, Thailand is now preparing a generic FDC as a second-line therapy, containing efavirenz, lopinavir and ritonavir.

ii. Dysfunctional Pharmaceutical Arbitrage of AIDS Drugs

a. Dysfunctional Arbitrage Is Rarely Observed

International arbitrage certainly seems to pose a plausible risk to pharmaceutical companies. The consumer retail price of a kilogram of the active ingredients in Combivir³⁰⁰ is about \$20,000 in the United States, but sells for as little as \$612 in Hyderabad and sub-Saharan Africa.³⁰¹ This price differential is equal to about twenty-five times the average per capita income in the lowest income countries. Neo-classical economic theory predicts that entrepreneurs³⁰² will divert these drugs from the poor and export them to wealthy countries where they will fetch higher prices. Domestic arbitrage occurs within the United States at much lower thresholds.³⁰³ Since the great majority of the world's AIDS patients are in poorer countries, if only a small percentage were diverted, significant volumes of ARVs could flow into high income country markets.³⁰⁴

300. Combivir is GlaxoSmithKline's best selling ARV drug, and the company holds a forty-five percent global market share in HIV/AIDS drugs. See Gautam Naik, *Glaxo's HIV Drugs Come Under Pressure: Competition, Calls for Price Cuts Weakens Company's Dominance of Maturing Market*, WALL ST. J., Sept. 22, 2003, at B3; GLAXOSMITHKLINE PLC, 2003 ANNUAL REPORT, Form 20-F, at 63, <http://www.sec.gov/edgar/searchedgar/companysearch.html> (total of all HIV sales).

301. The active ingredients in Combivir total 450 mg per tablet. A kilogram of active ingredients will create approximately 2222 tablets. The retail price of 2222 tablets of Combivir in the U.S. retail market exceeds \$20,000. See Drugstore.com, at <http://www.drugstore.com> (last visited July 9, 2004).

302. Or smugglers, depending upon your perspective.

303. See *infra* Subsection II.B.2. Jackie Judd, Senior Fellow with the Kaiser Family Foundation, Interview with Gilbert M. Gaul and Mary Pat Flaherty, Interview with *Washington Post* Reporters on Drug Safety Articles (Oct. 24, 2003), http://www.kaisernetwork.org/health_cast/hcast_index.cfm?display=detail&hc=1004 (describing significant arbitrage diversion within the U.S. market taking advantage of relatively modest price differentials).

304. The United States is a likely target market. The EU may not be as vulnerable to diversion because most of its citizens are covered by a third party prescription drug benefit and are not as price sensitive. DG TRADE, *supra* note 42, § 3.3. This conclusion might be true for ultimate consumers, but European intermediaries such as parallel traders could seek arbitrage earnings from this trade. The available evidence suggests that European parallel traders are closely scrutinized and do not knowingly participate in illegal diversions. See, e.g., *Glaxo Group Ltd. v. Dowelhurst Ltd.*, [2004] E.T.M.R. 39 (July 31, 2003), 2003 WL 21729286.

Further, criminal organizations might be attracted to the profits to be found in dysfunctional pharmaceutical arbitrage. The pricing ratios operating in the illegal cocaine market are broadly similar to ARV pricing ratios. The U.S. wholesale price of a kilogram of cocaine ranges from \$13,000 to \$25,000,³⁰⁵ comparable to the U.S. retail value of a kilogram of the active ingredients in Combivir.³⁰⁶ The U.S. retail price of a gram of cocaine is about \$100.³⁰⁷ The retail price of cocaine in Columbia is between three dollars and five dollars per gram,³⁰⁸ yielding a ratio of about 25:1.³⁰⁹ Since ARV arbitrage offers potentially higher profits than cocaine trafficking, one might expect criminal enterprises to enter the ARV business, especially since the risk of apprehension and punishment are so severe for cocaine trafficking, but relatively modest for prescription drug counterfeiting.³¹⁰

Given these facts, it would be striking if dysfunctional ARV arbitrage did not occur. And yet reality appears to depart from the neo-classical economic model, for there is quite limited evidence of dysfunctional arbitrage. It is notable that generic drugs have been produced in India for decades without apparently infiltrating or undermining Western markets.³¹¹ As of April 2002, both the European Commission and the pharmaceutical companies acknowledged that pharmaceutical arbitrage from poor countries into high income countries was "still largely theoretical."³¹² Only

305. U.S. DRUG ENFORCEMENT ADMIN., DRUG TRAFFICKING IN THE UNITED STATES (2001), <http://www.usdoj.gov/dea/pubs/intel/01020/index.html> (2000 data). Retail prices per gram are significantly higher, particularly for smaller quantities.

306. See *supra* note 301.

307. OFFICE OF NATIONAL DRUG CONTROL POLICY, TRENDS IN COCAINE PRICES (1981-2000) (price per gram for purchase of one to ten grams). The UK price for a gram in similar lots is around £50. Independent Drug Monitoring Unit Ltd., UK Drug Prices 2002, <http://www.idmu.co.uk/prices02.htm>.

308. This figure is from a hopelessly anecdotal source, a travel journal of an American using drugs in Columbia. David Ashley, Cocaine in Columbia (June 14, 2000), at <http://www.erowid.org/experiences/exp.php?ID=1796> (last visited Oct. 20, 2004).

309. The numerator is \$100 per gram and the denominator is \$4 per gram.

310. ALLIANCE AGAINST COUNTERFEITING & PIRACY, PROVING THE CONNECTION: LINKS BETWEEN INTELLECTUAL PROPERTY THEFT AND ORGANISED CRIME 7-8 (2002), <http://www.a-cg.com/index2.html> (last visited Oct. 20, 2004).

311. One would expect that over the past twenty years there would have been some significant reported court cases on illegal imports of Indian and other unlicensed generics if the problem were widespread. Andrew Farlow of Oxford finds little evidence of diversion. Andrew Farlow, Costs of Monopoly Pricing Under Patent Protection, Presentation at Columbia University slide 19 (Dec. 4, 2003).

312. DG TRADE, *supra* note 42, at § 3.3.

six months later, GlaxoSmithKline, the patent holder for several important AIDS drugs, brought the sensational charge that 36,000 packages of HIV/AIDS medicines worth approximately US\$18 million were found to have been diverted from West Africa to the EU.³¹³ GlaxoSmithKline sued several participants in the transactions, including a legal parallel trader in pharmaceuticals, Dowelhurst Ltd, for trademark infringement.³¹⁴

The Dowelhurst case unearthed several remarkable facts which undercut the public relations spin that Glaxo had put on the case. First, ninety-nine percent of the packages handled by Dowelhurst were not part of Glaxo's charitable access initiative for Africa, but were ordinary commercial sales to Africa, at prices approximating EU prices.³¹⁵ The Deputy Judge expressed keen displeasure upon finally understanding this point, as he had been led to believe that all of the packages were destined for charitable access programs.³¹⁶ Second, 99% of the packages had been sold within Europe, to addresses in France, and probably never made the trip to Africa.³¹⁷ The alleged diversions occurred in Europe, not in Africa. I say alleged diversions, because the case makes clear that the resale of the drugs was not proscribed by contract.³¹⁸ Third, by placing the packages into commerce within Europe, Glaxo exhausted its IP rights within Europe.³¹⁹ Finally, Glaxo sold the packages without any attempt to label them as ineligible for sale or reimportation into the EU. They were packaged in French, with EMEA license codes and nothing was done to indicate they

313. There were a number of media reports from three continents in October 2002. See, e.g., Sarah Boseley & Rory Carroll, *Profiteers Resell Africa's Cheap Aids Drugs*, THE GUARDIAN, Oct. 4, 2002, at P1; see also DUKES, *supra* note 84, at 50 & n.1.

314. Glaxo Group Ltd v. Dowelhurst Ltd, [2004] E.T.M.R. 39 (July 31, 2003), 2003 WL 21729286.

315. *Id.* ¶ 36.

316. *Id.* ¶ 46. The Deputy Judge imposed over ninety percent of the litigation costs on Glaxo, in part because he felt misled. Glaxo Group Ltd. v. Dowelhurst Ltd., [2003] E.W.H.C. 3060 (High Ct., Ch. Div. 2003), ¶¶ 10, 17.

317. Glaxo Group Ltd. v. Dowelhurst Ltd., [2004] E.T.M.R. 39 (2003), ¶¶ 66-76. Only one percent of the packages had actually been sold to a buyer in Africa, namely the packages involved in the access program.

318. *Id.* ¶ 39.

319. *Id.* ¶¶ 66-76. On appeal, the Court of Appeal upheld the Deputy Judge's rulings on summary judgment, permitting the trial to proceed on the question of compliance with EU rules for pharmaceutical parallel trade. Glaxo Group Ltd. v. Dowelhurst Ltd., [2004] E.W.C.A. Civ. 290 (App. Ct., Civ. Div., 2004). Specifically, the Court of Appeals upheld the exhaustion rule on 100% of the packages rather than just 99%. *Id.* ¶¶ 30-40.

were destined for a charitable access program.³²⁰ Legal European parallel traders were led to believe the drugs had been lawfully placed into European commerce. Indeed, the defendant suggested that Glaxo did so deliberately in order to generate the resulting publicity.³²¹ Within three weeks of the Glaxo diversion story, the European Commission announced plans to issue a regulation to curb such diversions.³²² The 2003 Council Regulation promptly required many modifications to packages and pills destined for essential access programs.³²³

The only other major media report of diversion of essential access drugs was in *Forbes* in April 2004, noting allegations of diversion in Indonesia, Chile, and Lebanon.³²⁴ This story parroted PhRMA's spin on the 2002 Glaxo case in Europe, but failed to mention any of the facts from the Dowelhurst case discussed above. The source of the report in Indonesia was a survey in Jakarta by a respected local health group, which found many donated drugs being either sold on the black market in Jakarta or available in the public health clinics for a price in excess of the statutory maximum.³²⁵ This is a simple case of local corruption, and there is no evidence that the drugs were leaving the immediate market. This situation might be regrettable, but it is not dysfunctional arbitrage; it does not replace commercial markets in the high income countries. Similar local diversions occur in the United States.³²⁶ The reports from Chile and Lebanon are sourced exclusively from local affiliates of PhRMA. Neither report was substantiated; nor do they suggest dysfunctional arbitrage as opposed to local movement of drugs within low or medium income countries. In sum, empirical evidence to date does not indicate a sizable

320. *Glaxo Group Ltd.*, [2004] E.T.M.R. at ¶¶ 46-50.

321. *Id.* ¶¶ 51-53.

322. *EU/WTO – Plan to Curb Illicit Medicines Trade*, EUR. REP., Oct. 26, 2002.

323. At present, the European Union Council Regulation only applies to “tiered price” pharmaceutical exports to seventy-six listed developing and least-developed countries and to “HIV/AIDS, malaria, tuberculosis and related opportunistic diseases” (a limitation which should be amended following Cancun). The European Union defines a “tiered price” pharmaceutical as being offered to the poor for either direct manufacturing cost plus no more than fifteen percent or at less than twenty-five percent of the OECD weighted average ex-factory price. Council Regulation 953/2003 To Avoid Trade Diversion into the European Union of Certain Key Medicines, art. 7, 2003 O.J. (L 135/6) art. 3(a) [hereinafter Council Regulation 953/2003].

324. Richard C. Morais, “*Pssst . . . Wanna Buy Some Augmentin?*,” *FORBES* 2000, Apr. 12, 2004, http://forbes.com/forbes/2004/0412/112_print.html.

325. *Id.*

326. Judd, *supra* note 303.

arbitrage market in ARVs from low income countries into the high income countries.

b. Measures To Hinder Dysfunctional Arbitrage

Possible reasons for the dearth of empirical evidence of dysfunctional pharmaceutical arbitrage include moral and legal sanctions within high income market countries. The impact of these norms is significant in pharmaceutical arbitrage markets. When pharmaceutical arbitrage is unmistakably legal, it flourishes, even at low differential pricing ratios. For example, the EU follows the “community exhaustion” rule, permitting parallel trade in patented and trademarked products within the European Economic Area. Differential pricing ratios of less than 2:1 have been sufficient to create a multi-billion euro legal arbitrage market within the EU,³²⁷ subject to complex rules on repackaging and trademark infringement devised by the European Commission and the European Court of Justice.³²⁸ In the European Union, illegal pharmaceutical arbitrage is rarely observed.³²⁹

Canada provides a contrasting example. Pharmaceutical arbitrage from Canada to the United States operated for years under legal ambiguity. Proponents touted the enhanced consumer access. The pricing differential is less than 2:1, but the arbitrage market now is in the range of \$600 million to \$1.1 billion a year.³³⁰

So the first imperative is to prevent any legal or moral uncertainty concerning dysfunctional arbitrage. At a minimum, diversion of drugs intended for the poor to high income country markets should be clearly illegal. The European Union, for example, promptly moved in this direction following media reports of the Glaxo diversion.³³¹ The United States should follow suit.

The second task is to modify the product to resist substitutability. The

327. PETER WEST & JAMES MAHON, BENEFITS TO PAYERS AND PATIENTS FROM PARALLEL TRADE (York Health Econ. Consortium, Working Paper, 2003) (estimating direct savings of 631 million in 2002 from legal pharmaceutical arbitrage (parallel trade) within the EU) (funded by a grant from European parallel traders). *But see* KANAVOS ET AL., *supra* note 54, at 15-16 (finding meager benefits to consumers from parallel pharmaceutical trade) (funded by an unrestricted grant from Johnson & Johnson).

328. For a recent discussion, see *Boehringer Ingelheim KG v. Swingward Ltd.*, [2004] E.T.M.R. 65 (2004), 2004 WL 343819, at ¶¶ 3-17.

329. *See supra* Subsection II.A.1.

330. *See infra* Subsection II.B.2.

331. *See* Council Regulation 953/2003, *supra* note 323, art. 7.

pharmaceutical manufacturing process could be altered to create multiple versions of any prescription drug, distinguished by radically different colors, shapes, names, sizes, and packaging. Markets must be segmented into commercial and charitable markets, and never the twain shall meet. The Cancun General Council Decision addresses this issue: Exporting countries must clearly identify the products through labeling or marking and through special coloring or shaping.³³² The EU Council Regulation follows this tact.³³³ GlaxoSmithKline and others are complying, altering both the packaging and the color of the product.³³⁴ These steps will eliminate the flow of improperly diverted essential access medicines through legal distribution channels such as parallel traders and distribution companies.

Third, the manufacturer also has the responsibility to deliver the essential medicines to a reputable supply chain located outside of the United States or European Union, in order to avoid domestic exhaustion.

Fourth, consumers in high income markets can be persuaded to resist substitution. Advertising could be directed to commercial market consumers, warning them never to take the red pills with labels in Swahili. This should not be an implicit safety warning that "those pills may not be safe," since Africans will be told exactly the opposite: "The red pills are safe and effective."³³⁵ Advertising should describe diversion as a moral and legal issue: High income patients who take pills intended for impoverished Africans are stealing from the poor.³³⁶ Under the European Union Council Regulation, all covered pharmaceuticals exported from the European Union will bear a special logo identifying the product as destined for the poor.³³⁷ In addition, domestic law within the high income countries should criminalize the practice.

The final front for anti-diversion measures is the borders of the high income countries. Pharmaceutical arbitrage may become dysfunctional

332. *Cancun General Council Decision*, *supra* note 137, ¶ 2(b).

333. *See* Council Regulation 953/2003, *supra* note 323, art. 7, ¶ 10. While the Council Regulation addresses importation in luggage for personal use, similar to the U.S. personal importation rule, it does not address (but probably covers) Internet sales. *Id.* art.10, ¶ 13. Seized product may be used for humanitarian purposes. *Id.* ¶ 14.

334. GLAXOSMITHKLINE PLC, 2003 ANNUAL REPORT, Form 20-F, at 29.

335. Vertical product differentiation based on quality is common in some products (regular versus premium gasoline), but is probably untenable in pharmaceuticals.

336. If the arbitrated drugs were voluntarily sold rather than stolen, then the moral claim weakens.

337. Council Regulation 953/2003, *supra* note 323, 2003 O.J at art. 7. The logo is found in Annex V of the regulation.

only when diversion occurs from low or middle income markets to high income markets. Trade among or between low and middle income markets is not dysfunctional.³³⁸ Thus, the key moment to control dysfunctional arbitrage is at the border of high income countries, not at the border of the exporting country. These protections can be put into place immediately by high income countries and do not depend upon reaching a multilateral agreement at the WTO. Furthermore, the high income countries possess the resources and infrastructure to make interdiction a reality. Indeed, the absence of observed dysfunctional arbitrage may well be a result of the border controls over the entry of drugs that many high income countries enjoy.

c. Low and Middle Income Markets Should Not Bear the Burden of Anti-Diversion Measures

The most striking aspect of these anti-diversion measures is that the responsibility for all of them logically rests upon the manufacturers and high income markets. None of the five measures require expenditure by low or medium income countries. Nevertheless, when PhRMA companies finally agreed to significant differential pricing of ARVs in low income countries, they insisted on strong anti-diversion protections and burden-sharing by the recipient countries.³³⁹ The Cancun General Council Decision requires importing countries to implement reasonable measures to prevent diversion and re-exportation. "Reasonable" measures must be "within their means" and "proportionate to their administrative capacities and the risk of trade diversion."³⁴⁰ Under Cancun, developing and least developed countries inappropriately bear these costs even if global pharmaceutical rents are supra-optimal.³⁴¹

Minor diversions at the clinic or patient level should not be an international enforcement focus. Given the difficulty in setting up a source collection system, it is unlikely that small batches or individual blister packs without packaging will filter back to high income country markets in significant quantities. Minor local diversions are likely to remain in the region and may well be re-sold to other poor patients outside of the

338. See *supra* Subsection II.A.2.iii.

339. Gellman, *supra* note 150.

340. *Cancun General Council Decision*, *supra* note 137, ¶ 4.

341. If global rents are supra-optimal, these costs could be borne by the PhRMA companies without harming innovation. Placing the burden on countries with annual per capita health budgets of \$100 or less is exceedingly unfair.

current distribution system.³⁴² This is not a best-case result, but preventing this arbitrage is not an enforcement priority. The priority should be to address weaknesses in the supply chain that allow large batches to be diverted in a single transaction. The risk may be greatest while the product is still outside of the recipient country.³⁴³

Finally, the heuristic suggests that some level of dysfunctional arbitrage may be tolerable from an innovation point of view. So long as commercial markets are not replaced, the practice will not harm innovation. Modest leakage from commercial markets may reduce rents, but will not harm innovation if overall rents remains supra-optimal.

iii. Counterfeit Drugs

In the debates over essential medicines, care must be taken to distinguish arbitrage from counterfeiting. The term “counterfeit” is often loosely applied in a manner which conflates several categories of products:³⁴⁴ safe and effective drugs that have entered the United States improperly, drugs that are intended to be legitimate but are “sub-standard” in some way, and those that are blatant attempts to defraud consumers by selling fake drugs.³⁴⁵ While all these types of products may raise concerns for consumers,³⁴⁶ I focus my analysis here on the third, narrow category—

342. This appears to be the case in Jakarta. See generally Morais, *supra* note 324.

343. Both conditions were present in the Glaxo case.

344. See *Prescription for Danger: Counterfeit Drug Trade Grows*, CBSNEWS.COM, Aug. 2, 2001, at <http://www.cbsnews.com/stories/2002/01/31/health/main327265.shtml> (“There is no single definition for counterfeit drugs. They may contain dangerous substitutes instead of the real ingredients. Or they may be much like ‘the real thing’—only expired, or not approved for sale in the [United States].”).

345. For discussion of the FDA’s definition, see *Examining the Implications of Drug Importation: Hearing Before the Senate Judiciary Comm.*, 108th Cong. (2004) (statement of William K. Hubbard, U.S. FDA Associate Commissioner for Policy and Planning), <http://www.fda.gov/ola/2004/importeddrugs0714.html>. See also Heather Won Tesoriero, *Fake-Drug Sites Keep a Step Ahead*, WALL ST. J., Aug. 10, 2004, at D4 (describing generic versions which were substituted for brand name drugs still patented in the United States as “counterfeits”); *Options for Safe and Effective Prescription Drug Importation: Hearing Before the Senate Comm. on Commerce, Science, & Transportation*, 108th Cong. (2004) (statement of Mark McClellan, Commissioner of the FDA), http://commerce.senate.gov/hearings/testimony.cfm?id=1105&wit_id=3132 (discussing “unapproved, imported pharmaceuticals” and “unsafe and illegal drugs” along with “ineffective, counterfeit” drugs).

346. See FDA, COUNTERFEIT DRUG TASK FORCE INTERIM REPORT 5-7 (2003), http://www.fda.gov/oc/initiatives/counterfeit/report/interim_report.html (noting that counterfeit drugs may “pose significant public health and safety concerns,” as they “may

products which do not contain the proper active ingredient (for example, where the counterfeit is essentially a placebo product). These “non-functional” counterfeits merit special consideration because of the particular incentives for, and dangers of, their production.³⁴⁷

Empirical evidence suggests that virtually none of the internationally arbitrated drugs arriving in the United States are non-functional counterfeits; their importation is most likely to simply violate technical restrictions on parallel importation, FDA approval or labeling, or other laws.³⁴⁸ Instead, most of the blatantly fake or nonfunctional counterfeit drugs in the United States have domestic origins or domestic networks.³⁴⁹ While the FDA still considers it a relatively rare practice,³⁵⁰ it is nevertheless growing rapidly in the United States and in other high income markets.³⁵¹ In 2000, the estimated value of EU pharmaceutical counterfeiting was

contain only inactive ingredients, incorrect ingredients, improper dosages, sub-potent or super-potent ingredients, or be contaminated.”); EUROPEAN FED’N OF PHARM. INDUS. & ASS’NS, INTERNATIONAL EXHAUSTION OF TRADE MARK RIGHTS 7 (2001) (describing the range of products that may be considered counterfeit by the WHO and the European pharmaceutical trade association and corresponding concerns).

347. My point is not to argue whose definition is “right,” but to demonstrate the analysis which is possible when focusing on this narrower category. In copyright and trademark practice, a “counterfeit” or “pirated” copy is one that was manufactured by an unlicensed source, but such copies are likely to be as functional as the genuine article: A counterfeit Gucci purse might nevertheless be a fully functional and stylish purse. In pharmaceuticals, non-functional counterfeits are, arguably, particularly likely. *See infra* note 353 and accompanying text.

348. *See, e.g.*, Press Release, FDA, Recent FDA/U.S. Customs Import Blitz Exams Continue To Reveal Potentially Dangerous Illegally Imported Drug Shipments (Jan. 27, 2004) (mentioning many categories of unapproved drugs but never indicating that any of them contained no active ingredient).

349. Gilbert M. Gaul & Mary Pat Flaherty, *U.S. Prescription Drug System Under Attack: Multibillion-Dollar Shadow Market Is Growing Stronger*, WASH. POST, Oct. 19, 2003, at A1.

350. FDA, *supra* note 346, at 3.

351. The FDA estimates that pharmaceutical counterfeiting has increased four-fold in the past few years. *See, e.g.*, Mary Pat Flaherty & Gilbert M. Gaul, *Anti-Counterfeit Steps Drugmakers Sought; Legislators’ Goal Is To Halt Illegal Sales*, WASH. POST, Jan. 17, 2004, at A11; Mary Pat Flaherty & Gilbert M. Gaul, *Miami Man Charged with Selling Counterfeit Lipitor*, WASH. POST, Dec. 6, 2003, at E1; Mary Pat Flaherty & Gilbert M. Gaul, *Lax System Allows Criminals To Invade the Supply Chain*, WASH. POST, Oct. 22, 2003, at A1. These articles were part of a series of articles on counterfeit drugs by Mary Pat Flaherty and Gilbert M. Gaul that ran in the *Washington Post* during Fall 2003/Winter 2004. The *Wall Street Journal* has also covered the story. *E.g.*, Anna Wilde Mathews & Heather Won Tesoriero, *Murky Channels: Bogus Medicines Put Spotlight on World of Drug Distributors*, WALL ST. J., Sept. 29, 2003, at A1.

more than 1.5 billion Euros. The United Kingdom-based Anti-Counterfeiting Group estimated in 2003 that 5.8% of pharmaceutical company annual revenue is lost due to counterfeiting.³⁵² If true, counterfeiting is a major threat not only to public health, but also to innovation, far outstripping the limited potential damage from dysfunctional pharmaceutical arbitrage.

Criminal enterprises are currently involved in pharmaceutical counterfeiting.³⁵³ Counterfeiting opportunities may explain the absence of criminal ARV arbitrage. In the illegal, nonprescription drug market, counterfeiting is a difficult practice: If users do not get high, the product will not sell, particularly in sales between repeat players. In prescription drugs, however, the opportunity for counterfeiting is much greater. Patients are often unable to tell whether a counterfeit pill contains the correct active ingredients. It may take weeks or months to notice that therapy is failing, and the cause of failure may not be linked with the counterfeits. Counterfeits may be introduced into legitimate supply chains, diluting therapy while making the counterfeiting more difficult to observe and trace. These information characteristics enable the criminal seller of counterfeit prescription drugs to act as if the transactions were discrete, rather than repeating.

While obtaining arbitrated ARVs might be possible, obtaining them in sufficient quantities would require a procurement team in the field (e.g., sub-Saharan Africa), with multiple diversions against alerted supply chains, followed by repackaging and illegal reverse supply chains back to high income country markets. Counterfeiting is arguably easier than diverting pills from Africa: Drugs labeled and packaged to look like the authentic licensed product (rather than distinguishable pills in packaging labeled for essential medicine programs) can be introduced into high income country supply chains directly. Counterfeiting dispenses with many costs. The per-pill cost to produce a placebo without active ingredients may be far cheaper than covert diversion and procurement, re-coloration, repackaging, and transportation. Finally, it is unlikely that anyone would bother to counterfeit a cheap generic drug. Expensive, patented drugs are the targets of counterfeiters; cheap generics are not.³⁵⁴ A criminal is

352. THE ANTI-COUNTERFEITING GROUP, WHY YOU SHOULD CARE ABOUT COUNTERFEITING 14, http://www.a-cg.com/docs/why_you_should_care.pdf (last modified June 22, 2004).

353. ALLIANCE AGAINST COUNTERFEITING & PIRACY, *supra* note 310, at 2 ("This document provides clear and unambiguous evidence of organised crime controlling, exploiting and benefiting from intellectual property fraud. It is on the increase.").

354. The examples of counterfeits in most media and FDA reports are of expensive

unlikely to counterfeit a pill and sell it as aspirin or Triomune, when it could be sold as Lipitor or Fuzeon. When low-cost generics are available on a nonrival basis, the threat of counterfeits recedes.

Counterfeits, not dysfunctional arbitrage, are the more immanent danger to both public health and PhRMA innovation. Counterfeiting will remain an issue so long as the actual product has a high value relative to the cost of manufacturing a plausible placebo. Taking all R&D cost recovery out of the price system will greatly reduce counterfeiting pressure, but so long as a placebo can be made for a fraction of the value of the actual pill, counterfeiting will remain an issue. The Hubbard-Love R&D Treaty thus would wipe out much of the current incentives to counterfeit, by removing R&D cost recovery from the retail sales price.³⁵⁵ Likewise, nonrival access in low and middle income countries would create the same conditions.

Additional anti-counterfeit measures in high income countries should include a pedigree system of tracing drugs from the manufacturer to the consumer. A pedigree system (or the European system of parallel traders giving notice of intent to trade) would also hinder arbitrage by making product movement transparent to the manufacturer. Most importantly, routine market sampling for counterfeits must be introduced, and sources of counterfeit drugs aggressively traced by law enforcement.³⁵⁶

iv. Implications for PEPFAR

When the Bush Administration established PEPFAR, it chose to largely bypass existing multilateral institutions such as the Global Fund. PEPFAR calls for only 6.3% of the \$15 billion to be placed with the Global Fund, with the remainder devoted to unilateral U.S. efforts.³⁵⁷ This move reflects the Bush Administration's penchant for unilateralism, even in the world of AIDS.

The Global Fund's procurement and supply management guidelines prioritize lowest price, assured quality, and legal compliance.³⁵⁸ Grant

patented drugs such as Lipitor, Epogen, Zyprexa, and Serostim. See Leila Abboud et al., *Fakes in the Medicine Chest; As Drug Counterfeiting Rises, FDA May Propose Changes in Sales, Distribution Network*, WALL ST. J., Sept. 22, 2003, at B1.

355. See *supra* note 213 and accompanying text.

356. Some steps towards an anti-counterfeiting policy are being taken by the FDA. FDA, *supra* note 346, at 18-22.

357. THE PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF: U.S. FIVE-YEAR GLOBAL HIV/AIDS STRATEGY 16 (2004).

358. THE GLOBAL FUND TO FIGHT AIDS, TUBERCULOSIS, & MALARIA, GUIDE TO THE GLOBAL

recipients retain flexibility in how they balance cost, quality, and access in the local context. For example, a recipient country could choose to rely on the WHO prequalification process as the quality mechanism on ARV drugs, or it could choose to impose different standards based on local collective preferences. Similar choices may be made between branded drugs and generics.

One way to understand PEPFAR is that it inverts the Global Fund's ARV procurement priorities and strikes a different balance between access and innovation. PEPFAR gives first priority to legal compliance (and highest quality) rather than lowest effective cost, shunning generics.³⁵⁹ PEPFAR requires approval by a "stringent regulatory authority" before procurement, meaning the NDRAs from the United States, EU, and Japan (the ICH), and possibly Canada.³⁶⁰ Critics attacked these standards as inappropriate barriers to rapid roll-out. On May 16, 2004 PEPFAR announced a new "fast track" FDA certification for generic ARVs, rather than following the WHO pre-certification process.³⁶¹ PEPFAR will impose "all FDA standards for drug safety, efficacy, and quality,"³⁶² even though existing studies have proven the efficacy of ARV treatment with generics.³⁶³

FUND'S POLICIES ON PROCUREMENT AND SUPPLY MANAGEMENT (2004), http://www.theglobalfund.org/pdf/guidelines/pp_guidelines_procurement_supplymanagement_en.pdf.

359. From the beginning, PEPFAR guidance to its field offices prohibited acquisition of cheaper generic FDCs. U.S. GEN. ACCOUNTING OFFICE, GAO-04-784, U.S. AIDS COORDINATOR ADDRESSING SOME KEY CHALLENGES TO EXPANDING TREATMENT, BUT OTHERS REMAIN 37 (2004). A cynic might view "highest quality" as merely a stalking horse for "highest price."

360. *Id.* at 19-37.

361. Tommy G. Thompson & Randall L. Tobias, HHS Proposes Rapid Process of Fixed Dose Combination and Co-Packaged Products: Joint Statement Issued by HHS Secretary Thompson and U.S. Global AIDS Coordinator Tobias (May 16, 2004), <http://www.state.gov/r/pa/prs/ps/2004/32503.htm>; see also Gautam Naik et al., *Viral Strain: In AIDS Fight, Ambitious Goals Meet Hard Realities - Millions of Ill in Poor Nations Fail To Get Drugs as Funds, Medical Systems Fall Short*, WALL ST. J., July 1, 2004, at A1; Sarah Lueck & Michael M. Phillips, *U.S. Awards Grants in AIDS Battle: Disbursement Is First Part of a \$10 Billion Pledge; Generics Issue Is Unresolved*, WALL ST. J., Feb. 24, 2004, at D5 (raising unresolved questions about whether the Office of the U.S. Global AIDS Coordinator will procure generic AIDS drugs at the lowest possible price).

362. *HIV/AIDS Fiscal Year 2005 Budget Request: Hearing of the Subcomm. on Foreign Operations of the Comm. on Appropriations*, 108th Cong. (2004) (statement of Randall L. Tobias, U.S. Global AIDS Coordinator), <http://appropriations.senate.gov/hearings/record.cfm?id=221702>.

363. S. PUJARI ET AL., SAFETY AND LONG-TERM EFFECTIVENESS OF GENERIC FIXED-DOSE FORMULATIONS OF NEVIRAPINE-BASED HAART AMONGST ANTIRETROVIRAL-NAÏVE HIV-INFECTED PATIENTS IN INDIA (2003) (background document for WHO meeting on Fixed Dose

PEPFAR is also creating its own supply chain management system, independent of The Global Fund.³⁶⁴ These efforts are duplicative and will inevitably raise costs and delay treatment.³⁶⁵ Amazingly, the United States funds both programs and remains the largest donor to The Global Fund.³⁶⁶

Three aspects of PEPFAR are worthy of detailed discussion. PEPFAR erects hurdles to procurement of generic ARVs in order to steer additional volume at higher prices to PhRMA companies. It establishes a separate supply chain, permitting the United States to maximize protection against diversion and arbitrage.³⁶⁷ And, finally, it controls quality hoping to delay the onset of resistance. My recommendations to PEPFAR's administrators are as follows:

a. Purchase Generics

The first goal is not legitimate on innovation grounds, since donor programs do not replace existing commercial markets for ARVs. PEPFAR's unilateralism is not needed for innovation, but imposes American notions of the appropriate quality-access balance upon desperately poor countries. Innovation does not require ignoring their collective preferences for low cost treatment under WHO prequalification. PEPFAR appears to operate in the mode of many bilateral aid projects, as a subsidy for domestic exports. The PEPFAR legislation requires fifty-five percent of the U.S. contribution to be used in treatment, and seventy-five percent of that amount (or 41.25% of the total) to be spent on ARVs for fiscal years 2006 through 2008.³⁶⁸ Blocking generic ARVs will funnel \$6.18 billion dollars in

Combinations for HIV/AIDS, Tuberculosis, and Malaria); Christian Laurent et al., *Effectiveness and Safety of a Generic Fixed-Dose Combination of Nevirapine, Stavudine, and Lamivudine in HIV-1-Infected Adults in Cameroon: Open-Label Multicentre Trial*, 364 THE LANCET 29 (2004); Gregory K. Robbins et al., *Comparison of Sequential Three-Drug Regimens as Initial Therapy for HIV-1 Infection*, 349 NEW ENG. J. MED. 2293 (2003); Robert W. Shafer et al., *Comparison of Four-Drug Regimens and Pairs of Sequential Three-Drug Regimens as Initial Therapy for HIV-1 Infection*, 349 NEW ENG. J. MED. 2304 (2003).

364. U.S. AGENCY FOR INT'L DEV., DRAFT STATEMENT OF WORK PUBLISHED FOR COMMENT—SUPPLY CHAIN MANAGEMENT SYSTEM FOR THE PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF (2004).

365. See, e.g., U.S. GEN. ACCOUNTING OFFICE, *supra* note 359; Mark Schoofs, *At Zimbabwe Clinic, Wait Is Long and U.S. Drug Cupboard Is Bare*, WALL ST. J., July 1, 2004, at A8.

366. See THE GLOBAL FUND TO FIGHT AIDS, TUBERCULOSIS & MALARIA, ANNUAL REPORT 2003 (2003).

367. U.S. AGENCY FOR INT'L DEV., *supra* note 364, at 6.

368. United States Leadership Against HIV/AIDS, Tuberculosis, and Malaria Act of 2003,

additional ARV sales to PhRMA companies,³⁶⁹ at a price much higher than generics.³⁷⁰ PEPFAR's stand also diverts those unit sales away from companies such as Cipla, another move advantageous to PhRMA companies.

b. Do Not Create Duplicate Supply Chains

The heuristic tells us that the second goal may be legitimate: Avoid arbitrage from donor programs to high-income markets. But the analysis is not so simplistic. PEPFAR costs are very significant, including both duplicated program expenses and indirect costs from delayed and constrained treatment. PEPFAR is devoting special multi-billion dollar efforts to minimize drug diversion within the recipient countries.³⁷¹ These costs should be balanced against the benefits of averted arbitrage. Most arbitrage is not harmful to innovation, and modest levels of dysfunctional arbitrage may be tolerable, particularly in conditions of supra-optimality.³⁷²

c. Generic FDCs Delay Resistance

PEPFAR's final goal is quality, together with the fear of mismanagement, leading to resistance. This is an important question. Unfortunately, it is likely that another parallel AIDS relief system will hinder the uniform management of the disease. A parallel system further complicates treatment in the field and confuses providers. Resistance must be managed globally. Furthermore, if PEPFAR's primary goal is the avoidance of resistance, it should not erect barriers to Triomune and other FDCs, which are the first-line treatments most effective in preventing the emergence of resistant strains, but which are only available as generic in FDC form. PEPFAR's insistence on U.S. FDA standards will delay the procurement of FDCs. If PEPFAR requires the same standards on anti-malarial FDCs (Fixed-dose Artesunate Combination Therapy or FACT), the most effective treatment for managing malarial resistance will be unnecessarily delayed,³⁷³ despite the fact that WHO has already

Pub. L. No. 108-25, § 403 (2003), 117 Stat. 711.

369. 41.25% of \$15 billion.

370. U.S. GEN. ACCOUNTING OFFICE, *supra* note 359, at 20; Ellen Nakashima & David Brown, *U.S. Rule on AIDS Drugs Criticized: Ban on Using Aid To Buy Foreign Generics Hinders Treatment, Experts Say*, WASH. POST, July 14, 2004, at A12 (quoting Rachel Cohen of MSF).

371. U.S. GEN. ACCOUNTING OFFICE, *supra* note 359, at 44.

372. *See supra* Subsection II.A.3.ii.

373. For a description of the effort to create FDCs for malaria in Africa and Latin

prequalified a FACT.³⁷⁴

B. Pharmaceutical Arbitrage from Canada

Pharmaceutical arbitrage is not just an issue in low and middle income countries; millions of U.S. residents are importing cheaper patented drugs from Canada and elsewhere—the so-called “Boston Tea Party of the 21st Century.”³⁷⁵ Drug imports from Canada should be a textbook example of pharmaceutical arbitrage at work, as PhRMA companies scramble to protect high-priced U.S. markets.

For the larger essential medicines debate, the most salient conclusion from the following analysis is that pharmaceutical arbitrage will flourish, even at relatively low arbitrage ratios below 2:1. Aggressive moves by PhRMA companies and the FDA have not succeeded in stopping the arbitrage. The key factor is the legal ambiguity and moral support for the practice of importing from Canada. Institutions and human behavior matter a great deal when neo-classical economic theory is applied in the real world. In the narrower context of U.S. drug prices, arbitrage from Canada seems unlikely to harm innovation, absent transparent access to PhRMA company data to prove the contrary.

1. The Opportunity for Arbitrage

Patented drug prices in the United States are generally the highest in the world.³⁷⁶ Most other OECD countries have regulatory structures that

America, see Press Release, DNDi, Malaria Patients Enter DNDi Clinical Trials (July 2, 2004), http://www.dndi.org/cms/public_html/insidearticleListing.asp?CategoryId=166&SubCategoryId=167&ArticleId=301&TemplateId=1; and DNDi, Fact Sheet, at http://www.dndi.org/cms/public_html/insidearticleListing.asp?categoryId=164&articleid=304&templateid=2 (last visited July 18, 2004).

374. WORLD HEALTH ORG., ACCESS TO ARTEMISININ-BASED COMBINATION ANTIMALARIAL DRUGS OF ACCEPTABLE QUALITY (2d ed. 2004), <http://mednet3.who.int/prequal/>. The Artemether/Lumefantrine FDC is manufactured by Novartis.

375. Senator Joe Lieberman, Democratic Presidential Debate in Goffstown, New Hampshire (Jan. 22, 2004), <http://www.washingtonpost.com/wp-dyn/articles/A39875-2004Jan22.html>, quoted in Donald L. Bartlett & James B. Steele, *Why We Pay So Much for Drugs*, TIME, Feb. 2, 2004, at 46.

376. United States patented prescription drug prices are the highest of any major market, with the possible exception of Japan. Danzon & Furukawa, *supra* note 173, exhibit 3. Generic drugs, unprotected by patents or exclusive marketing periods, are generally priced competitively in the United States. Comparisons of international drug prices should not conflate these categories. Danzon and Furukawa fault other studies for excluding

significantly limit prices for patented pharmaceuticals.³⁷⁷ Canadian price controls, including the Patented Medicine Prices Review Board³⁷⁸ help to keep Canadian prices significantly lower than U.S. prices for patented drugs.³⁷⁹ This significant differential pricing invites consumer arbitrage.

The first phase of the Canadian-U.S. arbitrage involved individuals purchasing drugs while traveling in Canada for other reasons, such as vacation or business. This arbitrage was usually limited to people who got

generics since they represent significant volumes in the OECD. *Id.* at 4. However, generics must be excluded when calculating rents or the potential for arbitrage in patented drugs. Canadian prices are sixty-four percent of U.S. prices for patented drugs, and somewhat higher for generics, yielding a net differential of six percent. *Id.* exhibit 4; *see also* Letter from William K. Hubbard, Associate Commissioner for Policy & Planning, FDA, to Ram Kamath & Scott McKibbin, Special Advocates for Prescription Drugs, State of Illinois (Nov. 6, 2003) (on file with author) [hereinafter Hubbard Letter] (noting that generics are generally cheaper in the United States compared to Canada). Thus the potential for arbitrage lies in the thirty-six percent differential in patented medications, not the six percent overall figure.

377. *See* ROTHNIE, *supra* note 66, at 491 (providing a general, but dated, discussion of EU pharmaceutical price controls); *see also* DANZON ET AL., *supra* note 186 (noting that pharmaceutical companies delay the launch of new drugs in EU countries with strict price controls to reduce the risk of parallel trade).

378. Since 1988, Canada has regulated patented drug prices through the Patented Medicine Prices Review Board, a quasi-judicial board with can bring proceedings against PhRMA companies that charge excessively high prices. MARIA BARRADOS ET AL., 1998 REPORT OF THE AUDITOR GENERAL OF CANADA ¶ 17.93 (1999), <http://www.oag-bvg.gc.ca>; Robert G. Elgie, Canada's Patented Medicine Prices Review Board: New Approaches, Address to Drug Industry Association Washington Conference on Pharmaceutical Pricing and Reimbursement: What New Variables are at Work? 3-4 (Apr. 16, 1999), <http://pmprrb.com/cmfiles/sp-dia-e14NRL-482003-7465.pdf>. The Board has constrained some patented drug prices in Canada. BARRADOS ET AL., *supra*, ¶17.25. Since the creation of the Board, patented pharmaceutical prices in Canada have increased only one percent per year on average. Elgie, *supra*, at 6. Nevertheless, Canada's system is not strictly a price control or rate-setting system, but a soft reference price system with a quasi-judicial process. BARRADOS ET AL., *supra*, ¶17.50 -17.56; Elgie, *supra*, at 6.

379. Many surveys have documented the price differential between U.S. and Canadian patented pharmaceuticals. *See, e.g.*, RAM KAMATH & SCOTT MCKIBBIN, ILL. OFFICE OF SPECIAL ADVOCATE FOR PRESCRIPTION DRUGS, REPORT ON FEASIBILITY OF EMPLOYEES AND RETIREES SAFELY AND EFFECTIVELY PURCHASING PRESCRIPTION DRUGS FROM CANADIAN PHARMACIES 79 (2003) (thirty-nine percent savings on the drugs that Illinois purchases that could be safely imported from Canada); Danzon & Furukawa, *supra* note 173, exhibit 4 (noting that patented drugs are thirty-six percent cheaper in Canada compared with the United States); *Savings Immense on Canadian Drugs*, WASH. TIMES, Nov. 5, 2003, at A15 (thirty-three percent to eighty percent cheaper for the ten most popular drugs).

sick while in Canada, or who unexpectedly exhausted their U.S. prescriptions while traveling. Marginal transaction costs were negligible for those persons already in Canada.

The second phase was more strategic on the part of consumers. Some U.S. consumers noticed the price differentials when filling prescriptions in Canada. People living close to the border could make short intentional trips to fill lower-cost prescriptions, with a transaction cost of a few dollars and a modest amount of time. Bus trips were subsequently organized for people living at greater distances, specifically to stock up on patented medications. Politicians, particularly those from states near Canada, began to sponsor the trips. The transaction costs for these trips were greater—several hundred dollars and significant time—but for some consumers, the cost savings were greater still. As consumers became more accustomed to mail order pharmacies, repeat customers could avoid the transaction costs of another trip and re-order by mail from Canada. Consumer arbitrage began to erode differential pricing between United States and Canadian drug prices.

These early forms of arbitrage were limited in several ways. Only drugs for outpatient non-emergency use could easily be substituted. The initial buyers were Americans who exhausted their personal drug supplies while traveling in Canada. The high transaction costs of travel to Canada limited the scope and potential expansion of this market. Information costs were also significant. Canadian pharmacies did not significantly advertise in the United States during this phase of the market. Knowledge of the arbitrage opportunity was largely gained by word of mouth or opportune discovery.

i. The Internet Enables More Extensive Arbitrage

The Internet dramatically altered the potential for pharmaceutical arbitrage. The transaction cost of importing a prescription from Canada dropped to a small fraction of the arbitrage savings.³⁸⁰ Many Canadian websites began to compete for the American consumer's attention. These factors multiplied the possible arbitrage market. The potential number of buyers for cross-border arbitrage jumped from several million Americans living near the Canadian border to the entire wired population of the United States. In the last several years, the potential number of buyers expanded again, as U.S.-based companies began to facilitate Internet ordering of pharmaceuticals for unwired consumers, particularly the

380. For a patient with annual out of pocket prescription costs of \$2000, a reasonable amount of search costs can be justified to save thirty percent.

elderly. Health insurers and some government officials began to encourage consumers to acquire cheaper medicines from Canada. The media devoted increasing attention to the phenomenon from 1999, raising awareness amongst consumers that arbitrage was an option. A large and growing portion of the most valuable market for patented pharmaceutical medications is now only a click away from arbitrage.

If this process continues unchallenged, one would expect institutions such as hospitals, nursing homes, and retail pharmacies to begin to source from Canada. Payors such as health plans³⁸¹ and governments³⁸² are now following suit. The State of Illinois recently recommended importing patented drugs from Canada for its employees and retirees. The State of Illinois estimates that \$250 million of its prescription drug costs could be sourced from Canada,³⁸³ with potential savings of \$90.7 million per year.³⁸⁴ Several other states are exploring similar programs.³⁸⁵ These state efforts are being blocked by the FDA.

The current level of arbitrage is already significant in the Canadian market. In 2004, the U.S. retail prescription drug market is an estimated \$207.9 billion.³⁸⁶ In October 2003, an FDA official estimated that three million U.S. prescriptions per year were being filled from Canada,³⁸⁷

381. United States-based PBMs are paying claims today from Canadian pharmacies, supporting the patient's decision to import, KAMATH & MCKIBBIN, *supra* note 379, at 13, as are some large health plans such as UnitedHealth, Thomas M. Burton, *The FDA Begins Cracking Down on Cheaper Drugs from Canada*, WALL ST. J., Mar. 12, 2003, at A1.

382. The State of Illinois is aggressively pursuing a plan to import patented medications from Canada beginning April 1, 2004, if FDA approval is given. KAMATH & MCKIBBIN, *supra* note 379, at 3, 30.

383. *Id.* at 79-81.

384. *Id.* at 19. This figure is based on the assumption that all employees and retirees will participate.

385. See, e.g., Fred Frommer, *Pawlenty Tries To Win FDA over on Drug Plan*, MINNEAPOLIS STAR TRIB., Jan. 16, 2004 (describing the Minnesota Governor's attempt to win FDA approval for a drug importation plan); Tony Leys, *Vilsack Offers Plan on Canadian Drugs*, DES MOINES REG., Jan. 22, 2004 (describing Iowa's plan); Katherine M. Skiba, *Doyle Makes Case for Buying Cheaper Drugs from Canada*, MILWAUKEE JOURNAL SENTINEL, Feb. 24, 2004, at 1A.

386. Heffler et al., *supra* note 80, exhibit 1. This number includes only retail sales of prescription drugs, excluding purchases of prescription drugs by institutions such as hospitals and nursing homes. The all-inclusive number for 2004 is closer to \$250 billion. SAGER & SOCOLAR, *supra* note 11, at 4 & n.25.

387. Transcript of Motion for Preliminary Injunction at 127-28, *United States v. Rx Depot, Inc.*, No. 03-CV-0616-EA (M) (N.D. Okla. Oct. 8-9, 2003) (examination of Thomas McGinnis, Director of Pharmacy Affairs, FDA) (on file with author) [hereinafter Rx Depot Transcript].

yielding an estimated arbitrage market size of \$600 to \$700 million in 2003.³⁸⁸ The IMS Health consulting agency estimates US\$1.1 billion (in U.S. prices) in 2003, an increase of seventy percent over 2002.³⁸⁹ The State of Illinois program alone could add \$250 million to this market, demonstrating the potential for growth. Canadian expenditures on prescribed pharmaceuticals in 2002 were CAN\$14.573 billion,³⁹⁰ thus the arbitrage market is already a significant part of the overall Canadian market.

Unlike ordinarily fleeting opportunities for financial arbitrage, this market is not self-correcting. Canadian prices will not increase much, given government regulation;³⁹¹ normal U.S. prices will not fall unless the PhRMA companies agree to reduce their monopoly price. If the supply of patented drugs in Canada remains sufficient, a permanent arbitrage opportunity results and will persist for as long as the patent remains in force.³⁹² With negligible transaction and information costs, a fungible

388. While the average size of U.S.–Canadian prescriptions is unknown, data from the State of Illinois describe consumer co-pays of at least \$40 per prescription, KAMATH & MCKIBBIN, *supra* note 379, at 5, implying a retail price of \$200 at a twenty percent co-pay. Recent Canadian estimates suggest a market of \$700 to \$800 million per year. Tamsin Carlisle, *Canada Cools to U.S. Drug Flow: Some Online Pharmacies Aren't Filling Big Orders Due to Fears of Shortages*, WALL ST. J., Dec. 26, 2003, at A9; Tony Pugh, *Low-Cost Drug Sales to U.S. Should Stop, Canadian Group Says*, PHILA. INQUIRER, Nov. 16, 2003, at A24. Other recent studies have reached similar estimates for the size of the Canadian arbitrage market. SAGER & SOCOLAR, *supra* note 11, at 4 (\$695 million in 2003, based on IMS data). The largest U.S. retail drug store chain, CVS, estimates that U.S. patients spend \$3 billion a year outside the United States. *US HHS Import Task Force Urged To Curb Overseas Pharma Price Controls*, PHARMA MARKETLETTER, May 17, 2004. By comparison, the domestic U.S. prescription mail order market was \$20.7 billion in 2001. NIHCM, *supra* note 63, at 9.

389. Morais, *supra* note 324 (quoting Paul Saatsoglou, IMS Health).

390. CAN. INST. FOR HEALTH INFO., *supra* note 79, at 66. Precise comparisons with U.S. pharmaceutical sales are difficult. The Canadian figures exclude sales to non-Canadians (including cross-border Internet sales) but include institutional sales (which are excluded from the comparable U.S. statistics).

391. PhRMA companies recently announced small price increases permitted by the Patented Medicine Prices Review Board. Bernard Simon, *Curtailing Medicines from Canada*, N.Y. TIMES, Nov. 11, 2003, at W1. These price increases were targeted against drugs in the U.S. arbitrage market. PhRMA companies are also attempting to limit the supply of drugs provided to Canada to hinder cross-border arbitrage, encouraging shortages and retail price increases. *Id.* Both actions are designed to hinder arbitrage.

392. A permanent arbitrage opportunity is also present in the EU parallel market, given national price controls and various legal restrictions which keep parallel trade to a manageable size. KANAVOS ET AL., *supra* note 54, at 136 (disproving the price convergence

product in abundant supply, and non-responsive pricing, one would expect a large portion of the available U.S. market to source from Canada, limited only by the capacity of the Canadian market to handle the volume.³⁹³

Theoretically, Canadian arbitrage may destroy the differential pricing system which kept U.S. drug prices the highest in the world. Erosion of differential pricing might shift consumer surplus from producers to consumers. American consumers might save many billions of dollars on pharmaceuticals, greatly improving financial access. The other side of the coin is that PhRMA companies may lose the lion's share of their worldwide profits.³⁹⁴ One unasked question is whether this process will result in sub-optimal pharmaceutical rents. Supporters of pharmaceutical companies simply assume that drug innovation will be hindered. So long as total pharmaceutical rents remains supra-optimal, Canadian arbitrage improves consumer welfare without harming innovation.

ii. Regulatory Arbitrage

A process similar to arbitrage also occurs between regulatory systems. Within the United States, if one particular state imposes draconian regulations upon businesses, the business owners may vote with their feet by relocating to a more attractive regulatory environment. If sufficiently important firms relocate, or credibly threaten to do so, then the state may reconsider its stance and ameliorate the harsh regulations.³⁹⁵

hypothesis).

393. A recent CBO issue brief suggests that the net effect on U.S. prices from Canadian arbitrage will be small. CONG. BUDGET OFFICE, WOULD PRESCRIPTION DRUG IMPORTATION REDUCE U.S. DRUG SPENDING? 15 (2004). The CBO assumed that arbitrage supplies would be successfully interdicted by PhRMA companies, capping the arbitrage at ten to fifteen percent of the U.S. market and assumed no competitive price reductions in the United States. *Id.* at 4-6. Even under the CBO's pessimistic assumptions, the ten year savings to U.S. consumers will be \$40 billion. *Id.* at 8. Put another way, PhRMA's displaced sales from legalizing OECD arbitrage will be \$40 billion over ten years.

394. Alan Sager and Deborah Socolar dispute this conclusion, claiming that Canadian arbitrage need not reduce the profits of PhRMA companies, but their conclusion requires that a high percentage of arbitrage purchases actually represent new aggregate demand. SAGER & SOCOLAR, *supra* note 11, at 1 ("We find that if new prescriptions' share of imports is 44.53 percent or more, importing actually increases drug makers' profits."). The question will turn on whether pharmaceutical demand is relatively inelastic. *Id.* at 11-13.

395. The classic work is Charles Tiebout, *A Pure Theory of Local Expenditures*, 64 J. POL. ECON. 416 (1956).

A variation of this process is at work in Canadian arbitrage. In the United States, pharmaceutical companies have been largely successful in blocking the adoption of price controls for their products.³⁹⁶ Other nations, such as Canada, have imposed more restrictive regulatory measures to reduce prices.³⁹⁷ One perspective on this cross-border arbitrage is that some Americans have imported Canada's pricing regulatory system into the United States for outpatient non-emergency pharmaceuticals.³⁹⁸ Regulatory arbitrage is at work between the United States and Canada.

Regulatory arbitrage encourages domestic political reaction. Constituents' demands for pharmaceutical arbitrage has led the Congress to pass the MEDS Act, which legalizes the process once the Secretary of Health and Human Services certifies its safety and cost savings.³⁹⁹ The certification proved to be the Achille's heel, since HHS has refused to issue the certification.⁴⁰⁰ The Medicare Prescription Drug and Modernization Act of 2003, as passed by the House of Representatives, permitted importation from Canada without requiring the Secretary's approval.⁴⁰¹ The Pharmaceutical Market Access Act of 2003, also passed by the House, permitted imports from twenty-five countries with effective NDRAs.⁴⁰² The

396. The industry strongly oppose price controls. See, e.g., Sidney Taurel, *Hands Off My Industry*, WALL ST. J., Nov. 3, 2003, at A14. Taurel is President, Chairman, and CEO of Eli Lilly.

397. Many discussions of Canada's patented pharmaceutical pricing system wrongly assume that it includes mandatory price controls. Canada's Patented Medication Prices Review Board uses soft reference prices and quasi-judicial processes to regulate the ex-factory prices within Canada. The Board also encourages R&D at a minimum level of ten percent of revenues and grants special pricing consideration to breakthrough drugs. BARRADOS ET AL., *supra* note 378, ¶ 17.56; Elgie, *supra* note 378, at 3-4. Thus, Canada's system is one attempt to optimize the appropriation of rents, striking a balance between cost, quality and access, based upon imperfect data.

398. The American Enterprise Institute identifies this as a major weakness of proposals to permit reimportation from Canada. JOHN E. CALFEE, *THE HIGH PRICE OF CHEAP DRUGS* (2003), http://www.aei.org/docLib/20030715_%2315530Calfeegraphics.pdf.

399. Medicine Equity and Drug Safety Act of 2000, Pub. L. 106-387, 114 Stat. 1549A-35 (codified at 21 U.S.C. § 384).

400. Sarah Lueck, *Senate Supports Wider Importing of Canada Drugs*, WALL ST. J., June 23, 2003, at A10.

401. Medicare Prescription Drug and Modernization Act of 2003, H.R. 1, 108th Cong. § 1121 (2003) (passed in the House on June 27, 2003). Another bill in the 108th Congress would have permitted reimportation from the EU as well. Save Our Seniors Act of 2003, H.R. 2769, 108th Cong. § 2 (2003).

402. Pharmaceutical Market Access Act of 2003, H.R. 2427, 108th Cong. (2003).

Senate version of the bill reinstated the certification requirement, effectively gutting Canadian importation under the Bush Administration.⁴⁰³ Most observers would not expect a majority of the U.S. Congress to enact Canada's price regulatory system for the United States; nevertheless, existing federal law (if certified by HHS) would achieve a similar result, in response to consumer exploitation of arbitrage opportunities.⁴⁰⁴

Another example of regulatory arbitrage involves the efforts of U.S. psychologists to obtain prescribing authority, currently denied to them under U.S. law. Some U.S. psychologists direct their patients to Canadian pharmacies, which accept prescriptions written by U.S. psychologists.⁴⁰⁵ This practice will provide empirical evidence of the medical efficacy of prescriptions by U.S. psychologists, a form of self-directed research.

In both cases, regulatory arbitrage focuses debate on the comparative advantages of alternative systems of regulation. This process should be encouraged, as it promotes competitive analysis of regulatory structures and allows market participants to influence the debates with diminished intermediation by interest groups.⁴⁰⁶

403. Medicare Prescription Drug Improvement and Modernization Act of 2003, 21 U.S.C. § 804 (2000); *see also* Sarah Lueck, *Senate Supports Wider Importing of Canada Drugs*, WALL ST. J., June 23, 2003, at A10. A subsequent administration could certify safety and cost-effectiveness and begin importation from Canada without additional Congressional legislation.

404. Henry J. Aaron, *Should Public Policy Seek To Control the Growth of Health Care Expenditures?*, W3 HEALTH AFF. 28-31 (2003) ("The chances that we will adopt the Canadian or French health care systems as a whole are about as good as those that we will join the British Commonwealth or adopt French as a second national language. Even adopting elements of foreign systems is problematic because important aspects of health care financing and delivery are mutually interrelated."). John Calfee of the American Enterprise Institute makes the point that reimportation of pharmaceuticals from Canada is equivalent to importing Canadian price controls. CALFEE, *supra* note 398.

405. Linda Temple, *Who Gets To Prescribe? Psychologists Send Drug Orders to Canada, Spark a Medical Debate*, USA TODAY, Dec. 18, 2003, at D10.

406. Alvarez and Trachtman note that regulatory arbitrage may or may not have positive effects, depending upon the condition of spillovers (negative externalities). Joel P. Trachtman, *Institutional Linkage: Transcending "Trade and . . ."*, 96 AM. J. INT'L L. 77, 84 (2002) (citing Joel P. Trachtman, *Regulatory Competition and Regulatory Jurisdiction*, 3 J. INT'L ECON. L. 331 (2000)). In the present case, pharmaceutical regulatory arbitrage is a response to the existing free rider problem of national drug price regulation. This response may well destabilize the system, and force OECD countries to re-allocate jurisdiction on drug price regulation. Efficient re-allocation of jurisdiction is the primary theme in Trachtman's article. Alan O. Sykes remarks that subjecting domestic regulatory systems to the pressures of global trade "need not be unfortunate. International regulatory competition may well

iii. Virtual Arbitrage

The closely-related concept of virtual arbitrage involves foregoing the actual importation of drugs, but using lower observed prices as an external reference price, whether by government regulation or in contract. The United States employs a virtual arbitrage system in requiring certain discounts for drugs purchased under Medicaid, discounts which reference other “best” prices.⁴⁰⁷ West Virginia recently established a state agency which adopted the Federal Supply Schedule as a soft reference price for drug purchases by the state.⁴⁰⁸ If West Virginia succeeds in lowering drug prices, many other states will likely follow suit.

Virtual arbitrage is preferred in any situation where physical arbitrage is acceptable. Virtual arbitrage is more efficient than physical arbitrage, since resources are not expended in transporting products or in policing against diversion.⁴⁰⁹ Virtual arbitrage is also safer than physical arbitrage since the supply chain is not needlessly articulated through intermediaries. Just as in physical arbitrage, virtual arbitrage from low income markets into high income markets must be blocked if differential pricing is to be supported for essential medicines.⁴¹⁰

Without clear data on the optimality of pharmaceutical rents, no conclusion can be reached as to whether other forms of virtual arbitrage harm innovation. All arbitrage, whether virtual or not, will reduce the surplus captured by the patent holder and shift surplus to the consumer and the arbitrageur; however it begs the question to assume that arbitrage will reduce pharmaceutical rents to a sub-optimal level. One should not assume that the externality is negative. It is possible that West Virginia’s use of an external reference price retains supra-optimal innovation incentives while dramatically lowering the state’s costs and improving access.

drive out foolish and wasteful regulations rather than undermine valuable regulations.” ALAN O. SYKES, *INTERNATIONAL TRADE AND HUMAN RIGHTS: AN ECONOMIC PERSPECTIVE* 17 (Univ. of Chi. John M. Olin Law & Econ., Working Paper No. 188, 2d Series, 2003).

407. 42 U.S.C. § 1396r-8 (2000) (using reference prices to calculate drug prices and drug rebates under Medicaid).

408. WEST VIRGINIA REPORT, *supra* note 175, at 1-7.

409. On the issue of the transaction costs of physical arbitrage, see the comments by Harvey E. Bale, Jr., the Director-General of the International Federation of Pharmaceutical Manufacturers Associations, in Harvey E. Bale, Jr., *The Conflicts Between Parallel Trade and Product Access and Innovation: The Case of Pharmaceuticals*, 1 J. INT’L ECON. L. 637 (1998). These claims are hotly disputed by proponents of parallel trade in pharmaceuticals. See, e.g., WEST & MAHON, *supra* note 327.

410. See *supra* Subsection II.B.2.

2. Responses to Canadian-U.S. Arbitrage

The current efforts to hinder Canadian arbitrage include legal interdiction, increasing transaction and information costs, and selectively controlling drug supplies shipped to Canada.

i. Reducing Arbitrage Demand

a. Legal Interdiction

If transaction costs are raised significantly, at some point the arbitrage transaction will become unrewarding and the market pressure on differential pricing will abate. For consumers, the transactions must be low-risk, particularly with regard to the legality of the transaction, eligibility for reimbursement from third parties, and the counterparty risk of fraud.⁴¹¹

In the first two phases of Canadian arbitrage,⁴¹² the transactions were clearly legal under U.S. and Canadian law. The consumer physically visited a Canadian pharmacy, presented a valid prescription, and received the product. When returning to the United States, most Americans were not searched or questioned about their pharmaceuticals. Even if they had been scrutinized, the federal government allowed them to import small amounts of pharmaceuticals for personal use.⁴¹³

When pharmaceutical arbitrage expanded to mail order and the Internet, Canadian pharmacies and their agents emphasized the personal use exception. Prior to 2003, federal officials did not vigorously challenge this practice. Federal officials did not lack statutory authority to block importation through the mails or package delivery services,⁴¹⁴ but enforcement was uncommon. This lack of enforcement, coupled with the claims of legality under the personal use exception, permitted consumers to believe that the transaction was legal and the risk of government sanction was small.

411. Virtual arbitrage partially escapes this condition since no additional transportation costs are incurred, and safety issues cannot be raised. Other transaction costs may still apply, such as the cost of observing prices and legal costs.

412. See *supra* Subsection II.B.1.

413. FDA, Personal Use Import Policy, <http://www.fda.gov/ora/import/pipinfo.htm> (Apr. 3, 1998).

414. See, e.g., *United States v. Ramsey*, 431 U.S. 606 (1977) (holding that customs officials are permitted to intercept mail for contraband).

Beginning in 2003, the enforcement environment changed.⁴¹⁵ Federal and state officials are currently attacking Internet pharmaceutical arbitrage on multiple fronts. The FDA is aggressively enforcing against U.S. companies involved in the trade.⁴¹⁶ The Customs Department has posted clarifications of the personal use exception to discourage importation.⁴¹⁷ Facilitators such as the Discount Prescription Center in West Virginia have been challenged by state Boards of Pharmacy as engaged in the unlicensed practice of pharmacy.⁴¹⁸ The FDA has sued regional facilitators such as Rx Depot for assisting in the importation of prescription drugs.⁴¹⁹ The FDA and state pharmacy investigators have also purchased prescription drugs in undercover operations.⁴²⁰ Direct interdiction would include enforcement actions against consumers, but arresting grandparents for purchasing Canadian Lipitor is not politically viable.

Canadian arbitrage was born in conditions of legal uncertainty, and continues with a zone of legal protection around the consumers. In addition, the consumers occupy the moral high ground of gaining access to an important drug at market rates. These conditions allowed arbitrage to take root and grow. Citizens and governments which would never consider importing cocaine are buying Canadian drugs over the Internet.

b. Raising Information and Transaction Costs

These enforcement actions, while significant, have not shut down the arbitrage trade. From the perspective of arbitrage, the more significant

415. Thomas M. Burton, *The FDA Begins Cracking Down on Cheaper Drugs from Canada*, WALL ST. J., Mar. 12, 2003, at A1.

416. Lolita C. Baldor, *FDA: Too Costly To Legalize Drug Imports*, LAS VEGAS SUN, Dec. 24, 2003 (describing confiscations of illegal mail-order drugs in New York); Gardiner Harris and Monica Davey, *U.S. Steps Up Effort Against Drug Imports*, N.Y. TIMES, Jan. 24, 2004, at C1; *Recent FDA/U.S. Customs Import Blitz Exams Continue To Reveal Potentially Dangerous Illegally Imported Drug Shipments*, FDA NEWS, Jan. 27, 2004.

417. U.S. Customs & Border Protection, Medication/Drugs, http://www.cbp.gov/xp/cgov/travel/alerts/medication_drugs.xml (last visited Feb. 15, 2004).

418. The West Virginia Circuit Court issued a preliminary injunction forbidding enforcement by the West Virginia State Board of Pharmacy against Discount Prescription Center, concluding that Discount Prescription Center was not a pharmacy and did not violate state law. *Becker v. W. Va. Board of Pharm.*, No. 03-C-1237, slip op. at 11-12 (W. Va. Cir. Ct. Nov. 3, 2003).

419. Rx Depot was shut down by a preliminary injunction granted by District Court Judge Claire V. Eagan on November 6, 2003. *United States v. Rx Depot, Inc.*, 290 F. Supp. 2d 1238 (N.D. Okla. 2003).

420. Rx Depot Transcript, *supra* note 387, at 16-40.

element is pairing enforcement action with widespread publicity to dampen consumer demand. The effect is to increase consumers' transaction costs and deter arbitrage without comprehensive direct interdiction.

Raising information costs may also support product differentiation and discourage substitution.⁴²¹ Pharmaceutical arbitrage occurs when the consumer considers the drugs to be substitutable. These consumers are generally not trained medical specialists, and are unable to evaluate safety or efficacy.⁴²² These consumers are relying on the effectiveness of Health Canada's Therapeutic Product Directorate (TPD), assuming that Canadian drugs are generally as safe as U.S. drugs regulated by the FDA. If the safety or equivalence of drugs from Canadian Internet pharmacies is in doubt, this assumption dissolves and risk-averse consumers are less likely to arbitrage. Supporters of importation take the opposite tack. In October, 2003, the State of Illinois released a major report in support of importing patented drugs from Canada. The report concluded that the Canadian drug supply was actually more secure than that of the United States.⁴²³

A major component of the assault on pharmaceutical arbitrage has been to question safety and equivalence. The FDA has publicly announced its lack of confidence in the Internet drug supply chain. Undercover operations and enforcement activities have highlighted the seizure of mislabeled, counterfeit, or out-of-date drugs.⁴²⁴ Questions have been raised as to whether the drugs are produced and transported under FDA standards of safety.⁴²⁵ Labeling issues, such as the Canadian label for Accutane, have been identified.⁴²⁶ The actual source of arbitrated drugs has also been publicly challenged by FDA officials who muse whether the drugs actually come from Canada at all; perhaps the true source is Thailand or India.⁴²⁷

421. PHILIPS, *supra* note 43, at 187-200.

422. Raising search costs for these consumers should hinder arbitrage and support differential pricing. *See* PHILIPS, *supra* note 43, at 187-200.

423. KAMATH & MCKIBBIN, *supra* note 379, at 11-16 (finding Canadian and U.S. systems equivalent for most aspects, but finding the Canadian system superior in preventing the introduction of counterfeit drugs and incident reporting for internal process errors).

424. *See, e.g., Recent FDA/U.S. Customs Import Blitz Exams Continue To Reveal Potentially Dangerous Illegally Imported Drug Shipments*, FDA NEWS, Jan. 27, 2004.

425. Rx Depot Transcript, *supra* note 387, at 16-158.

426. *Id.* at 77 l. 22 (cross-examination of Melvin Frank Szymanski, consumer safety officer, FDA).

427. *Savings Immense on Canadian Drugs*, WASH. TIMES, Nov. 5, 2003 ("It is not an answer to this problem to say go buy drugs from Canada, which may be coming from Pakistan and

At one level, these accusations prove too much. Counterfeit and unsafe drugs are found in the U.S. market generally and are not confined to the Internet supply chain.⁴²⁸ The FDA does not want to undermine consumer confidence in the U.S. drug supply, but to distinguish the U.S. domestic supply from international Internet sources. Thus, the FDA opposes all international pharmaceutical arbitrage into the United States.

c. The Special Case of Reimportation

Questions about production safety, equivalence, and labeling are reduced for a segment of this market known as reimportation. As a matter of production efficiency, pharmaceutical companies do not build plants in every country of the world. Many are located in the United States, including Puerto Rico, where the U.S. government has long encouraged pharmaceutical research and production through generous tax incentives under Section 936 of the Internal Revenue Code.⁴²⁹ Many drugs produced in these U.S. plants are both sold into the U.S. market as well as exported to nations like Canada. When these drugs make the return trip back to the United States, the process is called reimportation.

Concerns about production safety, equivalence, and labeling of re-imported drugs should be carefully scrutinized. The Canadian government is fully satisfied that these drugs are safe, efficacious, and properly labeled for Canadian use. The FDA worries about errors in shipping and handling from Canada to the consumer,⁴³⁰ but these questions are relevant to all mail order pharmaceuticals and are not endogenous to pharmaceutical arbitrage from Canada. The FDA correctly notes that some Canadian standards differ from FDA rules and forbids reimportation solely on that basis.⁴³¹ But the FDA has not shown that the Canadian drug supply is less safe. Rx Depot was one of the largest facilitators of importing prescription drugs from Canada. The FDA sued Rx Depot, demanding that importation cease. At the Rx Depot trial in October 2003, the FDA was unable to say that Canadian drugs were unsafe or had injured Americans.⁴³²

India and China and all those countries we have health concerns about.”) (quoting Sen. John B. Breaux, D-La); Hubbard Letter, *supra* note 376 (noting one instance of a Canadian website shipping an Indian drug).

428. ‘Lipitor’ Surfaces in Counterfeit Probe, WALL ST. J., Dec. 8, 2003, at B8; Daniel Yee, CDC: Seniors Prescribed Dangerous Drugs, LAS VEGAS SUN, Feb. 9, 2004.

429. Puerto Rico and Possessions Tax Credit, 26 U.S.C. § 936 (2000).

430. Rx Depot Transcript, *supra* note 387, at 29-31.

431. *Id.* at 28, 76-77.

432. *Id.* at 138-41. But see Hubbard Letter, *supra* note 376 (claiming that Internet sales

The most thorough recent analysis of this question concludes that the Canadian drug supply is actually safer on balance than that of the United States. The State of Illinois report recommends a controlled importation system, with extensive safety checks, that results in a high quality drug supply at substantial savings.⁴³³ The EU has many years of experience with parallel trade in pharmaceuticals, without significant safety issues.⁴³⁴

ii. *Reducing Arbitrage Supply*

Each arbitrage transaction lowers the average price. If the supply or demand of product available for arbitrage can be limited, the net financial impact on the producer will be less severe. In European markets, PhRMA companies successfully restrict supply to curb parallel trade.⁴³⁵ Conversely, theory suggests that if supply and demand are unlimited, differential pricing will disappear and a new equilibrium price will prevail in both markets, shifting surplus from the producer to the consumer.

a. *Targeting Canadian Internet Pharmacies*

Pharmaceutical companies have identified Canadian pharmacies that sell to the United States market. These pharmacies have been threatened with a refusal to deal unless the cross-border sales cease.⁴³⁶ This threat not only cuts off the supply for the patented drugs being arbitrated, but it also uses the entire product line as a weapon to enforce differential pricing.

This strategy may not wholly prevent arbitrage. Some doubt the effectiveness and legality of attempts to restrict supply to Canada.⁴³⁷

from Canada will be more open to counterfeiting).

433. KAMATH & MCKIBBIN, *supra* note 379, at 1-5.

434. WEST & MAHON, *supra* note 327.

435. Janice Haigh, *Parallel Trade: What Next?*, 9 PHARMA PRICING & REIMBURSEMENT 295, 297-98 (2004).

436. Tamsin Carlisle, *Pfizer Pressures Canadian Sellers of Drugs to U.S.*, WALL ST. J., Jan. 14, 2004, at A6; Tamsin Carlisle, *Some Online Pharmacies Aren't Filling Big Orders Due to Fears of Shortages*, WALL ST. J., Dec. 26, 2003, at A9; John O'Connor, *Canadians Warn of Rx Shortage*, CHI. SUN-TIMES, Nov. 13, 2003, at 18; Tony Pugh, *Low-Cost Drug Sales to U.S. Should Stop, Canadian Group Says*, PHILA. INQUIRER, Nov. 16, 2003, at 18. Similar restrictions have been employed for many years to hinder parallel trade in Europe, Maskus & Ganslandt, *supra* note 55, at 69-70, with limited effectiveness, WEST & MAHON, *supra* note 327. For the effects of the same tactic on a national level, see DANZON ET AL., *supra* note 186.

437. KAMATH & MCKIBBIN, *supra* note 379, at 22 ("[W]e do not feel the manufacturers['] rhetoric to restrict supply will ever materialize either broadly or consistently, and not at all in the Canadian pharmacies that are hybrid—internet and retail—for two reasons. First,

Members of Congress have asked the United States Attorney General to investigate whether antitrust laws are being violated,⁴³⁸ and traditional Canadian pharmacies are complaining about the impact of drug company restrictions on their domestic operations.⁴³⁹

Canadian pharmacies will still be able to purchase drugs for export but will be forced to purchase through intermediaries. Expenses and marginal cost are likely to rise, but given the significant price differentials between the United States and Canada, arbitrage opportunities will remain. Perverse effects should also be noted. By cutting off direct supplies to exporting pharmacies, the pharmaceutical companies force additional intermediaries into the supply chain, which increases safety and handling problems, increases inefficiencies, and increases the opportunity for spoilage and introduction of counterfeits.⁴⁴⁰ If the concern is truly for patient safety, supply restrictions are a crude and counterproductive tool.

b. Reducing Demand in the United States with a Medicare Prescription Drug Benefit

Pharmaceutical companies also restrict demand in the United States. The current market is mostly non-emergency outpatient drugs. For the Medicare population, these drugs have historically not been covered. In 2003, the U.S. Congress for the first time passed a Medicare prescription drug act, as PhRMA reversed its historic opposition and embraced a market-based third party reimbursement plan in Medicare for outpatient drugs.⁴⁴¹ The new Medicare drug benefit will reduce consumer demand for

limiting supply to Canadian pharmacies may risk their Canadian patent protection; second, as the Minnesota Attorney General and Illinois Attorney General are currently investigating any concerted effort by the pharmaceutical companies to limit supply may violate U.S. antitrust laws.”).

438. Gardiner Harris, *Some in Congress Seek Inquiry over Drug Supply to Canada*, N.Y. TIMES, Nov. 1, 2003, at C2.

439. Carlisle, *supra* note 436; O'Connor, *supra* note 436; Pugh, *supra* note 436.

440. KAMATH & MCKIBBIN, *supra* note 379, at 11-18 (explaining that Canada's drug distribution system does not rely on intermediaries to the same extent as the U.S. system and that increasing reliance on intermediaries increases the risk of counterfeit drugs).

441. Prescription Drug and Medicare Improvement Act of 2003, 42 U.S.C.A. § 1395, 1395-1430 (West 2004). This plan also sows the seeds of future government price controls. Once the federal government becomes the payor, price increases are directly translated into budget issues. Medicare providers such as physicians and hospitals were once paid on a fee-for-service market basis; after years of budgetary issues, Medicare now imposes price controls and rate setting for physician and hospital services. Pharmaceuticals may well

arbitrage in an important population and thus support differential pricing.

3. Implications of Optimality for Canadian-U.S. Arbitrage

Mindlessly blocking pharmaceutical arbitrage between high income countries needlessly sacrifices cost and financial access on the altar of quality. Wonder drugs are useless if they are too expensive to be taken as prescribed. The government's regulatory power should not be used to force consumers into grey markets.

The United States should permit functional pharmaceutical arbitrage, particularly with countries with NDRAs similar to the FDA. Regulatory resources would be devoted to coordination with these governments to ensure the integrity of the supply chain. PhRMA companies bemoan this approach as destructive of long-term research incentives. This is an overly simplistic assessment, for it assumes that pharmaceutical rents would be sub-optimal at undifferentiated high income market prices. If, in fact, Canadian prices are supra-optimal, then Canada is not free riding on American R&D.⁴⁴² Optimal pharmaceutical rents would be achieved at prices between current U.S. and Canadian prices. PhRMA companies would be able to compensate for reduced unit prices by increasing volume.

If Canadian prices currently result in supra-optimal pharmaceutical rents, then extending Canadian prices to the United States will do no harm to innovation. This astonishing possibility would greatly reduce U.S. pharmaceutical access issues without any decline in innovation. Price controls in Canada do not appear to have stifled innovation, as Canadian pharmaceutical R&D is robust and growing.⁴⁴³ If optimality lies somewhere between U.S. and Canadian prices, then U.S. prices could be decreased by some amount without harming innovation. Modest levels of arbitrage and additional price transparency may achieve this result.

Finally, the Canadian experience suggests that PhRMA companies will react to reduced unit prices by stimulating demand for their products. In Canada, despite stable to declining Canadian unit prices for patented pharmaceuticals, national drug expenditures per capita have been rising

follow the same trend line.

442. See Kevin Outterson, *Free Trade Against Free Riders*, 9 PHARMA PRICING & REIMBURSEMENT 254-55 (2004).

443. BARRADOS ET AL., *supra* note 378, ¶ 17.11 (noting that Canadian drug companies agreed to increase R&D to ten percent of sales by the end of 1996). For current data on Canadian pharmaceutical R&D, see Rx&D, at <http://www.canadapharma.org> (the official trade association website).

by 10.2% annually.⁴⁴⁴ Companies increase their profits in declining unit price markets by increasing unit sales⁴⁴⁵ and by developing new drugs.⁴⁴⁶ If profits are stable or increasing, innovation is not harmed. It may be possible to reduce prices, increase access and improve human health simultaneously—the Holy Grail of health policy.

The major barrier to empirically proving any of these three conditions is the lack of independent and reliable data on actual R&D expenditures and profits. Erosion of the high income market internal differential pricing system would put the ball in PhRMA companies' court to demonstrate whether the resulting pharmaceutical rents were globally sub-optimal. For perhaps the first time, these decisions could be made on the basis of actual data rather than imprecise estimates and secret company data.

CONCLUSION

The head of the U.S. global AIDS effort, Ambassador Randall Tobias, is the former CEO of Eli Lilly & Co. When asked about the essential medicines access issue, he claimed it was “yesterday’s issue” and that “from a price point of view, there’s no longer that much difference.”⁴⁴⁷ I beg to differ. Not only are ARVs still not widely available at marginal cost in developing countries, but drug pricing remains unaffordable for other global diseases such as cancer and heart disease in low and middle income markets. The industry prefers to turn off the media spotlight and assume that access problems were adequately addressed at Doha and Cancun, or will be dealt with by PEPFAR. Meanwhile, global public health catastrophes continue to mount. For some of these conditions, we possess effective therapies that can be provided on nonrival terms but are withheld from the poor because of fears of inadequate pharmaceutical rents.

Health care public policy should not be chained to innovation; it must also champion access, whether in Africa or Akron. The theory and praxis of pharmaceutical arbitrage suggests that pharmaceutical access may be greatly improved, at a modest cost, without damaging optimal innovation.

444. CAN. INST. FOR HEALTH INFO., *supra* note 79, fig.18 (reporting, based on 2002 data, the stable to declining Patented Medicine Price Index since the introduction of the Patented Medicine Prices Review Board).

445. *Id.* fig.14 (reporting annual growth rate of per capita prescribed drug expenditures of 10.2% from 1997-2000).

446. *Id.* at 33-43.

447. Robin Wright, *A CEO To Direct the AIDS Battle: Former Eli Lilly Chief Comes out of Retirement*, WASH. POST, Feb. 13, 2004, at A25.

COMMENTARY

Using Reciprocity To Motivate Organ Donations

Mark S. Nadel, J.D.* and Carolina A. Nadel, M.D.†

New drugs and techniques have been steadily increasing the number of patients able to benefit from organ transplants,¹ but the supply of organs has not kept pace with demand. While about 39,000 candidates join waiting lists for organs in the United States every year,² only about 14,000 deaths occur in a manner leaving organs usable for transplants³ and only

* Attorney, Federal Communications Commission. The views expressed in this Commentary are solely the personal views of the author (and his co-author) and are unrelated to his work at the Federal Communications Commission.

† Consultant, medical informatics. The authors wish to thank Elaine Berg, Leonard Bucklin, Arthur Caplan, Mark Cherry, Peter Cohen, Gail Daubert, Francis Delmonico, Douglas Hanto, Julie Hecht, Eric Johnson, Jonathan Kaufelt, Tom Koch, Adam Kolber, Emily Marcus Levine, Andrew Mirsky, Anita Molzahn, Eugene Nadel, Thomas Overcast, Thomas Peters, Jeffrey Prottas, Christopher Robertson, Michael Shapiro, Rochelle Sharpe, Dave Undis, Stephen Wigmore, and Donna Henry Wright for their helpful comments on earlier drafts.

1. See David Hamilton, *Kidney Transplantation: A History*, in KIDNEY TRANSPLANTATION: PRINCIPLES AND PRACTICE 1 (Peter J. Morris ed., 5th ed. 2001); United Network for Organ Sharing (UNOS), Timeline of Key Events in U.S. Transplantation and UNOS History, at <http://www.unos.org/inTheNews/factSheets.asp> (last visited Nov. 04, 2004).

2. Organ Procurement & Transplant Network, Waiting List Additions, at <http://www.optn.org/latestData/rptData.asp> (last visited Nov. 04, 2004).

3. See OFFICE OF INSPECTOR GEN., U.S. DEP'T OF HEALTH & HUMAN SERVS., VARIATION IN ORGAN DONATION AMONG TRANSPLANT CENTERS 1 (2003) (estimating 12,000 to 15,000 potential donors annually for August 2001 to November 2002), <http://oig.hhs.gov/oei/reports/oei-01-02-00210.pdf> [hereinafter 2003 HHS OIG REP.]; Ellen Sheehy et al., *Estimating the Number of Potential Organ Donors in the United States*, 349 NEW ENG. J. MED. 667, 671 (2003) (estimating 13,300-13,800 annually for 1997-1999). This number may increase, however, because thirty-one states have repealed their laws requiring motorcyclists to wear

about half of those organs, approximately three per cadaver, are actually donated.⁴ Lack of permission to use the remaining suitable organs leads to about sixteen deaths daily in the United States⁵ and is why over 85,000 candidates remain on transplant waiting lists.⁶ The majority are waiting for kidneys,⁷ resulting in increased use of dialysis, which is not only burdensome for patients but also costs taxpayers tens of millions of dollars per year.⁸ This Commentary contends that a reciprocity policy could dramatically increase donations and thereby decrease associated deaths. Under the policy, those who committed to donate organs would be granted a preference in the event that they later required a transplant.

helmets, many recently. Matthew L. Wald, *As Risks Make a Helmut More Vital, Fewer Motorcyclists Wear One*, N.Y. TIMES, June 14, 2004, at A13.

4. See 2003 HHS OIG REP., *supra* note 3, at 4 (fifty-one percent donation rate at sample of transplant hospitals, forty-seven percent at other hospitals for August 2001 to November 2002 data); Sheehy, *supra* note 3, at 671 (forty-two percent for 1997-1999 data).

5. See United Network for Organ Sharing, at <http://www.unos.org> (last visited Nov. 4, 2004). This count ignores those removed from the waiting list before they die due to their health and others who are never added for health or financial reasons. See Teri Randall, *Too Few Human Organs for Transplantation, Too Many in Need . . . and the Gap Widens*, 265 JAMA 1223, 1223 (1991); Jonathan D. Sackner-Bernstein & Seth Godin, *Increasing Organ Transplantation—Fairly*, 77 TRANSPLANTATION 157, 157 (2004); see also *Assessing Initiatives to Increase Organ Donations: Hearing Before the House Subcomm. on Oversight & Investigations of the House Comm. on Energy & Commerce*, 108th Cong. 37 (2003) [hereinafter *2003 House Hearing*] (almost sixty percent of those on the waiting list today will die before receiving a transplant). Still, some of these deaths are due to unrelated conditions and many would still die even if all suitable donors donated their organs. See Anthony J. Langone & J. Harold Helderman, *Disparity Between Solid-Organ Supply and Demand*, 349 NEW ENG. J. MED. 704 (2003).

6. See United Network for Organ Sharing, at <http://www.unos.org> (87,271 candidates waiting as of November 4, 2004). Moreover, many patients needing organ transplants are not listed due to financial constraints, see Randall, *supra* note 5, at 1223, or screening standards, see Sackner-Bernstein & Godin, *supra* note 5, at 157 (suggesting that ten times as many listed are excluded); see also DAVID L. KASERMAN & A.H. BARNETT, *THE U.S. ORGAN PROCUREMENT SYSTEM: A PRESCRIPTION FOR REFORM* 26 (2002).

7. See Nat'l Kidney Found., *Transplant Waiting List*, at <http://www.kidney.org/atoz/atozItem.cfm?id=114> (last updated May 17, 2004).

8. See KASERMAN & BARNETT, *supra* note 6, at 64-68 (estimating the social welfare cost of the present system at one billion dollars per year); see also Leonard H. Bucklin, *Woe Unto Those Who Request Consent: Ethical and Legal Considerations in Rejecting a Deceased's Anatomical Gift Because There is No Consent by the Survivors*, 78 N.D. L. REV. 323, 343 (2002) (estimating taxpayer savings of \$500 million over twenty years if transplants replaced dialysis in one thousand cases).

Before discussing the proposal, Part I identifies the two main reasons that so many suitable organs are not donated. Part II then reviews efforts intended to address these issues, including those currently in place in the United States and the two major proposals—presumed consent and financial incentives—now receiving the most attention. Finally, Part III describes the reciprocity proposal advocated here: III.A explains how it works, III.B. describes some of its likely effects, and III.C responds to the major criticisms of the proposal.

I. TWO MAIN PROBLEMS

It has long been argued that organ donation should be motivated solely by altruism, but relying only on such generosity leaves half of the suitable organs in cadavers unused. Sadly, approximately 6000 deaths occur annually due to lack of an organ.⁹ There are two main reasons why suitable organs are not transplanted. First and foremost, most people are not sufficiently motivated to commit to donate. Although more than two-thirds of Americans express a willingness to donate their own organs,¹⁰ less than half of the public has formally committed to do so.¹¹ Many are

9. See *supra* note 5 and accompanying text.

10. See THE GALLUP ORG., THE AMERICAN PUBLIC'S ATTITUDES TOWARD ORGAN DONATION AND TRANSPLANTATION 15 (1993) [hereinafter GALLUP POLL], http://www.transweb.org/reference/articles/gallup_survey/gallup_index.html (reporting that fifty-five percent were willing to donate their organs); Princeton Survey Res. Assocs., Organ Donation Survey (May 1999) [hereinafter 1999 Princeton Survey], <http://www.pollingreport.com/health2.htm#organ> (reporting that forty-two percent of respondents "very likely" to want to donate and twenty-five percent "somewhat likely"). But see Laura A. Siminoff et al., *Public Policy Governing Organ and Tissue Procurement in the United States*, 123 ANNALS INTERNAL MED. 10, 15 (1995) (suggesting that such figures probably reflect bias due to the high social desirability of the answer).

11. See Cindy Bryce et al., Do Incentives Matter? Providing Benefits to Families of Organ Donors (2004) (unpublished manuscript, on file with authors) (finding that, in a survey of residents of Pennsylvania, forty-five percent reported that they had committed to donate on a drivers' license or donor card); GALLUP POLL, *supra* note 10, at 15 (only twenty-eight percent of those surveyed said they had formally committed to donate); 1999 Princeton survey, *supra* note 10 (reporting that forty-two percent had committed to donate on a drivers' license or donor card); see also Laura A. Siminoff, *American Beliefs and Attitudes About Death*, in THE DEFINITION OF DEATH: CONTEMPORARY CONTROVERSIES 183, 189 (Stuart J. Youngner et al. eds., 1999) (finding data on drivers' license requests consistent with 1993 poll); cf. ENVIRONICS RESEARCH GROUP, ORGAN AND TISSUE DONATIONS: PUBLIC AWARENESS, KNOWLEDGE AND ADVERTISING RECALL 11 (2002) [hereinafter 2002 CANADIAN SURVEY] (prepared for Health Canada) (finding that about forty percent of Canadians reported

apathetic or reluctant to contemplate their own mortality.¹² They may prefer to avoid the stress¹³ or even the physical effort required to sign up.¹⁴ Many, at least partially influenced by film and television fiction, fear that their organs will be removed prematurely,¹⁵ i.e., that some in the medical community will view them merely as potential suppliers of organs.¹⁶ Others perceive favoritism in the allocation of organs to celebrities.¹⁷ Still others

having signed a donor card or registering with an organ registry).

12. See GALLUP POLL, *supra* note 10, at 13 (reporting that thirty-six percent of the public found it uncomfortable to think about their own death); see also Lloyd R. Cohen, *Increasing the Supply of Transplant Organs: The Virtues of a Futures Market*, 58 GEO. WASH. L. REV. 4, 10, 13 (1989); Jesse Dukeminier, *Supplying Organs for Transplantation*, 68 MICH. L. REV. 811, 829-30 (1970) (predicting, therefore, disappointing results for organ donations).

13. See Mary Frances Luce, *Choosing to Avoid: Coping with Negatively Emotion-Laden Consumer Decisions*, 24 J. CONSUMER RES. 409 (1998).

14. See William Samuelson & Richard Zeckhauser, *Status Quo Bias in Decision Making*, 1 J. RISK & UNCERTAINTY 7 (1988).

15. See GALLUP POLL, *supra* note 10, at 33 (finding that six percent of respondents feared premature removal of organs); 2002 CANADIAN SURVEY, *supra* note 11, at 45 (nineteen percent feared being declared dead prematurely); Laura A. Siminoff & Kata Chillig, *The Fallacy of the "Gift of Life,"* HASTINGS CENTER REP., Nov.-Dec. 1999, at 34, 36; John F. Neylan, *Transplant Dramas on the Critical List*, TV GUIDE, Apr. 25, 1998, at 50; Mike Holloway, *The Campaign Against Organ Donation (Winter 1996-1997)*, at <http://www.transweb.org/class/holloway.htm>. For a recent example, see *Colo. Dispute over Organ Donor Brain Death, All Things Considered* (Nat'l Pub. Radio broadcast, Oct. 8, 2004) (reporting that a Colorado coroner ruled that organs were removed from a man for donation before medical personnel proved he was brain dead, although donor officials and the medical community counter that standard guidelines for recognizing brain death were used in the case), available at <http://www.npr.org/templates/story/story.php?storyId=4077549>. See also RONALD MUNSON, *RAISING THE DEAD: ORGAN TRANSPLANTS, ETHICS, AND SOCIETY* 173-76 (2002). But see Cohen, *supra* note 12, at 9. This situation also arises outside the United States and Canada. See Catalina Conesa et al., *Psychosocial Profile in Favor of Organ Donation*, 35 TRANSPLANTATION PROC. 1276, 1279-80 (2003).

16. See A. BRUCE BOWDEN & ALAN R. HULL, *CONTROVERSIES IN ORGAN DONATION: A SUMMARY REPORT* 23, 95-96, 98 (1993) (report for the National Kidney Foundation); Deborah L. Seltzer et al., *Are Non-Heart-Beat Cadaver Donors Acceptable to the Public?*, 11 J. CLINICAL ETHICS 347, 354 (2000) (reporting that between eighteen percent and forty-four percent of respondents worry that if doctors know they are donors the doctors may do less to save their lives); Laura A. Siminoff & Mary Beth Mercer, *Public Policy, Public Opinion, and Consent for Organ Donations*, 10 CAMBRIDGE Q. HEALTH & ETHICS 377, 384 (2001) (finding that while only twenty-one percent of whites were concerned that doctors would do less to save their lives if they knew their patient was an organ donor, fifty-two percent of non-whites felt that way).

17. See MUNSON, *supra* note 15, at 36-37; Siminoff & Mercer, *supra* note 16, at 384

prefer to be buried intact for personal or religious reasons¹⁸ (although all major religions permit, if not encourage, life-enhancing donations¹⁹). Some fear making death or funerals more difficult for their families, among other reasons.²⁰

Second, hospitals and doctors also often fail to honor a deceased's directions to donate. In some cases they may lack easy access to a patient's driver's license or organ donor card and a relevant organ donor registry may not exist.²¹ Yet, even when a deceased's wishes are clear, medical personnel routinely seek out surviving family members and defer to their decision, even if it overrides the deceased's directive.²² Healthcare

(reporting that more than sixty-seven percent of donors and seventy-five percent of non-donors believe that rich or famous people have an advantage in obtaining a needed organ); *Liver Allocation and Organ Donation: Public Hearing Before the Dep't of Health & Human Servs* 87 (Dec. 10-12, 1996) [hereinafter *1996 HHS Hearings*] (testimony of Dr. Sollinger on December 10) (noting that following Mickey Mantle's liver transplant, and the controversy over favoritism, relatives were eight times more likely to refuse to donate organs).

18. See GALLUP POLL, *supra* note 10, at 5, 31, 37 (finding that seventeen percent of respondents found it important for a person's body to be intact when buried and five percent believed their religion required this).

19. See ORGAN AND TISSUE DONATION: A REFERENCE GUIDE FOR CLERGY, at V-2 to V-5 (M. Lisa Cooper Hammon & Gloria J. Taylor eds., 4th ed. 2000), <http://www.redcross.org/donate/tissue/relgstmt.html>; John Gillman, *Religious Perspectives on Organ Donation*, CRITICAL CARE NURSING Q., Nov. 1999, at 19. Religious leaders of some denominations oppose donations primarily because they reject the concept that a person can be "brain dead" even though their heart can continue to beat. See, e.g., Debra Nussbaum Cohen, *New Front in Fight over Organ Donation*, JEWISH WEEK, May 14, 2004, at 1.

20. See ENVIRONICS RESEARCH GROUP, ORGAN AND TISSUE DONATIONS: CANADIAN PUBLIC AWARENESS, KNOWLEDGE AND ATTITUDES 25-26 (2001) [hereinafter *2001 CANADIAN SURVEY*] (prepared for Health Canada) (finding twenty-percent and sixteen percent consider it important that donation would make death or funeral arrangements more difficult). In addition, twenty-two percent considered it important that recipients may not deserve an organ transplant, see *id.* at 25, and some may unintentionally prevent donations by executing an advanced directive explicitly demanding the withholding of life support. See Eric F. Galen, *Organ Transplantation at the Millennium: Regulatory Framework, Allocation Prerogatives, and Political Interest*, 9 S. CAL. INTERDISC. L.J. 335, 346-47 (1999); Dave Wendler & Neal Dickert, *The Consent Process for Cadaveric Organ Procurement*, 285 JAMA 329, 333 (2001).

21. Organ donor registries are discussed *infra* notes 62-63 and accompanying text.

22. See Laura A. Siminoff & Renee H. Lawrence, *Knowing Patients Preferences about Organ Donation: Does It Make a Difference?*, 53 J. TRAUMA 754, 756 (2002) (finding that ten percent of families who knew the deceased had chosen to donate still overrode that choice); Wendler & Dickert, *supra* note 20, at 331; GALLUP POLL, *supra* note 10, at 26 (twenty-four

professionals may fear upsetting families,²³ possibly leading to harmful publicity or litigation.²⁴ This is so even though legal penalties are highly unlikely due to statutory immunity provisions,²⁵ and laws may even prohibit overriding the decedent's intent.²⁶ Unfortunately, many believe the often publicized myth that family consent is legally required irrespective of the donor's wishes.²⁷

percent of those who would not donate themselves would also overrule a family member's known preference to donate); see also Kathryn Schroeter & Gloria J. Taylor, *Ethical Considerations in Organ Donation for Critical Care Nurses*, 19 CRITICAL CARE NURSE 60, 64 (1999); Siminoff et al., *supra* note 10, at 16; Donna H. Wright, *Advance Directives and Donor Card Effectiveness Survey Report* (1998) (prepared for UNOS).

23. See Jeffrey M. Prottas, *The Rules for Asking and Answering: The Rule of Law in Organ Donation*, 63 U. DET. L. REV. 183, 186 & n.11 (1985). Also, hospital chaplains seem to define success in dealing with organ donation in terms of whether the family was able to grieve successfully, regardless of whether a decision to donate organs was made. See Ann Mongoven, *Giving in Grief: Perspectives of Hospital Chaplains on Organ Donation*, in *CARING WELL: RELIGION, NARRATIVES AND HEALTH CARE ETHICS* 170, 183-84 (David H. Smith ed., 2000).

24. See Ann C. Klassen & David K. Klassen, *Who Are the Donors in Organ Donation? The Family's Perspective in Mandated Choice*, 125 ANNALS INTERNAL MED. 70, 71-72 (1996); Wendler & Dickert, *supra* note 20, at 332; Wright, *supra* note 22. But see Bucklin, *supra* note 8, at 339-40 (observing that honoring a donor's intent to improve another's life would seem more likely to generate good, rather than bad, publicity); Schroeter & Taylor, *supra* note 22, at 67 (same).

25. See *infra* note 44 and accompanying text. The immunity provision encourages judges to block suits on "summary judgments" without trials, and this shield has not been pierced. See Bucklin, *supra* note 8, at 334-36; Prottas, *supra* note 23, at 190. Still, the medical community greatly fears litigation for overriding the wishes of the deceased's family, *id.* at 190-91; Bucklin, *supra* note 8, at 339 n.145, and recent data support that, see Wright, *supra* note 22, at 8 (reporting survey finding that five of forty-one organ procurement organizations, or OPOs, had been sued for organ removals).

26. Some states have adopted laws to this effect. See, e.g., VA. CODE ANN. § 54.1-2984 (Michie 2004) ("In no case shall the agent refuse or fail to honor the declarant's wishes in relation to anatomical gifts or organ, tissue or eye donation."); see also Bucklin, *supra* note 8, at 339 n.148, 343-48; Daniel Jardine, Comment, *Liability Issues Arising Out of Hospitals' Organ Procurement Organizations: Rejection of Valid Anatomical Gifts: The Truth and Consequences*, 1990 WIS. L. REV. 1655.

27. The myth that family consent is legally required has even been spread by those seeking to increase donations. See Robert E. Sullivan, *The Uniform Anatomical Gift Act*, in *ORGANS AND TISSUE DONATION: ETHICAL, LEGAL, AND POLICY ISSUES* 19, 30-31 (Bethany Spielman ed., 1996) [hereinafter *ORGANS AND TISSUE DONATION*]. For example, a senior organ donation administrator, writing a column titled "Legally Speaking," in the nationally respected publication *RN*, advised nurses in 1987: "[A]ny family has the legal right to say

Where there is no formal directive, families, who often have not discussed the issue with the deceased,²⁸ are forced to make quick decisions in moments of grief and anguish. About half of families asked to donate refused.²⁹ In addition to the reasons noted above, some families are unwilling to delay funerals, and many act out of concern that the deceased “has already suffered enough.”³⁰ Others fear disfiguring the bodies of loved ones.³¹ Many likely view the deceased’s donation directive as a nonbinding charitable impulse.

II. EFFORTS TO ADDRESS THE PROBLEMS

A. *The Current System*

To better understand policies for increasing organ donations, it is useful to consider the current organ allocation system. Those requiring an organ from a cadaveric donor must be listed on the United Network for Organ Sharing (UNOS) waiting list.³² This generally requires that they meet the medical suitability standards of a transplant center and demonstrate their ability to finance the transplant.³³ Medicare generally

‘No’ [to donation] even though the patient was carrying a donor card permitting the retrieval of his organs for use in transplants.” John Kiernan, *If You Have to Ask for an Organ Donation*, RN, Oct. 1987, at 112, 114. Assertions that “family consent is required” have also been made by UNOS, *see* Jardine, *supra* note 26, at 1658 n.17, and by the U.S. General Accounting Office, *see* U.S. GEN. ACCOUNTING OFFICE, ORGAN TRANSPLANTS: INCREASED EFFORT NEEDED TO BOOST SUPPLY AND ENSURE EQUITABLE DISTRIBUTION OF ORGANS 17 (1993) [hereinafter 1993 GAO REPORT]. *See also* Bucklin, *supra* note 8, at 328-34 (discussing legislative efforts to clarify that family consent was not required where an individual had previously stated his or her desire to donate his or her organs). *But see infra* note 49.

28. *See* GALLUP POLL, *supra* note 10, at 19-20 (finding that about fifty percent of respondents had not discussed their preferences regarding donations with their family).

29. *See* Sheehy, *supra* note 3, at 671; Siminoff et al., *supra* note 10, at 14.

30. *See* Siminoff & Chillig, *supra* note 15, at 36; Siminoff & Lawrence, *supra* note 22, at 756.

31. *See* GALLUP POLL, *supra* note 10, at 38 (reporting that nineteen percent of respondents feared disfigurement from a donation).

32. Patients can avoid the UNOS waiting list process by receiving a “directed donation” from a willing and compatible donor. Such directed donations to named individuals are legal throughout the United States. *See* 1987 UAGA, § 6(a); 1993 GAO Report, *supra* note 27, at 63-64; *see also* ROBERT M. VEATCH, TRANSPLANTATION ETHICS 303-04, 388-411 (2000).

33. *See* WILLIAM J. CURRAN ET AL., HEALTH CARE LAW AND ETHICS 768 (1998); MUNSON, *supra* note 15, at 51-52.

covers the bulk of the costs of kidney transplants for its beneficiaries, and Medicaid may cover some transplants for the poor in some states.³⁴ Some patients, however, are forced to pursue loans, grants, or donations,³⁵ and many, like Denzel Washington's character's son in the 2002 film *John Q*, fall short and are thus excluded by this so called "green screen."³⁶

The allocation of organs among those on the UNOS waiting is based, to a large degree, on compatibility.³⁷ For example, for kidneys, a standardized formula awards points to potential recipients based on factors like tissue type, immune status, time on the waiting list, and distance from the donor.³⁸ For most organs, consideration is first given to recipients located within the same donation service area (DSA) as the donor. Nationwide, there are fifty eight DSAs, which are regional combinations of organ procurement organizations (or OPOs) and their transplant center networks. The organ is given to the person in the DSA with the highest UNOS score.³⁹ If there are no suitable recipients in the donor's DSA, the organ is offered next to the candidates in the donor's OPO region (there are eleven OPO regions nationwide), again, based on their scores. If there are no suitable recipients in that region, then the organ is offered nationwide based on those UNOS scores.⁴⁰ This "local first" policy has been

34. See UNOS, Financing a Transplant, <http://www.transplantliving.org/beforethetransplant/finance/funding.aspx#medicare> (last visited Nov. 17, 2004).

35. *Id.*

36. See TOM KOCH, SCARCE GOODS: JUSTICE, FAIRNESS, AND ORGAN TRANSPLANTATION 131-50, 175-98 (2002); Clive O. Callendar et al., *Blacks and Whites and Kidney Transplantation: A Disparity! But Why Won't It Go Away?*, 16 TRANSPLANTATION REV. 163, 171 (2002) (discussing the "green screen").

37. See MUNSON, *supra* note 15, at 47-51. OPOs, however, cannot consider an organ seeker's ethnicity, gender, or religion, and some OPOs also have policies against discrimination against prison inmates. See, e.g., James Sterngold, *Inmate's Transplant Prompts Questions of Costs and Ethics*, N.Y. TIMES, Jan. 31, 2002 at A18.

38. UNOS Organ Distribution Policies § 3.5-3.9 (July 2004) (rules for kidneys, livers, and hearts), <http://www.unos.org/policiesandbylaws/policies.asp?resources=true>; see also Marlies Ahlert et al., *Kidney Allocation in Eurotransplant*, 23 ANALYSE & KRITIK 156 (2001); Johan De Meesters et al., *The New Eurotransplant Kidney Allocation System*, 66 TRANSPLANTATION 1154 (1998).

39. With some exceptions (e.g., special priority is given to O-type recipients, see Galen, *supra* note 20, at 357-58), the organ is offered first to the transplant team of the person on the top of the list from the DSA. Meanwhile, doctors of the patients scoring highest will decline an organ when their patient is not willing and healthy enough to undergo major surgery immediately or insufficiently compatible with the donor.

40. See 1993 GAO REPORT, *supra* note 27, at 18-19.

widely criticized.⁴¹

B. Policies Already in Place To Increase Organ Donations

The problem of enforcing a deceased's express wish to donate was first addressed by the model 1968 Uniform Anatomical Gift Act (UAGA), which all states adopted.⁴² It makes such decisions irrevocable after a donor's death⁴³ and grants immunity from liability to those who act in good faith to honor those wishes.⁴⁴ When, despite this, few OPOs were willing to take organs based solely on a deceased's written directive, a 1987 revision was offered.⁴⁵ Its more explicit language states that: "An anatomical gift that is not revoked by the donor before death is irrevocable and does not require

41. The justification given for the "local first" policy is that organs deteriorate rapidly and that the policy encourages local donors. Livers, however, are generally offered to the medically suitable patient with the most urgent need nationwide, rather than local, subject to travel time constraints. *See infra* notes 139-142 and accompanying text.

42. *See* UNIF. ANATOMICAL GIFT ACT (UAGA) prefatory note (1987), 8A U.L.A. 4 (2003) [hereinafter 1987 UAGA].

43. The 1968 UAGA § 2(e) stated that: "The rights of the donee [OPO] created by the gift are paramount to the rights of others except as provided in Section 7(d)," where 7(d) states that the UAGA is subject to state laws regarding autopsies. UNIF. ANATOMICAL GIFT ACT (UAGA) §§ 2(e), 7(d) (1968), 8A U.L.A. 116, 146 (2003) [hereinafter 1968 UAGA]. In addition, the official comment to the subsection explained "Subsection (e) recognizes and gives legal effect to the right of the individual to dispose of his own body without subsequent veto by others." *Id.* § 2(e) cmt.

44. 1968 UAGA § 7(c), 8A U.L.A. 146 (2003). That provision was slightly clarified in the 1987 UAGA § 11(c), 8A U.L.A. 64 (2003), and now reads: "A hospital, physician, . . . or other person, who acts in accordance with this Act . . . or attempts in good faith to do so is not liable for that act in a civil action or criminal proceeding." And, absent a factual dispute about whether consent was given, such immunity has been upheld by courts on summary judgment. *See, e.g.,* Lyon v. U.S., 843 F. Supp. 531 (D. Minn. 1994); Nicoletta v. Rochester Eye & Human Parts Bank, 529 N.Y.S.2d 928 (N.Y. Sup. Ct., 1978); Carey v. New England Organ Bank, 17 Mass. L. Rptr. 582, 2004 WL 875623, at *9 (Mass. Super. 2004).

45. 1987 UAGA § 2(h) was an attempt to respond to the medical community's failure to take advantage of the 1968 UAGA. *See* Thomas D. Overcast et al., *Problems in the Identification of Potential Organ Donors: Misconceptions and Fallacies Associated with Donor Cards*, 251 JAMA 1559, 1561-62 (1984) ("The evidence suggests . . . that . . . family consent is still required in [all states except California, Colorado, Florida, and Wyoming]. . . . In the majority of instances, this policy is based on fear of prosecution. The medical community does not think that the provisions of the UAGA provide sufficient protection."); *see also supra* note 25; *infra* note 49.

the consent or concurrence of any person after the donor's death."⁴⁶ Although only thirty-four states have adopted that revision,⁴⁷ the effort to pass legislation which can overcome the resistance of transplant professionals⁴⁸ is now gaining greater attention under an initiative entitled "donor designation."⁴⁹

The U.S. Department of Health & Human Services (HHS) is focusing its efforts on helping hospitals to improve their ability to convince the families of dead or dying patients to donate. HHS created a "Gift of Life Initiative," which includes an "Organ Donation Breakthrough Collaborative" to identify and promote the best practices for requesting donations from family members.⁵⁰ It builds on experiences, particularly

46. 1987 UAGA § 2(h).

47. See Advisory Committee on Organ Transplantation (ACOT), U.S. Dep't of Health & Human Servs., Recommendations to the Secretary app.6 (2003), <http://organdonor.gov/acotapp6.html>. States have opposed the revised UAGA for various reasons. See Ann McIntosh, Comment, *Regulating the Gift of Life*, 65 WASH. L. REV. 171, 176 (1990).

48. See *supra* text accompanying notes 23-26.

49. The AOPO, UNOS, and HHS ACOT have all endorsed implementing the 1987 UAGA provision, i.e., the "donor designation" policy, in all states. See U.S. Dep't of Health & Human Servs. Advisory Comm. on Organ Transplantation (ACOT), Summary Notes from Meeting, Wash. DC 4, 9-10, 11-13 (May 22-23, 2003), available at <http://www.organdonor.gov/acot5-03.html> [hereinafter ACOT May 2003 Notes]. This has been the rule in four states since 1985. See Overcast et al., *supra* note 45, at 1562; see also David A. Peters, *A Unified Approach to Organ Donor Recruitment, Organ Procurement, and Distribution*, 3 J.L. & HEALTH 157, 185-87 (1988) (noting that families "should be considerably informed that retrieval procedures will be implemented in deference to their loved one's prior decision."). OPOs that fail to abide by donor directives could even be penalized with a temporary suspension of federal funds or of accreditation by the Joint Commission on Accreditation of Healthcare Organizations (JCAHO). This option was pointed out to the authors by Peter Cohen. See E-mail from Peter Cohen to the author (Nov. 21, 2003) (on file with authors). Moreover, this would appear practical to enforce where there was a disgruntled family member, angry that the rest of the family had overridden the deceased's wishes.

50. See U.S. DEP'T OF HEALTH & HUMAN SERVS. & SERVS. ADMIN., OFFICE OF SPECIAL PROGRAMS, DIV. OF TRANSPLANT, THE ORGAN DONOR BREAKTHROUGH COLLABORATIVE: BEST PRACTICES FINAL REPORT (2003), <http://www.organdonor.gov/bestpractice.htm> (identifying seven overarching principles and fifteen specific practices for increasing organ donations); U.S. Dep't of Health & Human Servs. Advisory Comm. on Organ Transplantation (ACOT), Summary Notes from Meeting, D.C., May 6-7, 2004, <http://www.organdonor.gov/acot5-04.htm> [hereinafter ACOT May 2004 Notes]; 2003 House Hearing, *supra* note 5, at 36, 41-42 (testimony of Michelle Snyder, Dir. Office of Special Programs, HRSA, discussing HHS's Gift of Life Initiative); Laura Siminoff et al., *Factors Influencing Families' Consent for Donation*

with the Spanish Model—whereby a specially trained team, separate from the medical/transplant teams, is responsible for increasing organ donations⁵¹—and also with the “Donor Action” diagnostic review protocols.⁵² It seeks to raise the average donation rate to the seventy-five percent level now achieved by the most successful hospitals.⁵³

Efforts to increase donor consent rates have also long included attempts to educate the public, and over the last decade public service announcements promoting organ donation in the United States have used about half a billion dollars in free television time.⁵⁴ In addition, special organ donation programs have been initiated by the American Medical Association (AMA), HHS, the American Society of Transplant Surgeons (ASTS), and UNOS.⁵⁵ HHS, for example, is promoting major public

of Solid Organs for Transplantation, 286 JAMA 71 (2001).

51. See Blanca Miranda et al., *Optimizing Cadaveric Organ Procurement: The Catalan and Spanish Experience*, 3 AM. J. TRANSPLANTATION 1189 (2003). But see George E. Chang et al., *Expanding the Donor Pool: Can the Spanish Model Work in the United States*, 3 AM. J. TRANSPLANTATION 1259 (2003) (suggesting that the protocol may primarily represent a higher utilization of marginal donors).

52. See Leo Roels & Celia Wight, *Non-Exploited Potential for Organ Donation: Aggregated Data from the Donor Action Database*, 2 AM. J. TRANSPLANTATION 375 (Apr. supp. 2002); Leo Roels & Celia Wight, *Donor Action: An International Initiative to Alleviate Organ Shortage*, 11 PROGRESS TRANSPLANTATION 90 (2001).

53. See Press Release, Sec’y’s Donation Initiative, U.S. Dep’t of Health & Human Servs., HHS Expands Organ and Tissue Donation Initiative, Promotes Living Donation Safety and Awareness (June 3, 2003), <http://www.organdonor.gov/secgrndbrk.htm>.

54. Telephone conversation with Melissa Devanny, Assistant Director, Coalition on Donation (Oct. 1, 2003); see also Thomas J. Cossé & Terry M. Weisenberger, *Words Versus Actions About Organ Donation: A Four-Year Tracking Study of Attitudes and Self-Reported Behavior*, 50 J. BUS. RES. 297 (2000).

55. See Am. Med. Ass’n, D-370.992, at http://www.ama-assn.org/apps/pf_new/pf_online?f_n=browse&doc=policyfiles/DIR/D-370.992.HTM; Tommy Thompson, U.S. Dep’t of Health & Human Servs., Organ Donation, <http://www.organdonor.gov/workplace.htm> (describing the Workplace Partnership for Life); First Family Pledge, What’s New at First Family Pledge, at <http://familypledge.org/WhatsNew.asp>. Other successful efforts include the National Minority Organ Tissue Transplant Education Program (MOTTEP). See LISA GILMORE ET AL., U.S. DEP’T OF HEALTH & HUMAN SERVS., STATE STRATEGIES FOR ORGAN AND TISSUE DONATION: A RESOURCE GUIDE FOR PUBLIC OFFICIALS 53-77 (2001); Clive O. Callender & Patrice V. Miles, *Obstacles to Organ Donation in Ethnic Minorities*, 5 PEDIATRIC TRANSPLANTATION 383 (2001) (addressing the particular difficulty recruiting minorities); see also Amitai Etzioni, *Organ Donation: A Communitarian Approach*, 13 KENNEDY INST. ETHICS J. 1, 5-7 (2003).

education initiatives.⁵⁶ Unfortunately, evidence from the substantial national educational campaigns in the United States, Canada, Sweden, the Netherlands, Australia, and England indicates that none have significantly increased organ donation rates.⁵⁷ Then again, it could be that the primary impact of such programs is offsetting the negative impact of the chilling, fictional media broadcasts noted above.⁵⁸

Another related set of efforts include “mandated choice,” requiring individuals to decide in advance whether they will donate;⁵⁹ “required request” laws, which command hospitals to ask patients or their families

56. The HHS organ donation website, Organ Donation, at <http://organdonor.gov>, is being redesigned and will feature HHS’s educational initiatives prominently, according to ACOT’s recommendation #11. See 2003 House Hearing, *supra* note 5, at 36-37, 80.

57. See 2002 CANADIAN SURVEY, *supra* note 11 (finding little change in Canadian attitudes after a 2001 to 2002 media campaign); Cossé & Weisenberger, *supra* note 54; Mehmet C. Oz et al., *How To Improve Organ Donation: Results of the ISHLT/FACT Poll*, 22 J. HEART & LUNG TRANSPLANT 389, 393 (2003) (noting the 1998 Netherlands campaign); see also B. Cuzin & J. M. Dubernard, *The Media and Organ Shortage*, in ORGAN SHORTAGE: THE SOLUTIONS 287, 288-89 (J.L. Touraine et al. eds., 1994) (reporting on a campaign in Australia); BILL NEW ET AL., KINGS FUND INST., A QUESTION OF GIVE AND TAKE: IMPROVING THE SUPPLY OF DONOR ORGANS FOR TRANSPLANTATION (1994); Margaret S. Verble & Judy Worth, *The Case Against More Public Education to Promote Organ Donation*, 6 J. TRANSPLANT COORDINATION 200 (1996). But see Ian Kennedy et al., *The Case for “Presumed Consent” in Organ Donation*, 351 THE LANCET 1650, 1650 & 1652 nn.6-8 (1998) (“Supply can be increased by energetic educational campaigns . . .”); Cuzin & Dubernard, *supra*, at 289-90 (reporting on some success in Saudi Arabia). There is some evidence that direct personal contacts, such as speaking at drivers’ education classes and church functions, is most effective. See Cuzin & Dubernard, *supra*, at 292-93; Ellen G. Lanser, *Sharing the Gift of Life: Your Role in Raising Organ Donor Awareness*, HEALTHCARE EXEC., Nov./Dec. 2001, at 20, 23-25.

58. See *supra* note 15 and accompanying text.

59. See Klassen & Klassen, *supra* note 24, at 72 (critiquing mandated choice and noting that when the Virginia DMV began using mandatory choice forty-five percent of subjects registered as nondonors and twenty-four percent as undecided); Siminoff & Mercer, *supra* note 16, at 380 (noting that when Texas attempted mandated choice in the early 1990s there was an eighty percent refusal rate and the Texas legislature repealed the law); Aaron Spital, *Mandated Choice for Organ Donation: Time To Give It a Try*, 125 ANNALS INTERNAL MED. 66 (1996); Monique C. Gorsline & Rachelle L.K. Johnson, Note, *The United States System of Organ Donation, the International Solution, and the Cadaver Organ Donor Act: “And the Winner is . . .”*, 20 J. CORP. L. 5, 38-48 (1994). At least five states ask their residents whether they are willing to be organ donors. See GILMORE ET AL., *supra* note 55, at 43; Andrew C. MacDonald, *Organ Donation: The Time Has Come to Refocus the Ethical Spotlight*, 8 STAN. L. & POL’Y REV. 177, 183 (1997) (interpreting Colorado’s Department of Motor Vehicles (DMV) preference request to be an example of mandated choice).

about donating;⁶⁰ as well as driver's license applications that invite drivers to check off a box to donate.⁶¹ At least thirty states have created donor registries,⁶² which facilitate hospital access to patient choices, and Congressional bills have proposed a national registry.⁶³

The introduction of live donors for kidneys, as well as for liver or lung parts, has reduced the organ shortage.⁶⁴ In addition, transplants of organs

60. Such laws were federalized in 1986. Pub. L. No. 99-509, § 9318(a), 100 Stat. 2009 (1986) (codified as amended at 42 U.S.C. § 1320b-8(a)(1)(A) (2000)). The statute conditions eligibility for Medicare and Medicaid hospitals on employment of some form of required request. Federal regulations require hospitals to notify an OPO of all imminent or recent deaths and ensure that families of potential donors are asked to donate by a specially trained representative. 42 C.F.R. § 482.45(a)(3) (2003). In addition, by 1992, almost all states had enacted some form of 1987 UAGA § 5, which requires that potential donors be asked about donation. Fred H. Cate, *Human Organ Transplantation: The Role of Law*, 20 J. CORP. L. 69, 73 & n.48 (1994).

61. See Overcast et al., *supra* note 45; Editorial, *The Virginia DMV's Noble New Cause*, ROANOKE TIMES & WORLD NEWS, May 20, 1999, at A20 (noting a jump from 16,000 to 64,000 registering to donate organs in March 1999 after the Virginia DMV began orally asking customers to do so).

62. See Ass'n for Organ Procurement Orgs., Donor Registry Information by State (Dec. 2003) (on file with authors); see also UNOS, Donor Designation (First Person Consent) Status by State (May 2004), at <http://www.unos.org/Resources/factsheets.asp?fs=6>. See generally LEWIN GROUP, GUIDELINES FOR DONOR REGISTRY DEVELOPMENT CONFERENCE, FINAL REPORT app. B (2d rev. 2002) (prepared for the Div. Transplantation, U.S. Dep't of Health & Human Servs.), <http://www.organdonor.gov/nfdrguidelines.html>; OFFICE OF INSPECTOR GEN., U.S. DEP'T OF HEALTH & HUMAN SERVS., ORGAN DONOR REGISTRIES: A USEFUL, BUT LIMITED TOOL (2002), <http://oig.hhs.gov/oei/reports/oei-01-01-00350.pdf>.

63. See, e.g., Motor Donor Act, S.788, 107th Cong. (2001); Motor Donor Act, H.R. 2645, 107th Cong. (2001); Donate Act, S.1062, 107th Cong. (2001); Organ Donor Enhancement Act H.R. 955, 107th Cong. (2001). There are also registries like Living Bank, at <http://www.livingbank.org>. Belgium has had a national registry since 1986, and Austria and Sweden established them in 1996. See Paul Michielsen, *Informed or Presumed Consent Legislative Models*, in ORGAN AND TISSUE DONATION FOR TRANSPLANTATION 344, 345 (Jeremy Chapman et al. eds., 1997) [hereinafter CHAPMAN].

64. In fact, in 2001 and 2002 there were more live kidney donors than cadaver donors, although more organs came from the latter. See Alvin E. Roth et al., *Kidney Exchange*, 119 Q. J. ECON. 457, 458 (2004); see also Denise Grady, *Transplant Frontiers: A Special Report; Healthy Give Organs to Dying Raising Issue of Risk and Ethics*, N.Y. TIMES, June 24, 2001, § 1, at 1. Moreover, kidneys from live donors appear to produce significantly better results. See Sundaram Hariharan et al., *Improved Graft Survival After Renal Transplantation in the United States*, 342 NEW ENG. J. MED. 605 (2000). This is leading to increased focus on "paired exchanges." See Francis L. Delmonico, *Exchanging Kidneys—Advances in Living-Donor Transplantation*, 350 NEW ENG. J. MED. 1812 (2004); Roth, *supra*. Yet, the better results for

that were previously considered unusable are now possible because of new drugs, technologies, and methods.⁶⁵ Research continues on more controversial options like using animal organs, known as xenotransplantation, and cloning.⁶⁶

Despite all of these current efforts, however, half of the usable organs in cadavers continue to go undonated, leading to thousands of unnecessary deaths annually.⁶⁷ While some current initiatives—such as the HHS Breakthrough Collaborative—are certainly promising,⁶⁸ it seems worthwhile to also consider other options.⁶⁹

recipients of organs from live donors may actually be due to the better health of such recipients. They tend to be younger and have spent less time on waiting lists than those receiving cadaveric donations. See Alex Tabarrok, *Life-Saving Incentives: Consequences, Costs and Solutions to the Organ Shortage*, LIBR. ECON. & LIBERTY, Apr. 5, 2004, at n.3, at <http://www.econlib.org/library/Columns/y2004/Tabarrokorgans.html>. Also researchers have estimated that one in three liver donors suffers a medical complication and half of those are serious. See Laura Meckler, *Living Organ Donors Often Oblivious to Risks They Run*, L.A. TIMES, Aug. 10, 2003, at A1. See generally David Steinberg, *An "Opting In" Paradigm for Kidney Transplantation*, AM. J. BIOETHICS, Dec. 2004, at 1, 1-5 (discussing the drawbacks of live donation).

65. See Robert J. Stratta, *Expanded Criteria Donors in Kidney Transplantation: A Treadmill or Bandwagon Effect?*, MEDSCAPE TRANSPLANTATION, Sept. 16, 2004, <http://www.medscape.com/viewarticle/488926>. But see B. K. Rayburn et al., *Are Efforts at Expanding the Donor Pool Misdirected?*, 17 J. HEART & LUNG TRANSPLANTATION 998 (1998).

66. See JOINT COMM. ON ACCREDITATION OF HEALTHCARE ORGS., HEALTH CARE AT THE CROSSROADS: STRATEGIES FOR NARROWING THE ORGAN DONATION GAP AND PROTECTING PATIENTS 35 (2004) [hereinafter 2004 JCAHO REPORT], available at <http://www.jcaho.org/news+room/press+kits/organ+donation+white+paper/organ+donation+white+paper.pdf>; MUNSON, *supra* note 15, at 220-64. The use of organ donations from condemned prison inmates, however, appears to be beyond the pale. See Phyllis Coleman, *"Brother, Can You Spare A Liver?" Five Ways To Increase Organ Donation*, 31 VAL. U. L. REV. 1, 26-34 (1996). But see Craig S. Smith, *On Death Row, China's Source of Transplants*, N.Y. TIMES, Oct. 18, 2001, at A1.

67. See *supra* note 4-5 and accompanying text.

68. See ACOT May 2004 Notes, *supra* note 50, at 24-26 (reporting promising preliminary results from the Collaborative Breakthrough).

69. One option that is beyond the scope of this analysis is the one voiced by Tom Koch, among others, that organ transplantation should be suspended in the nation until the fundamental social and geographical inequalities of the current system are remedied. See KOCH, *supra* note 36.

C. Other Proposed Policies: Presumed Consent & Financial Incentives

At least nineteen nations have legislated a policy of “presumed consent.”⁷⁰ Under that policy, an individual is treated as having consented to donate organs absent express instructions to the contrary.⁷¹ It appears to be the preferred approach of many, if not most, transplant professionals,⁷² and the HHS Advisory Committee on Organ Transplantation (ACOT) is considering whether to recommend the policy to HHS.⁷³ Not only have data indicated that a presumed consent default could save lives by increasing actual donations by sixteen percent or more,⁷⁴ but the policy also relieves many grieving relatives of the burden of deciding whether or not to donate a loved one’s organs.

On the other hand, many medical professionals are concerned that strictly enforcing presumed consent tramples the autonomy, if not civil liberties, of individuals who prefer not to donate but fail to formally opt

70. See Ronald W. Gimbel et al., *Presumed Consent and Other Predictors of Cadaveric Organ Donation in Europe*, 13 *PROGRESS TRANSPLANTATION* 17, 19 (2003) (listing eighteen nations). Singapore is another example. See *infra* note 97.

71. See Carl Cohen, *The Case for Presumed Consent To Transplant Human Organs After Death*, 24 *TRANSPLANT PROC.* 2168 (1992); Jesse Dukeminier & David Sanders, *Organ Transplantation: Proposal for Routine Salvaging of Cadaver Organs*, 279 *NEW ENG. J. MED.* 413 (1968); William N. Gerson, Note, *Refining the Law of Organ Donation: Lessons from the French Law of Presumed Consent*, 19 *N.Y.U. J. INT’L L. & POL.* 1013 (1987); Kennedy et al., *supra* note 57; Oz et al., *supra* note 57; *infra* note 78. A version whereby consent is presumed absent informed rejection by donor or family has also been suggested. Arthur J. Matas et al., *A Proposal for Cadaver Organ Procurement: Routine Removal with Right of Informed Refusal*, 10 *J. HEALTH POL. POL’Y & L.* 231 (1985). As of 1995, twenty-one states applied presumed consent to corneal tissue. See Cate, *supra* note 60, at 84 & nn.115-16. See generally The Presumed Consent Foundation, at <http://www.presumedconsent.org> (last visited Nov. 11, 2004).

72. See Oz et al., *supra* note 57, at 391. But see J.D. Jasper et al., *Altruism, Incentives, and Organ Donation: Attitudes of the Transplant Community*, 42 *MED. CARE* 378, 383 (2004) (finding support for presumed consent/ mandatory donation from only nineteen percent of surgeons, seven percent of transplant center coordinators, and five percent of nurses).

73. See ACOT May 2003 Notes, *supra* note 49, at 5-9. But see *infra* note 78.

74. See ALBERTO ABADIE & SEBASTIEN GAY, *THE IMPACT OF PRESUMED CONSENT LEGISLATION ON CADAVERIC ORGAN DONATION: A CROSS COUNTRY STUDY* (Harvard Univ. John F. Kennedy Sch. of Gov’t, Working Paper No. RWP04-024, 2004), <http://ssrn.com/abstract=562841> (finding a twenty-five to thirty percent increase for a survey of twenty-two nations over ten years); Gimbel et al., *supra* note 70 (finding more than a fifty percent increase for a broad survey of European nations); Eric J. Johnson & Daniel Goldstein, *Do Defaults Save Lives?*, 302 *SCIENCE* 1338, 1339 (2003).

out.⁷⁵ In fact, personal autonomy is valued so highly that no nation has been willing to override it, even to save lives, as by requiring that all usable organs of the dead be made available for transplants.⁷⁶ Accordingly, in France, Greece, Hungary, and Italy, among other nations with presumed consent laws, medical professionals often enforce a de facto “informed consent” policy, deferring to families to determine whether the deceased had preferred not to donate even where no formal record suggests this.⁷⁷ Furthermore, in the United States, there is both significant public opposition to presumed consent⁷⁸ and good reason to question whether it would be effective.⁷⁹

A second, controversial proposal for increasing organ donations is the

75. See Robert M. Veatch & Jonathan B. Pitt, *The Myth of Presumed Consent: Ethical Problems in New Organ Procurement Strategies*, 27 TRANSPLANTATION PROC. 1888 (1995) (expressing particular concern about the impact of presumed consent on the uneducated). Some doctors find it awkward, if not immoral, to take organs absent express permission of the deceased or next of kin. See Cohen, *supra* note 12, at 19-20; James F. Childress, *Ethical Criteria for Procuring and Distributing Organs for Transplantation*, 14 J. HEALTH POL. POL'Y & L. 87, 95-98 (1989). This led many nations to shift to an opt-in approach. See Troy R. Jensen, Comment, *Organ Procurement: Various Legal Systems and Their Effectiveness*, 22 Hous. J. INT'L L. 555, 558-67 (2000).

76. See, e.g., John Harris, *In Praise of Unprincipled Ethics*, 29 J. MED. ETHICS 303, 304-05 (2003). Many advocates of drafting organs from dead bodies note that nations are willing to force the young to risk their lives in war. See, e.g., Theodore Silver, *The Case for a Post-Mortem Organ Draft and a Proposed Model Organ Draft Act*, 68 B.U. L. REV. 681 (1988). Still, nations generally permit conscientious objectors. On the paramount status of autonomy see Raanan Gillon, *Ethics Needs Principles – Four Can Encompass the Rest – and Respect for Autonomy Should Be “First Among Equals,”* 29 J. MED. ETHICS 307, 310-11 (2003).

77. See, e.g., Gerson, *supra* note 71, at 1024; Gimbel et al., *supra* note 70, at 19 (listing France, Greece, Hungary, Italy, Luxembourg, Norway, and Slovenia in this group); Michielsen, *supra* note 63. Also, in presumed consent nations many may register as non-donors in panic. See Jensen, *supra* note 75, at 572-73; Siminoff et al., *supra* note 10, at 16.

78. See VEATCH, *supra* note 32, at 170 (concluding that from the empirical evidence of limited public support for actually donating their organs, “there can be no basis for presuming consent”); TASK FORCE ON ORGAN TRANSPLANTATION, U.S. DEP'T OF HEALTH & HUMAN SERVS., ORGAN TRANSPLANTATION: ISSUES AND RECOMMENDATIONS 30-31 (1986) [hereinafter 1986 HHS TASK FORCE] (rejecting presumed consent due to polling data); Seltzer et al., *supra* note 16, at 354 (reporting that only fourteen to thirty-six percent of respondents supported presumed consent); Siminoff & Mercer, *supra* note 16, at 380 (reporting that only twenty-three percent of respondents supported presumed consent). But see 2001 CANADIAN SURVEY, *supra* note 20, at 45 (half support presumed consent).

79. See CURRAN ET AL., *supra* note 33, at 751 (noting that in the first year after medical examiners were given presumed consent authority in Texas, it was only used twice).

use of financial incentives. There have long been strong objections to using monetary incentives to procure organs, even to pay for funeral expenses.⁸⁰ Many worry that this would lead to exploitation of the poor.⁸¹ An aversion to treating body parts as commodities sold for profit led the 1984 National Organ Transplant Act (NOTA) to prohibit donors from being offered any “valuable consideration,”⁸² and many states followed suit.⁸³ A U.S. Congressional hearing on this issue in June 2003 confirmed strong ongoing and widespread opposition to direct financial incentives.⁸⁴

80. For a general review of the issue of monetary compensation for organs, see Donald Joralemon, *Shifting Ethics: Debating the Incentive Question in Organ Transplantation*, 27 J. MED. ETHICS 30 (2001); Abdullah S. Daar, *Paid Organ Donation: Towards an Understanding of the Issues* in CHAPMAN, *supra* note 63, at 46; and Abdullah S. Daar, *Rewarded Gifting*, 24 TRANSPLANTATION PROC. 2207 (1992).

81. See MUNSON, *supra* note 15, at 116-19; Madhav Goyal et al., *Economic and Health Consequences of Selling a Kidney in India*, 288 JAMA 1589 (2002); Nancy Scheper-Hughes, *Keeping an Eye on the Global Traffic in Human Organs*, 361 THE LANCET 1645 (2003).

82. 42 U.S.C. § 274e (2000) (making it illegal “for any person to knowingly acquire, receive, or otherwise transfer any human organ for valuable consideration for use in human transplantation if the transfer affects interstate commerce”). The provision appears to have been a reaction to a Virginia physician’s efforts to address the organ shortage by brokering living donors’ kidneys in a manner designed to earn a profit. See Cate, *supra* note 60, at 80. Instead of specific justifications, the reports of the House, Senate, and the task force they established to inquire further into these policy matters all offer only conclusory condemnations of organ sales. See, e.g., S. REP. NO. 98-382 (1984), at 16, *reprinted in* 1984 U.S.C.C.A.N. 3975, 3982; H.R. CONF. REP. NO. 98-1127 (1984), at 16, *reprinted in* 1984 U.S.C.C.A.N. at 3989, 3992; 1986 HHS TASK FORCE, *supra* note 78, at 96. The 1968 UAGA had intentionally left open the issue of payment. See E. Blythe Stason, *The Uniform Anatomical Gift Act*, 23 BUS. L. 919, 927-28 (1968).

83. The 1987 UAGA § 10, 8A U.L.A. 29, 64 (2003), adopted by many states, prohibits the sale of organs. See Radhika Rao, *Property, Privacy, and the Human Body*, 80 B.U. L. REV. 359, 376 n.58 (2000).

84. See 2003 House Hearing, *supra* note 5, at 5, 21, 64-67; see also Arnold et al., *supra* note 88 (position of ASTS); *id.* at 1362-63 (position of Pope John Paul II); Thomas J. Cossé & Terry M. Weisenberger, *Encouraging Human Organ Donation: Altruism Versus Financial Incentives*, J. NON-PROFIT & PUB. SECTOR MARKETING, Sept. 1999, at 77; Francis L. Delmonico et al., *Ethical Incentives – Not Payment – For Organ Donors*, 346 NEW ENG. J. MED. 2002 (2002); Jasper, *supra* note 6, at 384 (reporting that a \$1500 cash payment was only supported by only sixteen percent of surgeons, seven percent of transplant center coordinators, and nine percent of nurses); J.D. Jasper et al., *The Public’s Attitudes Toward Incentives for Organ Donation*, 31 TRANSPLANTATION PROC. 2181, 2183 (1999) (reporting that forty-three percent of respondents found a direct payment of \$1500 to be morally inappropriate while only thirty percent found it morally appropriate); Oz et al., *supra* note 57, at 391, 393 (finding that sixty-six percent of those surveyed opposed direct compensation for organs); see also

Although the sale of human organs for transplants is also illegal in almost all nations (with the apparent exceptions of Iran, Kuwait, and the Philippines), such sales have been tolerated with little secrecy in Israel, India, China, and Russia, where there may be little or no penalties for violating the law⁸⁵ (although, that may be changing⁸⁶).

Motivated by the desire to save some of the thousands of lives lost annually under current policies, proposals for limited financial incentives or even restricted markets have been made in books and scholarly journals,⁸⁷ as well as legislative bills.⁸⁸ All recognize the need to address the

KASERMAN & BARNETT, *supra* note 6, at 89-99 (finding that the medical community has a financial incentive to maintain the current rules). *But see* Bryce, *supra* note 11, tbl.3 (fifty-three percent support direct payment).

85. *See* Goyal et al., *supra* note 81, at 1590; Nancy Scheper-Hughes, *The Global Traffic in Human Organs*, 41 CURRENT ANTHROPOLOGY 191 (2000); Michael Finkel, *This Little Kidney Went to Market*, N.Y. TIMES, May 27, 2001, § 6, at 28; Abraham McLaughlin et al., *What is a Kidney Worth?*, CHRISTIAN SCI. MONITOR, Jun. 9, 2004, at 1; Larry Rohter, *Tracking the Sale of a Kidney on a Path of Poverty and Hope*, N.Y. TIMES, May 23, 2004, § 1, at 1; David J. Rothman & Shelia M. Rothman, *The Organ Market*, N.Y. REV. BOOKS, Oct. 23, 2003, at 49; Furthermore, convictions for organ trafficking appear nonexistent. Scheper-Hughes, *supra* note 81, at 1646.

86. *See* McLaughlin et al., *supra* note 85; Rohter, *supra* note 85; Rothman & Rothman, *supra* note 85 (reporting on the scandal at Bangkok's Vachiraprakarn General Hospital; Michael Wines, *14 Arrested in the Sale of Organs for Transplant*, N.Y. TIMES, Dec. 8, 2003, at A6.

87. *See* KASERMAN & BARNETT, *supra* note 6, at 69-88; GREGORY E. PENCE, RE-CREATING MEDICINE: ETHICAL ISSUES AT THE FRONTIER OF MEDICINE 33-62 (2000); William Barnett II & Michael Saliba, *A Free Market for Kidneys: Options, Futures, Forward & Spot*, 30 MANAGERIAL FIN. 38 (2004); Janet Radcliffe-Richards et al., *The Case for Allowing Kidney Sales*, 351 THE LANCET 1950 (1998). The first major proposals appear to have been Marvin Brams, *Transplantable Human Organs: Should Their Sale Be Authorized by State Statute?*, 3 AM. J.L. & MED. 183 (1977); and Richard Schwindt & Aidan R. Vining, *Proposal for a Future Delivery Market for Transplant Organs*, 11 J. HEALTH POL. POL'Y & L. 483 (1986). *See also* Gregory S. Crespi, *Overcoming the Legal Obstacles to the Creation of a Futures Market in Bodily Organs*, 55 OHIO ST. L.J. 1 (1994) (comparing four proposals).

88. *See, e.g.*, Gift of Life Tax Credit Act of 2001, H.R. 1872, 107th Cong. (2001) (granting a \$10,000 tax credit for donated organs); *Putting Patients First: Increasing Organ Supply for Transplantation: Hearing Before the House Subcomm. on Health & Environment, House Comm. on Commerce*, 106th Cong. 87-126 (1999) [hereinafter 1999 House Hearing] (describing \$10,000 insurance policy for donors known as Project Donor); H.R. 540, 98th Cong. (1983) (allowing organ donors to deduct \$25,000 per organ from their estate). Pennsylvania enacted a statute creating a special state fund that could provide donors and donors' families up to \$3000 to pay for reasonable hospital, medical, and funeral expenses incurred in connection with organ donation. 20 PA. CONS. STAT. § 8622(b)(1) (2003), discussed in Robert M. Arnold et al., *Financial Incentives for Cadaver Organ Donation: An Ethical*

ethical dangers such incentives produce; suggested strategies have included providing appropriate, continuing medical aftercare to living donors and preventing sales that would merely enable creditors to squeeze a bit more out of debtors.⁸⁹ Assuming that this could be done, many have noted that permitting sales might actually aid the disadvantaged by allowing them to avoid even less attractive options, like taking a life-threatening job or being forced to watch a child die for lack of funds for medical care.⁹⁰ Thus, the AMA, ASTS, and UNOS/OPTN (Organ Procurement Transplant Network)⁹¹ all now support the *study* of financial options.⁹² Since 1994, Georgia has reduced its drivers' license fees for those who agree to donate their organs.⁹³

Reappraisal, 73 TRANSPLANTATION 1361, 1366 (2002). Fearing that a \$3000 payment could be coercive, the designated organ donor committee recommended payments of \$300. Due to state officials' concern that this statute may violate federal law, however, funds collected for this program are, instead, being used to offset travel and lodging expenses of live donors. See Christopher Snowbeck, *Organ Donor Funeral Aid Scrapped*, PITTSBURGH POST-GAZETTE, Feb. 1, 2002, at B1.

89. See, e.g., Gloria J. Banks, *Legal & Ethical Safeguards: Protection of Society's Most Vulnerable Participants in a Commercialized Organ Transplantation System*, 21 AM. J.L. & MED. 45, 83-107 (1995).

90. See Rohter, *supra* note 85 (quoting Orley de Santana, a twenty-six-year-old Brazilian laborer, who stated "in order not to have to steal or kill, I thought it better to sell my kidney" for \$6,000); cf. Nicholas D. Kristof, *Inviting All Democrats*, N.Y. TIMES, Jan. 14, 2004, at A19 (describing the dangerous, uncomfortable, and very low-paying work that many Cambodians engage in because they have no better options, possibly because some better options had been prohibited by well meaning, but naïve, social liberals). In fact, a 2001 study found that about sixty-four percent of non-whites supported direct payments to families who agreed to donate a kin's organ. See Bryce, *supra* note 11, tbl. 3.

91. UNOS is the contractor that HHS selected to administer the Organ Procurement and Transplantation Network.

92. See AM. MED. ASS'N., CODE OF ETHICS § E-2.15 (Aug. 7, 2002), <http://www.ama-assn.org/ama/pub/category/8445.html> (Financial Incentives for Organ Donation); Francis L. Delmonico & James A. Schulak, Letter to the Editor, WASH. POST, May 12, 2002, at B06 (position of the ASTS); Press Release, UNOS, OPTN/UNOS Board Endorses Studies of Incentives to Increase Donation (June 28, 2002), <http://www.unos.org/news/newsDetail.asp?id=1>.

93. See GA. CODE ANN. § 40-5-25(d)(2) (2003), adopted in 1994 Ga. Laws 1200. The maximum \$8 fee for donors (versus \$15 for others) was upheld as constitutional in *Barnhill v. State*, 575 S.E.2d 460 (Ga. 2003). Nevertheless, the Georgia governor proposed repealing the discount, claiming that it had little impact: although it raised donation rates by thirty-three percent. See Brian Basinger, *Organ Donor Discount Could End*, SAVANNAH MORNING NEWS, Feb. 6, 2003, <http://www.savannahnow.com/stories/020603/LOCXGRDonor>

In summary, current efforts leave half of all usable organs from cadavers unused, and proposals for presumed consent and markets in organs face stiff political opposition. Against this background it is useful to consider a less controversial option—a reciprocity policy—which is also compatible with both presumed consent and financial incentives.⁹⁴

III. A RECIPROCITY PROPOSAL

Fortunately, a relatively simple adjustment to the organ donation rules would likely alleviate the two central problems with the current system by inducing many more commitments to donate and deterring families from challenging those wishes. Instead of asking individuals to act purely altruistically, UNOS/OPTN⁹⁵ should formally recognize those who commit to donate organs at death by significantly increasing such individuals' chances of receiving an organ should they later need one.

Variations of this idea have been proposed periodically over the last twenty years, apparently beginning with Jonathan Kaufelt's 1986 letter in the *Wall Street Journal*.⁹⁶ One version of this proposal was adopted by

Discounts.shtml. In addition, the firm Administrative Resource Options (ARO) has a program to reimburse every one of its employees for the cost of their drivers' licenses if they sign up to be an organ donor. Memorandum from Jenn Hirjak, Donate for Life Benefit Program (July 8, 2004) (on file with authors). See also *Marketing Organ Donations: Give Speeders a Break?*, ASSOCIATED PRESS (June 28, 2004), <http://www.wcpo.com/wcpo/localshows/healthyliving/3aa3fad8.html> (describing a billboard in Cleveland, which calls out "Hey policeman," with an arrow pointing to a donor insignia on a young man's license, "give this guy a break").

94. In fact, one commenter supports a reciprocity policy as a useful adjunct to presumed consent. See Stephanie Eaton, *The Subtle Politics of Organ Donation: A Proposal*, 24 J. MED. ETHICS 166 (1998); see also *infra* note 97 (discussing Singapore law).

95. UNOS identifies criteria that may be used for allocating organs. See 1993 GAO REPORT, *supra* note 27, at 18; see also 42 U.S.C. § 273(b)(3)(E) (2002); Jeffrey Prottas, *Rationing Human Organs for Transplant*, in TRANSPARENCY IN PUBLIC POLICY: GREAT BRITAIN AND THE UNITED STATES 70, 76-77 (Neal D. Finkelstein ed., 2000). In contrast, in England, individual surgeons set their own allocation criteria. *Id.* at 82-83.

96. See Jonathan D. Kaufelt, Letter to the Editor, *Meeting the Need for Human Organs*, WALL ST. J., May 15, 1986, at 29; Peters, *supra* note 49, at 177-82; Irwin Kleinman & Frederick H. Lowy, *Cadaveric Organ Donation: Ethical Considerations for a New Approach*, 141 CANADIAN MED. ASS'N. J. 107, 109-10 (1989); James Muyskens, *Procurement and Allocation Policies: Should Receiving Depend on Willingness To Give?*, 56 MT. SINAI J. MED. 202 (1989); Irwin Kleinman & Frederick H. Lowy, *Ethical Considerations in Living Organ Donation and a New Approach: An Advance-Directive Organ Registry*, 152 ARCHIVES INTERNAL MED. 1484, 1486-87 (1992); James Muyskens, *Should Receiving Depend upon Willingness To Give?*, 24

Singapore in 1987.⁹⁷ Although a 1993 UNOS Committee Report recommended wider discussion of a priority program,⁹⁸ it has generally been overlooked by policymakers. The idea was never raised during either the 1999 or 2003 hearings in the House of Representatives on increasing organ donations,⁹⁹ nor was it identified in either the 1993 General Accounting Office (GAO) review of alternatives for achieving this goal¹⁰⁰ or

TRANSPLANTATION PROC. 2181 (1992); Rupert Jarvis, *Join the Club: A Modest Proposal To Increase Availability of Donor Organs*, 21 J. MED. ETHICS 199, 202-03 (1995); Richard Schwindt & Aidan Vining, *Proposal for a Mutual Insurance Pool for Transplant Organs*, 23 J. HEALTH POL. POL'Y & L. 725 (1998); Gundolf Gubernatis & Hartmut Kliemt, *A Superior Approach to Organ Allocation and Donation*, 70 TRANSPLANTATION 699 (2000); Alexander Tabarrok, *The Organ Shortage: A Tragedy of the Commons*, in ENTREPRENEURIAL ECONOMICS: BRIGHT IDEAS FROM THE DISMAL SCIENCE 107, 109-10 (Alexander Tabarrok ed., 2002); Adam J. Kolber, *A Matter of Priority: Transplanting Organs Preferentially to Registered Donors*, 55 RUTGERS L. REV. 671 (2003); Sackner-Bernstein & Godin, *supra* note 5; Steinberg, *supra* note 64; Michael J. Booker, *Justice and the Macroallocation of Human Donor Organs* 146-54 (1990) (unpublished Ph.D. dissertation, University of Tennessee) (on file with authors); *see also* Dukeminier, *supra* note 12, at 848 (priority for family if they agree to donate). In addition, a 1993 survey found that seventy percent of those eighteen to twenty-four were either very or somewhat interested in using a form of preference. *See* BOWDEN & HULL, *supra* note 16, at 117. *But see* Marlies Ahlert et al., *Common Sense in Organ Allocation*, 23 ANALYSE & KRITIK 221, 226-27 (2001) (finding that a majority of German students in one study oppose such a preference); Jasper, *supra* note 72, at 384 (reporting that a priority policy was supported by only forty-five percent of surgeons, thirty-four percent of transplant coordinators, and forty percent of nurses); Siminoff & Mercer, *supra* note 16, at 380 (reporting that only twenty-five percent of respondents believed that people who have signed organ donor cards should get a preference).

97. Singapore's Human Organ Transplant Act, adopted July 16, 1987, establishes both a system of presumed consent, 131A C.A.P. §§ 5, 9-11 (Sing.), <http://statutes.agc.gov.sg>, and priority in receipt of organs for those who have not opted out., *id.* § 12. Despite the latter and financial incentives, by 1997 only three percent of Muslims—exempt from the presumed consent provision—had registered to donate. *See* Volker H. Schmidt & Lim Chee Han, *Organ Transplantation in Singapore: History, Problems, and Policies* 6-7 (Aug. 2003) (unpublished manuscript, on file with authors). Still, the priority policy is probably not publicized widely in Singapore, given that most of the population is subject to presumed consent. Thus, most Muslims may be unaware of the policy, and it seems unlikely that any significant efforts were made to inform Muslims and encourage individuals to register.

98. *See* JAMES F. BURDICK ET AL., *PREFERRED STATUS FOR ORGAN DONORS: A REPORT OF THE UNITED NETWORK FOR ORGAN SHARING ETHICS COMMITTEE* (1993), <http://www.unos.org/resources/bioethics.asp?index=5> (concluding that the idea required further discussion). Such efforts, however, appear to have been neglected in favor of other priorities.

99. *See* 1999 House Hearing, *supra* note 88; 2003 House Hearing, *supra* note 5.

100. *See* 1993 GAO REPORT, *supra* note 27, at 61-65.

the June 2004 Joint Committee on Accreditation of Healthcare Organizations (JCAHO) report entitled *Strategies for Narrowing the Organ Donation Gap and Protecting Patients*.¹⁰¹

While almost all of the suggestions for this approach have been offered in general, conceptual terms, this Commentary attempts to offer a detailed proposal that can be tested. This Commentary also offers specific reasons why this approach should be effective and attempts to respond comprehensively to potential counter-arguments.

A. How the Proposal Would Work

Under the reciprocity policy proposed here, those who committed to donate would receive a significant advantage in the organ allocation process, if they later needed a transplant. This would enable them, like military veterans seeking a government job, to be placed ahead of non-donors of slightly superior qualifications on the waiting list. For kidneys, where potential organ recipient scores are in the range of about ten to twenty-five, and former live kidney donors receive four extra points,¹⁰² committed donors might receive up to two points on their kidney score.¹⁰³ The bonus would be phased in, based on how long a patient had been registered as willing to donate (similar to the "time on waiting list" criteria now used).¹⁰⁴ Individuals, including young adults who had been registered

101. See 2004 JCAHO REPORT, *supra* note 66.

102. The point system for kidney allocation is based on time on the waiting list (1 point for each year and up to 1 point on each list), quality of "antigen" match (2, 1, or 0 points), the presence of reactive antibodies (4 points), and age (4 points if 3-11 years old, 3 if 11-18). Former donors receive 4 extra points and medical urgency is considered. See UNOS Organ Distribution Rules, *supra* note 38, § 3.5.11.6 (Point System for Kidney Allocation).

103. Selecting an appropriate size preference is important, see BURDICK ET AL., *supra* note 98, and, given the time and data, one might seek the pareto optimal level that maximizes the number of lives saved while not leaving any non-donor worse off, see Kolber, *supra* note 96, at 704-14, or simply maximizes the number of lives saved. Two points is suggested here as a reasonable estimate of the optimal value, which would appear to be between four points and zero.

104. See Hartmut Kliemt, Clubs and Reciprocity in Organ Transplantation 9-10 (2003), http://www.indiana.edu/~workshop/colloquia/papers/kliemt_paper.pdf; see also *supra* note 102. The full bonus might be reached ten years after one had committed to donate. Alternatively, there could be a waiting period before the bonus took effect or those who had not committed to donate before they needed an organ could be denied any bonus at all. Some such policy is needed to encourage healthy people to commit to donate. Singapore uses a two year waiting period, 131A C.A.P. § 12(1)(b) (Sing.), as does Peters, *supra* note 49, at 180.

by their parents, would be permitted to change their minds, but anyone who removed themselves from the committed-to-donate list would lose credit for the time they had already been listed, even if they later re-registered.¹⁰⁵ For livers and hearts, committed donors might be granted first priority within their “status” group (i.e., 1A, 1, 2, etc.) and ranked within the group based on how long they had been on the committed-to-donate list.¹⁰⁶

To motivate those who expect to be denied access to a transplant due to the green screen, the preference could also include a chance to benefit from funds set aside to cover at least one “free” organ transplant annually.¹⁰⁷ These patients would be given a contingent status on the waiting list—only considered for a transplant if funds were available at the time an organ was available.

Individuals would continue to record their commitments in a manner similar to the way they currently do—through license renewals at offices of state DMVs or by filling out organ donor forms made available elsewhere, including health care facilities, voter registration offices, or other social service agencies. For individuals to receive preferences, their commitments would have to be recorded in registries—databases maintained by individual states for their residents¹⁰⁸ or in a national database for residents of states without their own databases, which would also linked to existing state registries.¹⁰⁹

The status of those whose medical condition, e.g., those with HIV or

105. Babies could be enrolled by parents. See Coleman, *supra* note 66, at 40-41; Raanan Gillon, *On Giving Preference to Prior Volunteers When Allocating Organs for Transplantation*, 21 J. MED. ETHICS 195, 195 (1995); Aidan R. Vining & Richard Schwindt, *Have a Heart: Increasing the Supply of Transplant Organs for Infants and Children*, 7 J. POL’Y ANALYSIS & MGMT 706, 708 (1988). Given the burden of reconsidering the decision to donate, see Johnson & Goldstein, *supra* note 74, it would seem unlikely that many would change their minds. On the other hand, the proposal would subject any person who attempted to gain the preference for registering while using some other legal device to nullify that commitment in the case that they died, to a significant fine for fraud. Furthermore, it would impose criminal penalties on anyone who conspired to organize multiple frauds of this kind.

106. Those on the UNOS heart and liver waiting lists are given a status, e.g., 2, 1, 1A, depending on their condition. If the medical community believed that giving committed donors first priority in their status group was too great a bonus, it could subdivide the status group or award a set number of relevant points.

107. These funds might come from private donations or NIH; alternatively, UNOS could add a \$500 charge for each organ transplanted.

108. Cf. *supra* note 62.

109. Cf. *supra* note 63.

Hepatitis C, makes them unacceptable donors raises a difficult question. This proposal would permit such individuals to get equal credit for agreeing to donate their body for medical research on transplantation. Other options might be to permit those unable or unwilling to donate their organs to make alternative efforts to increase the supply of organs, as by helping to educate the public at health fairs; however, this would raise many administrative questions about precise standards.¹¹⁰

Donors' commitments would effectively represent organ insurance,¹¹¹ not unlike the former "family credit" blood donor systems, under which a blood donor's contribution served to cover his or her family's annual blood needs.¹¹² The proposal advocated here would operate somewhat differently than "club" systems,¹¹³ like Singapore's,¹¹⁴ or "LifeSharers," the provocative directed donation entity.¹¹⁵ Rather than offering only a limited preference for committed donors,¹¹⁶ club proposals favor a minimally medically compatible club member over non-members who are much

110. Individuals currently HIV positive or with Hepatitis C might be asked to provide similar service and this might also be offered to those with other objections. See Abdullah S. Daar, *Altruism and Reciprocity in Organ Donation: Compatible or Not?*, 70 *TRANSPLANTATION* 704, 704-05 (2000); Peters, *supra* note 49, at 180-82. Then again, Illinois recently recognized that HIV positive patients may donate to other HIV positive patients. 2004 Ill. Legis. Serv. 93-737 (West) (codified at ILL. COMP. STAT. 20 § 2310-330(c-5)).

111. See PAUL RAMSEY & MARGARET A. FARLEY, *THE PATIENT AS PERSON: EXPLORATIONS IN MEDICAL ETHICS* 212 (2d ed. 2002) ("This practice [of rewarding blood donors with insurance against their future needs] of giving and receiving, not buying and selling, is the one that should be extended to other tissue."); Muyskens, *supra* note 96, at 2182; Schwindt & Vining, *supra* note 96; Tabarrok, *supra* note 96, at 109.

112. See DOUGLAS STARR, *BLOOD: AN EPIC HISTORY OF MEDICINE AND COMMERCE* 190, 250-56 (1998); RICHARD M. TITMUSS, *THE GIFT RELATIONSHIP: FROM HUMAN BLOOD TO SOCIAL POLICY* 82 (1971).

113. See Jarvis, *supra* note 96.

114. See *supra* note 97.

115. Open to all willing donors, this program requires members to agree to donate their organs (upon death) to another member of the club if a member is a medically acceptable recipient. See LifeShares, *How LifeSharers Works*, at <http://www.lifesharers.com/howitworks.htm> (last visited Nov. 18, 2004); see also Chris Fusco, *An Organ Transplant is a Mouse Click Away*, CHI. SUN-TIMES, Nov. 23, 2002, at 3. LifeSharers members make directed donations, which appear to be legal. See *supra* note 32. However, this has been subject to criticism. See Sheldon Zink et al., *Examining the Potential Exploitation of UNOS Policies* (Sept. 2004) (unpublished manuscript, on file with authors) (criticizing the fairness of directed donations other than those to family members).

116. Organ seekers receiving small preferences may still face long waits. See Delmonico et al., *supra* note 84, at 2004.

better matches,¹¹⁷ in the same manner as the current “local first” preference rules favor local recipients over better-matched recipients outside the local area.¹¹⁸ There would certainly be a greater incentive for people to register under a club system rather than a bonus system, yet it is not clear that the incremental benefit from a marginally increased incentive justifies the cost of favoring a barely compatible recipient over one who was an excellent match.

B. Likely Effects of the Proposal

There are good reasons to believe that, by making it in a person’s self interest to commit to organ donation, a priority policy would produce significantly more donations. In fact, the policy would respond to both current problems deterring donations: It should convince more people to sign up to donate and make it more likely that those wishes will be honored, even if the donors’ families would prefer to override them.

First, the policy would appear to significantly increase the likelihood that individuals would sign up to donate when they were seeking a driver’s license renewal or during a visit to their doctor. With respect to the former, it is reasonable to assume that a significant number of individuals who presently decline to check the box for organ donor on their driver’s license renewal are neutral or only slightly predisposed against signing up. Some may have slight concerns that registering as donors would lead doctors to work less hard to save their lives, but even a small doubt might be enough to outweigh an even smaller expected benefit from acting altruistically. For many of such current borderline non-donors, a small, but significant health benefit should lead them to choose to donate.

This effort might also be aided by a new marketing approach. While the most effective publicity in the past has involved celebrity athletes¹¹⁹ or poignant stories about children,¹²⁰ a different tactic might well better motivate visitors to the DMV. Instead of relying solely on the positive

117. See Timothy F. Murphy & Robert M. Veatch, *Members First: The Ethics of Donating Organs and Tissues to Groups* (2004) (unpublished manuscript, on file with authors).

118. See Schwindt & Vining, *supra* note 96, at 736; *infra* notes 139-142 and accompanying text.

119. For example, the NBA star Alonzo Mourning has brought considerable attention to the topic of organ donation. See Chris Broussard, *Dozens Offer a Kidney to Mourning*, N.Y. TIMES, Nov. 26, 2003, at D1; Maureen Dowd, *Give Thanks and Life*, N.Y. TIMES, Nov. 27, 2003, at A39.

120. See, e.g., REG GREEN, *THE NICHOLAS EFFECT* (1999), <http://www.nicholasgreen.org/contents.html>.

feelings people should get from donating, which might be too weak to trigger registration, instructions about registering to donate on driver's license forms could highlight how non-donors could lose out. For example, instructions might note that "failure to agree to donate could permit those who have committed to donate to move ahead of you on the organ wait list if you later need an organ." Studies have shown that individuals are much more likely to act to avoid a bad outcome ("loss aversion") than to obtain a comparable good result.¹²¹

The health benefit from committing to donate should also make it more likely that doctors and nurses would place donor registration forms in their waiting rooms and, if there was time at the end of check-ups, recommend donation, possibly right after they typically now suggest how patients might improve their diets and exercise regimes. While patients concerned about their health—particularly those whose test results served as a wake-up call of potential danger—may find it difficult to maintain their good intentions regarding diet and exercise for a few weeks or even days, registering to donate would require no ongoing motivation; a simple recommendation to act should often be enough to trigger a registration.

Also, since those entitled to this preference would be less likely to die for lack of an organ, life insurance companies might well offer them a discount.¹²² Some individuals who noticed this when purchasing life insurance or comparing policy prices might find it sufficient motivation to register to donate.

A priority policy should also help to address the second problem with donation: enforcing a donor's wishes against family opposition. Today, family members may well regard a donor's decision to donate as a unilateral charitable impulse, whose revocability should continue after their death, even though the law is otherwise. Once a transplant specialist had politely informed them about the basic concept of a priority policy, however, most family members would likely recognize that the donor's decision to donate was part of a quid pro quo agreement. Most would probably understand that it would be wrong for them to try to renege on the donor's death-triggered promise. Thus, one would expect fewer families to attempt to override a donor directive, and it should be easier

121. This psychological phenomenon is called "anticipatory regret," which appears to be the same as "loss aversion," discussed in Alexander J. Rothman et al., *The Systematic Influence of Gain- and Loss-Framed Messages on Interest in and Use of Different Types of Health Behavior*, 25 PERSONALITY & SOC. PSYCHOL. BULL. 1355 (1999).

122. Given how long it took life insurance companies to give non-smokers a discount, however, this would likely be a long time in coming.

for transplant specialists to overcome any resistance offered.

Finally, although non-donors on the waiting list would sometimes be bypassed by a patient with a bonus, a substantial increase in the total supply of organs triggered by this policy should more than offset that loss, actually increasing even non-donors' chances to receive an organ. Of course, one's chances would still be better if one committed to donate.

C. Responses to Main Criticisms

The reciprocity policy has been subject to a number of criticisms, but none appear to be very persuasive.

The most significant charge is that the policy would not produce more donations. Although there is good reason to believe that the proposal would increase the supply of transplantable organs,¹²³ it would certainly be sensible to test it—in a state with an existing database of committed donors¹²⁴—before adopting it more widely. At least four types of effects would deserve to be evaluated. First, it would be useful to review DMV records to measure the effect of a short statement on drivers' license forms that explained the benefit of a preference and how those who did not sign up could be bypassed on the waiting list by others who had signed up. Second, it would be important to survey primary care physicians to determine whether a reciprocity policy led any of them to make a greater effort to encourage their patients to sign up, such as providing forms in their waiting rooms and encouraging patients to fill them out. Third, it would be relevant to see whether the policy led a smaller percentage of families to seek to override a donor's directive after being informed of the quid pro quo nature of the priority policy. Fourth, it would be useful to try to determine whether the type of people who were spurred to register to donate by this policy were demographically similar to current donors or whether they were more (or less) likely die in a manner that led them to be suitable donors.

A second complaint about a reciprocity policy is that it would threaten the purity of altruistic efforts. Thus, an UNOS Committee evaluating the reciprocity concept in a 1993 report found "the most important negative aspect of the idea" is that, like "all other forms of inducement, [a preferred status priority system] is likely to be seen by some as inherently compromising the altruism" of the current voluntary system.¹²⁵ Yet public

123. See *supra* Subsection III.B.

124. See UNOS, *supra* note 62.

125. See Burdick et al., *supra* note 98.

health is rooted in enlightened self-interest, i.e. utilitarian principles;¹²⁶ society does not expect transplant or other healthcare professionals to be motivated solely by altruism. Moreover, a priority policy would actually represent a form of "reciprocal altruism."¹²⁷ Granting an optional preference to committed organ donors seems no more morally harmful than making charitable contributions tax deductible. Furthermore, like the latter, it should increase, not decrease the incentive to donate.¹²⁸ Finally, an excellent, detailed examination of the significance of altruism in the context of organ donations exposed the inconsistencies in the arguments that incentives, like a priority system, are detrimental to altruism or contribute to inhumane "commoditization" of the human body.¹²⁹

A third concern may be that a preference might be considered "valuable consideration" for an organ donation, which arguably would violate the current law,¹³⁰ but that seems very unlikely for two reasons. First, as a technical matter, there would be no actual exchange of organ for value. The deceased parties who actually donated their organs would not receive any compensation and those who benefited from the preference would not have donated their organs.¹³¹ Second, prosecutors and legal counsel for UNOS already seem to recognize that the ban on compensation for organ donors does not apply to the current UNOS policy of rewarding live kidney donors (or paired partners) with a preference,¹³² and both should regard this policy the same way.¹³³ Still, to

126. See A.H.M. Kerkhoff, *Origin of Public Health and Preventative Medicine*, in *ETHICAL DILEMMAS IN HEALTH PROMOTION* 35 (Spyros Doxiadis ed., 1987).

127. See Robert Trivers, *The Evolution of Reciprocal Altruism*, 46 Q. REV. BIOLOGY 35-57 (1972); cf. Ernst Fehr & Simon Gächter, *Fairness and Retaliation: The Economics of Reciprocity*, 14 J. ECON. PERSP. 159 (2000).

128. The even more provocative policy of financial incentives would seem to produce an even greater net gain. See, e.g., Bryce, *supra* note 11, tbls. 2, 4 (indicating that seventeen percent are more willing to donate; less the eight percent who are less willing to donate will yield a nine percent net gain); GALLUP POLL, *supra* note 10, at 43 (demonstrating a net gain of seven percent).

129. Kolber, *supra* note 96, at 714-37; see also Henry Hansmann, *The Economics and Ethics of Market for Human Organs*, 14 J HEALTH POL. POL'Y & L. 57, 68, 74-78 (1989); Julia D. Mahoney, *The Market for Human Tissue*, 86 VA. L. REV. 163, 215-20 (2000). But see Arnold et al., *supra* note 88.

130. See *supra* note 82-83.

131. See also Kolber, *supra* note 96, at 698-700.

132. That policy is noted *supra* note 102 and *infra* note 138. The legislative history does not indicate any opposition to this practice. See *supra* note 82. Nor was there any proposal to ban it, despite clear notice, when Congress considered a revision to the definition of

avoid any confusion, laws that now ban compensation for organs should be amended to add this form of reciprocity/insurance to the list already exempted from such bans.¹³⁴

Fourth, some argue that it is critical for organ allocations to avoid the corrupting influence of non-medical issues,¹³⁵ but there are three responses to this point. First, it is not clear what should be considered as “medical” criteria and why such criteria do not raise ethical issues. To the extent that medical criteria focus on not “wasting” a scarce organ on a likely medical failure, then a commitment to donate comes close to satisfying that criteria, by helping to reduce the waste of scarce organs. Granted, it is not a pure medical factor, but it appears much closer to one than to a subjective criteria like social worth, which requires subjective judgments and ethical questions about their relevance.

In addition, many features of the current organ allocation system are justified principally by their impact on the organ supply or on non-medical social values.¹³⁶ Some may consider “time on waiting list” as a proxy for

“valuable consideration” in 2004. *See also* Williams Mullen, Legal Memorandum to UNOS, Intended Recipient Exchanges, Paired Exchanges and NOTA §301 (Mar. 7, 2003), http://asts.org/ezefiles/UNOSSection_301_NOTA_.pdf (explaining why 42 U.S.C. §274(e) does not apply to such exchanges).

133. *See also* Burdick et al., *supra* note 98 (citing UNOS Ethics Committee report that a trial of priority incentives “could be implemented without requiring any alteration in existing legislation, unlike other mechanisms under discussion”).

134. 42 U.S.C. § 274e(c)(2) now reads: “valuable consideration does not include the reasonable payments associated with the removal, transportation, implantation, processing, preservation, quality control and storage of a human organ or the expenses of travel, housing, and lost wages incurred by the donor of a human organ in connection with the donation of an organ.” S.573, 108th Cong. (2003), passed by the Senate, would have amended that provision by adding, at the end of it: “Such term does not include familial, emotional, psychological, or physical benefit to an organ donor, recipient, or any other party to an organ donation event.” The version that became law, however, did not include that provision. *See* Pub. L. No. 108-216, 118 Stat. 584 (2004) (codified at 42 U.S.C. §§ 273a, 274f). State laws should also be amended. *See* ACOT May 2003 Notes, *supra* note 49, at 4.

135. *See* 42 C.F.R. § 121.8 (2003); Gillon, *supra* note 103, at 196 (contending that the Achilles heel of the club proposal is creating a slippery slope of using non-medical criteria for allocating organs); *see also* Alexander M. Capron, *More Blessed to Give Than Receive?*, 24 TRANSPLANTATION PROC. 2185 (1992).

136. *See* Stefanos A. Zenios et al., *Dynamic Allocation of Kidneys to Candidates on the Transplant Waiting List*, 48 OPERATIONS RES. 549, 564-66 (2000); Mark E. Votruba, Efficiency-Equity Tradeoffs in the Allocation of Cadaveric Kidneys 49-52 (Nov. 15, 2001 draft) (unpublished Ph.D. dissertation, Princeton University), http://www.princeton.edu/~mvotruba/KA_text.pdf. The policy of favoring those on the list longest, separate and apart

urgency, but the failure to replace it with a better metric for medical urgency is probably due to the view that it is only fair to favor those who have waited longest, even though this is a biased statistic.¹³⁷ As noted earlier, some paired partners of living kidney donors already receive priority access to an organ in consideration for their partners' contribution to the supply of organs.¹³⁸ Also, as mentioned above, while the rapid deterioration of organs justifies a preference for shorter transport times,¹³⁹ the current "local first" preference is much greater than medically justified.¹⁴⁰ The rationale offered is that more individuals will donate organs if they know that they will most likely be aiding someone in their

from their medical condition, appears to be due to "fairness," see Childress, *supra* note 75, at 104-05, since the data do not support such a large preference for them, INSTITUTE OF MEDICINE, ORGAN PROCUREMENT AND TRANSPLANTATION: ASSESSING CURRENT POLICIES AND THE POTENTIAL IMPACT OF THE DHHS FINAL RULE 90 (1999) [hereinafter 1999 IOM REPORT]; Votruba, *supra*, at 38. Also, the list of UNOS objectives includes some non-medical goals. See UNOS, UNOS RATIONALE FOR OBJECTIVES OF EQUITABLE ORGAN (1994), <http://www.unos.org/resources/bioethics.asp?index=8>.

137. See Gabriel M. Danovitch et al., *Waiting Time or Wasted Time? The Case for Using Time on Dialysis To Determine Waiting Time in the Allocation of Cadaveric Kidneys*, 2 AM. J. TRANSPLANTATION 891 (2002).

138. See Lainie Friedman Ross & Stefanos Zenios, *Practical and Ethical Challenges to Paired Exchange Programs*, 4 AM. J. TRANSPLANTATION 1553 (2004) (noting that, in 2001, region one of UNOS developed a program, now called "list pair exchange," whereby those seeking an organ could go to the head of the recipient line if they found a live person willing to donate an organ on their behalf); David Wessel, *Easing the Kidney Shortage*, WALL ST. J., Jun. 17, 2004, at B1. HHS supports such preferences. See ACOT May 2003 Notes, *supra* note 49, at 2 (HHS supports ACOT recommendation #5). Furthermore, research indicates that such programs produce a net gain of organs. See STEFANOS ZENIOS ET AL., PRIMUM NON NOCERE: AVOIDING HARM TO VULNERABLE WAIT LIST CANDIDATES IN AN INDIRECT KIDNEY EXCHANGE (Graduate Sch. of Bus., Stanford Univ., Research Paper No. 1684, 2001), <http://gobi.Stanford.edu/ResearchPapers/Library/RP1684.pdf>. Careful structuring can even yield a net gain for blood type O organ recipients. See Lainie Friedman Ross & Stefanos Zenios, *Restricting Living-Donor-Cadaver-Donor Exchanges To Ensure that Standard Blood Type O Wait-List Candidates Benefit*, 78 TRANSPLANTATION 641 (2004).

139. The maximum allowable transport time for organs removed for transplant (also known as cold ischemic time) limits how far they can be sent to recipients. See Introduction to Transplants, at http://www.ustransplant.org/primer_intro.php (last updated July 9, 2004). There is also a cost advantage to minimizing transport time. See Mark A. Schnitzler et al., *The Economic Impact of Preservation Time in Cadaveric Liver Transplantation*, 1 AM. J. TRANSPLANTATION 360 (2001).

140. See Votruba, *supra* note 136, at 112. Thus, some suggest accounting for travel time directly. See Sackner-Bernstein & Godin, *supra* note 5, at 158..

own “community,”¹⁴¹ but the evidence does not support this.¹⁴²

Finally, while some might perceive a preference policy as favoring committed donors due to their moral superiority over non-donors, that is not the case: the preference is based solely on a person’s willingness to participate in a reciprocal system designed to increase donor incentives and thus the supply of organs. Thus an unemployed ex-convict who committed to donate would get the preference, while a Nobel Peace Prize winner who did not commit to donate would not. It is not an inherently subjective, and thus problematic, policy; it is objective and treats all individuals on the same terms.

As a fifth matter, a reciprocity system could be seen as unfairly punishing those currently receiving the worst health care, many of whom would fail to commit to donate out of ignorance of the policy. Yet this seems no different from the impact of the current preference for those who have been on the waiting list longest. After all, those now receiving the worst health care are likely to be late in discovering their need for a transplant and thus not enroll on the UNOS list until months, if not years, after those with the identical condition who receive superior healthcare.¹⁴³ Meanwhile, unlike the current “local first” policy (which favors those who can afford to register at multiple locations¹⁴⁴), a reciprocity policy would treat rich and poor equally¹⁴⁵ (except for those unable to finance a

141. See MUNSON, *supra* note 15, at 49; 1999 House Hearing, *supra* note 88, at 48-52, 54-56, 72, 77; 1996 HHS Hearings, *supra* note 17, at 76-77 (testimony of Dr. D’Alessandro).

142. See 1999 IOM REPORT, *supra* note 136, at 52-53 (1999) (reporting that both a 1998 Gallup poll and a 1995 Southeastern Inst. of Research poll found little patient preference for local recipients over more needy patients in the nation). On the other hand, a local preference probably serves to improve the morale and motivation of those involved in encouraging organ donation in each community. See KOCH, *supra* note 36, at 74, 97-99. This policy also reflects the efforts of smaller, local transport centers to protect themselves and their patients. See Jeffrey Prottas, *The Politics of Transplantation*, in ORGANS AND TISSUE DONATION, *supra* note 27, at 3, 17.

143. To avoid the bias against the disadvantaged caused by using time on wait list, UNOS should require that OPOs that desire to use such a metric to use time on dialysis instead. See Danovitch et al., *supra* note 137.

144. The current UNOS system permits wealthy or well-insured organ seekers to increase their chances of receiving an organ by registering at multiple transplant centers. See Robert M. Merion et al., *Prevalence and Outcomes of Multiple-Listing for Cadaveric Kidney and Liver Transplantation*, 4 AM. J. TRANSPLANTATION 94 (2004); Tracy E. Miller, *Multiple Listing for Organ Transplantation: Autonomy Unbounded*, 2 KENNEDY INST. ETHICS J. 43 (1992).

145. The 1993 UNOS Report found this aspect of a preference system admirable. See Burdick et al., *supra* note 98.

transplant operation¹⁴⁶), and the system would not encourage black market donations.¹⁴⁷ Certainly society should work to provide the most disadvantaged with healthcare more fully and effectively, but the flaws in the current system are no more of a justification for rejecting a preference policy than they are for rejecting the use of the “time on waiting list” statistic for allocating organs.

A sixth complaint might be that the system would discriminate against those who refused to donate for religious or other reasons,¹⁴⁸ but this would not appear to create unfairness. Religions that forbid organ donations would seem, almost necessarily, to reject organ transplantation generally, and thus their believers would not desire organs at all, certainly not a preference over others who had chosen not to donate. It should also be noted that veterans’ preferences already discriminate against pacifists, and that fifty-nine percent of transplant professionals surveyed would go so far as to refuse access to the donor pool to those who refuse to donate because of religious reasons.¹⁴⁹ Finally, the preferences here would not be based on an individual’s minority group status, but rather, only on their actual willingness to aid the organ donor pool.¹⁵⁰

CONCLUSION

The substantial health benefit of a system of reciprocal organ donation incentives and its minimal cost (for maintaining registries) should combine to lead many people—encouraged by their families, their physicians, and the media—to overcome the factors that currently inhibit organ donation. In addition, families should be less likely to attempt to override a deceased’s decision to donate if they understand it as a binding portion of an “insurance” arrangement, based on reciprocity.¹⁵¹ Relying purely on altruism for organ donations would certainly be ideal, but it is

146. Still, a chance at a free transplant would begin to alleviate the inequality for those otherwise neglected by the system. See *supra* text accompanying notes 34-36.

147. See Finkel, *supra* note 85; Goyal et al., *supra* note 81; Rohter, *supra* note 85; Christian Williams, Note, *Combating the Problems of Human Rights Abuses and Inadequate Organ Supply Through Presumed Donative Consent*, 26 CASE W. RES. J. INT’L L. 315, 321-27 (1994); see also *DIRTY PRETTY THINGS* (Miramax 2003) (illustrating the tragedy in the black market in organ sales).

148. See Robert A. Sells, *Donation: Will the Principle of “Do As You Would Be Done By” Be Enough?*, 70 TRANSPLANTATION 703, 703 (2000).

149. See Oz et al., *supra* note 57, at 394.

150. See Gubernatis & Kliemt, *supra* note 96, at 700-01.

151. See Siminoff & Chillig, *supra* note 15, at 35.

USING RECIPROCITY TO MOTIVATE ORGAN DONATIONS

not worth the loss of thousands of lives annually.

BOOK REVIEW

Sex, Fear, and Public Health Policy

John G. Culhane, J.D.*

Gay Bathhouses and Public Health Policy. Edited by William J. Woods & Diane Binson. New York: Harrington Park Press, 2003. Pp. 253.

Looking into the AIDS abyss in the mid-1980s, public health officials sometimes succumbed to the same impulses—notably, panic and scapegoating—that activated politicians, judges, and the public itself. Among the best-known results of these impulses were city-by-city efforts to shut down gay bathhouses. No one disputed that sexual activity went on in the bathhouses, but it was—and remains—unclear whether closing them would help stop the transmission of HIV, hinder that effort, or have no net effect. *Gay Bathhouses and Public Policy*,¹ a collection of essays on this topic, comes two decades after the hardest-fought bathhouse closure battles. William J. Woods and Diane Binson, the book's editors (and contributors), have skillfully amassed a group of works that provides a mix of historical depth, reportorial analysis, statistical research, and legal background to the battle over the bathhouses. The authors' stated purpose is to fill a void in knowledge, information, and understanding of the bathhouse question. The bathhouse wars are thereby given historical and cultural context that is perhaps only possible twenty years after these battles were conducted.

In this mission, the book succeeds. The volume, simultaneously published as two issues of the *Journal of Homosexuality*, collects legal, public health, and reportorial papers about the controversy over gay bathhouses and their role in the prevention or spread of HIV.

* Professor of Law, Widener University School of Law; Lecturer, Yale University School of Public Health. This Book Review is dedicated to my public health law students, both at Widener and at Yale.

1. GAY BATHHOUSES AND PUBLIC HEALTH POLICY (William J. Woods & Diane Binson eds., 2003) [hereinafter GAY BATHHOUSES].

Two of the essays foreground the discussion of HIV transmission in bathhouses. One deals with the history of gay bathhouses² while the other, the volume's most compelling read, provides a detailed account of the social, political, and legal battle to close the San Francisco bathhouses at the height of AIDS hysteria in 1984.³ Later in the book, this battle is brought back to life through reprints of two articles from a San Francisco gay monthly. These accounts, which first ran in 1984, were delivered by journalists whose news-gathering techniques included participation in the sexual culture they were describing—in the bathhouses⁴ and in other commercial settings in which sex between men took place.⁵ Indeed, the editors of *Gay Bathhouses* note that the "spark" for the book was the idea to simply reprint these two articles.⁶ But the volume expanded as noted above, and that expansion was broad enough to take in two final essays examining behaviors and interventions in bathhouses today, when the horror of certain death from AIDS has receded enough to enable sober discussion.

In its overall impact, *Gay Bathhouses and Public Policy* supports the conclusion that, although time and distance can impart rationality and depth to the discussion of charged public health issues, sensible solutions and approaches will remain elusive. Indeed, public health law and policy are replete with instances where initiatives that could save lives are swallowed whole by the scapegoating and sloganeering logic of politics. A commonly cited example of this phenomenon involves needle exchange programs: Despite clear and consistent evidence that such programs both reduce the incidence of disease transmission and provide good opportunities for addiction intervention and treatment,⁷ both Congress and the executive branch have refused to support their funding.⁸ Worse,

2. Allan Bérubé, *The History of Gay Bathhouses*, in GAY BATHHOUSES, *supra* note 1, at 33.

3. Christopher Dismar, *The San Francisco Bathhouse Battles of 1984: Civil Liberties, AIDS Risk, and Shifts in Health Policy*, in GAY BATHHOUSES, *supra* note 1, at 71.

4. Michael Helquist & Rick Osmon, *Sex and the Baths: A Not-So-Secret Report*, in GAY BATHHOUSES, *supra* note 1, at 153.

5. Michael Helquist & Rick Osmon, *Beyond the Baths: The Other Sex Businesses*, in GAY BATHHOUSES, *supra* note 1, at 177.

6. William J. Woods & Diane Binson, *Public Health Policy and Gay Bathhouses*, in GAY BATHHOUSES, *supra* note 1, at 1, 7.

7. By this time, the effectiveness of well designed needle-exchange programs is beyond reasonable doubt. For an article citing a few of the many studies on this point, see *Needle-Exchange Programs Are Slowly Finding Greater Acceptance*, AIDS ALERT (Am. Health Consultants), June 1, 2002, at 69 [hereinafter *Needle-Exchange Programs*].

8. During the Clinton Administration, the Secretary of Health and Human Services

they have often spun the tragedy of drug abuse into partisan gold by fretting about the signal of government acceptance that such programs supposedly send.⁹

The spin is “logical”—in a perverse sense—because the direct benefits of such public health programs most often go to minority groups: sexual, racial, and economic.¹⁰ It is, in the short run, cheaper and easier to blame the victims than to engage in the more complex task of selling a policy whose broader societal benefits, including lowered incidence and prevalence of serious diseases, will be realized only over the long term.

The needle exchange illustration shows that even in cases where the rational public health arguments all come down on one side, expediency is sometimes prioritized over good policy decisions. The bathhouse issue, by contrast, is *not* simple. Because our coarse political discourse does not handle nuance well, needed debate sputters and often stalls. Moreover, bringing difficult issues into the open risks their immediate conversion into politically expedient sound bites. But as the issues gain some distance from the eyes of political storms, activists and scholars become less reluctant to talk honestly about problems and limitations on all sides of a debate.

Through its aggregative approach, *Gay Bathhouses* implicitly makes this

did acknowledge the many benefits of such programs, but the Administration did not recommend funding them at the national level. Editorial, *Widen the War on AIDS*, L.A. TIMES, July 3, 1998, at B8. The message, shaped by political reality, was that such programs should be funded at the local level. Many cities do have such programs today, *see* Mike Chalmers, *Needle-Exchange Programs Urged for Wilmington*, NEWS J., Aug. 3, 2004, at A17, but the lack of a clear and consistent national commitment to the needle exchange initiative has left such programs without a strong rhetorical anchor; too often, this translates into a lack of funding. The Bush Administration opposes such programs, Nicole Foy, *Would Needle Exchanges Fly in S.A.?*, SAN ANTONIO EXPRESS-NEWS, May 13, 2004, at 1A, and Congress has gone so far as to ban funding for needle-exchange programs in the District of Columbia, Spencer S. Hsu, *Republicans Revive D.C. Voucher Plan*, WASH. POST, Nov. 17, 2003, at B1.

9. *See* Kathy Kiely, *AIDS Activists Storm Office*, DAILY NEWS (New York), July 21, 1998, at 16 (noting that many in Congress oppose needle-exchange programs because “they send the wrong signal”).

10. The benefits are often, but not always, most apparent to those at the margins. An important exception is the decades-old commitment to immunization of children against infectious diseases that once killed, injured, or seriously disabled millions. *See* John G. Culhane, *Tort, Compensation, and Two Kinds of Justice*, 55 RUTGERS L. REV. 1027, 1092 (2003). Other examples, such as the need for sanitation, are by now so much a part of the landscape that they are taken for granted in the absence of a crisis. Moreover, good public health outcomes ultimately redound to the benefit of all.

point about the value of perspective. Reading the essays in the order in which they were presented, I was able to appreciate fully the cumulative effect of this volume—part history, part ethnographic study, part political journal. Doing so allowed me to appreciate the book's final line: "Given the solid position that bathhouses hold within gay sexual cultures, there is a compelling obligation to understand them and to use these unique environments to promote health and safety among their patrons."¹¹ Yet the overall effect of this book is to produce a sense of the frustration born of complexity. Even though sober discussion has by now become possible, solutions remain elusive. The intractable difficulties of predicting and affecting human behavior—and then pursuing public health policies consistent with any conclusions reached—may be greatest where sexual desire and impulse collide with public disapproval and the reality of disease and mortality. No one intelligent and thoughtful enough to complete this sometimes disturbing collection of readings could wholly subscribe to either of the extreme positions that defined the bathhouse debate twenty years ago when the first signs of HIV infection heralded certain death. The harder question is always what to do in the face of such intractable uncertainty.

The polar positions are easy enough to state. Those who favored closing the bathhouses—including then-Mayor of San Francisco (and now U.S. Senator) Dianne Feinstein—simply took the position that the sexual practices carried on in the bathhouses led to the transmission of HIV, and the bathhouses must therefore be closed. Christopher Disman's account of the San Francisco "bathhouse wars" is the book's most compelling story. As he relentlessly establishes in *The San Francisco Bathhouse Battles of 1984: Civil Liberties, AIDS Risk, and Shifts in Health Policy*,¹² evidence that called into question the prevalence of "unsafe sex" in the bathhouses was not honestly assessed,¹³ nor was much thought given to the question of whether such practices would simply move elsewhere in the absence of these institutions.¹⁴ Indeed, the centripetal political forces were strong enough to effect the conversion of San Francisco's Public Health Director, Dr. Mervyn Silverman, from a position opposing the closure of the bathhouses

11. Matt G. Mutchler, *Comparing Sexual Behavioral Patterns Between Two Bathhouses: Implications for HIV Prevention Intervention Policy*, in GAY BATHHOUSES, *supra* note 1, at 221, 240.

12. See Disman, *supra* note 3.

13. *Id.* at 90-91, 99.

14. *Id.* at 98.

to one favoring it.¹⁵ Those familiar with the routine capitulation of the Centers for Disease Control and Prevention (CDC) to the executive and legislative branches of government will hardly find Silverman's change of heart surprising,¹⁶ but Disman's nuanced account makes clear that Silverman's conversion is not so easily or neatly explained. He may have honestly come to believe that the impossibility of regulating behavior within the bathhouses made closing them the only workable solution.

On the other side of the bathhouse debate is the following rote account of the role of these institutions and the projected effect of their forced closure: Bathhouses provide a safe space for the expression of gay sexuality that is otherwise consigned to such places as outdoor spaces and public restrooms.¹⁷ Since patrons are in a more welcoming and friendly place, they may be more receptive to interventions—such as condom use, other safe sex practices, and HIV testing—than would be possible, let alone practical, in less controlled settings.¹⁸ Without the bathhouses, the sexual conduct will simply disperse to the places it had been previously.¹⁹ HIV transmission will therefore increase and public health will suffer. This cluster of related arguments, while facially plausible, is difficult to prove or disprove empirically.

That said, the book's final essay, *Comparing Sexual Behavioral Patterns Between Two Bathhouses: Implications for HIV Prevention Intervention Policy*,²⁰ suggests that the "bathhouse means prevention" argument is no more convincing than its counterpart. Although one should not draw any solid conclusions from a short-term study of behavior in only two bathhouses—a caveat the authors themselves express²¹—it is impossible to tell the "bathhouses mean prevention story" with the same confidence after reading this essay. In fact, one of the bathhouses studied by this consortium of local and federal public health specialists could be

15. *Id.* at 79-109.

16. For example, the CDC has been known to buckle to political pressure in removing links to websites that right-wing organizations, such as Focus on the Family, find objectionable. See Online Policy Group, Action Alert: Urge CDC and USDA To Provide Same-Sex Info to Youth, at <http://www.onlinepolicy.org/action/cdcusdaalert.shtml> (last visited Sept. 13, 2004).

17. Bérubé, *supra* note 2, at 35-37.

18. See Freya Spielberg, *Designing an HIV Counseling and Testing Program for Bathhouses: The Seattle Experience with Strategies To Improve Acceptability*, in GAY BATHHOUSES, *supra* note 1, at 203.

19. Bérubé, *supra* note 2, at 49.

20. Mutchler, *supra* note 11, at 221.

21. *Id.* at 240.

characterized as a mecca for the transmission of infectious disease.²² In short, *Gay Bathhouses* makes clear that the two extreme accounts of the risks and benefits of bathhouses are too simple. The remainder of this Review focuses on the difficulty of the policy choices facing public health officials and judges, as well as the owners and patrons of the bathhouses. As with other tough public health issues, though, government usually has the last word.

Disman's account highlights Feinstein's position on the bathhouses, which was a matter of record: "My own opinion is that if this was a heterosexual problem, they would have been closed."²³ The connection this position bore to her purportedly negative attitudes towards matters of men's sexuality and sexual creativity is unclear,²⁴ but certainly many of those who favored closing the bathhouses were influenced by the "ick" factor—the equation of "distaste with immorality"²⁵—a tendency particularly prevalent in discussions about homosexual sex. And Feinstein certainly had plenty of cover from the gay community itself; as Disman reminds us, some gay activists feared that if the AIDS epidemic broke out widely in the heterosexual population, failure to close the bathhouses would make it easier to blame the gay "lifestyle."²⁶

Given the epidemiology of the disease and the undisputed higher risk that those engaging in anal as opposed to vaginal sex will contract it, blaming the gay community for the HIV epidemic was likely inevitable whatever the fate of the bathhouses. But because the bathhouses—which, according to Dr. Silverman's estimates, were frequented by only five to ten percent of the gay male community during the mid-1980s²⁷—are such powerful cultural and political signifiers, debates about public policy toward them assume a disproportionate significance. Yet the San Francisco debate simplified bathhouse culture in a way that ignored the bathhouses' important and beneficial role in the gay community. Under this view, it was

22. For a fuller discussion of the point, see *infra* notes 62-67 and accompanying text.

23. Disman, *supra* note 3, at 90 (quoting Larry Liebert & Hsu, *Feinstein Would Shut Bathhouses*, S.F. CHRON., Apr. 5, 1984).

24. See Disman, *supra* note 3, at 90.

25. Arielle Goldhammer, *A Case Against Consensual Crimes: Why the Law Should Stay Out of Pocketbooks, Bedrooms, and Medicine Cabinets*, 41 BRANDEIS L.J. 237, 241 (2002) (quoting Michael Nava & Robert Dawidoff, *Created Equal: Why Gay Rights Matter to America*, in *TAKING SIDES: CLASHING VIEWS ON CONTROVERSIAL MORAL ISSUES* 165, 170 (Stephen Satris ed., 6th ed. 1998)).

26. Disman, *supra* note 3, at 89 (discussing letter from gay author Frank Robinson to Public Health Director Silverman).

27. *Id.* at 77.

all about sex.

Allan Bérubé's *The History of Gay Bathhouses*²⁸ provides a helpful corrective to this narrowing impulse. The origin of the gay bathhouse around a century ago (as the evolutionary offspring of Turkish baths, public baths, and spas) offered a nervous and nervy "contradict[ion to] these stigmas" of gays as criminally diseased sinners "and gave Gay Americans a sense of pride in themselves and their sexuality."²⁹ Admittedly, the point about stigma is arguable; after all, acting in illicit places may reinforce the very stigma patrons are trying to overcome. Nonetheless, it is almost certainly true that the bathhouses were useful to gay men, at least when they were not being raided. As Bérubé notes, the bathhouses offered a sort of democratic camaraderie,³⁰ an important zone of privacy, and a relatively comfortable social environment. Each of these was valuable in a society that was many decades away from its still-grudging willingness to "see" gay people, but the bathhouses' chief benefit was their safety: Compared to public parks, with their potential for injury and death, not to mention blackmail, the bathhouses must have seemed like home. Raids were always possible, but, depending on the bathhouse, these were less of a threat than the arrests for public sex that had been a constant fear.³¹ Bérubé notes that establishments catering to the "best citizens" were often left alone.³²

Further, the bathhouse culture was in a constant state of evolution. At least in the meccas of New York City and San Francisco, gay sexuality had taken hold in other commercial establishments in the years immediately

28. Bérubé, *supra* note 2.

29. *Id.* at 34.

30. Nakedness is democratic in one sense because class distinctions diminish or evaporate, but it substitutes a hierarchy of its own based on the relative beauty of bodies. This point is well articulated in PATRICK MOORE, *BEYOND SHAME* 32-33 (2004).

31. *Id.* at 37.

32. *Id.* The ever-present possibility of raids extended to all establishments frequented by gays, whether or not the patrons were actually looking for sex. *Id.* at 41 (noting that one goal of the periodic anti-bath and anti-bar campaigns included preventing gay men and women from having a place to socialize). For a literary account of the socializing, sex, and fear that were the lot of the gay underground in 1949, see ETHAN MORDDEN, *HOW LONG HAS THIS BEEN GOING ON?* 3-91 (1995). This fear of the gay "other" continues to influence every issue from gay marriage to judicial disputes about child custody and visitation. In one case, for example, a trial judge forbade a mother to "expose" her children to anyone "known by [her] to be lesbian"—whether or not this person had any sexual or romantic tie to the mother. *DeLong v. DeLong*, No. WD 52726, 1998 Mo. App. LEXIS 69, at *8 (Mo. Ct. App. Jan. 20, 1998), *superceded by* *J.A.D v. F.J.D.*, 978 S.W.2d 336 (Mo. 1998).

preceding the AIDS crisis.³³ Given the drift toward greater acceptance, Feinstein's crusade—which involved sending undercover officers to bathhouses and then misleadingly reporting what they had found³⁴—might not have caught fire in a less combustible situation. But AIDS was decimating the gay male population of San Francisco, and the ravaged wraiths awaiting death were a constant, terrible reminder of the disease's toll. While people were beginning to understand that different sexual behaviors carried different levels of risk,³⁵ lack of confidence in the science of transmission occluded the debate and enabled Silverman to complete his 180 degree turn, made official by this statement: “[A]ll sexual activity between individuals [is to] be eliminated in public facilities in San Francisco where the transmission of AIDS is likely to occur.”³⁶

This statement makes no sense on its face; the transmission of AIDS (more precisely, HIV) is “likely to occur” only in the presence of specific sexual conduct, so if *those* risks—certainly less than all sexual conduct—are eliminated, no transmission will occur. Silverman's proclamation can only be understood in a non-contradictory way by assuming that the bathhouses *themselves* are responsible for transmission. So by this time the public health community had gotten behind the reductive idea that bathhouses were all about sex and that *they* were the problem. Shortly after this statement, Silverman declared that the bathhouses were public nuisances³⁷ and ordered their closure. Thus was a difficult public health problem “solved” by fiat.

Nonetheless, Disman points out that the often-reported story that the San Francisco bathhouses were closed by court order is false. In fact, while the city was able to obtain a temporary restraining order forcing the

33. See MOORE, *supra* note 30 (discussing bathhouses, sex clubs, and dance clubs that permitted sexual conduct that flourished during this time).

34. Disman, *supra* note 3, at 106-07 (discussing Silverman's generic and outraged description of every imaginable and “unimaginable” sexual activity even though actual evidence was less dramatic).

35. See *id.* at 97 (noting that suggestions for baths “failed to mention AIDS-risk levels”).

36. *Id.* at 90.

37. States and cities have the authority to declare anything that injures or threatens the public health, safety, and welfare a public nuisance. See John G. Culhane & Jean Macchiaroli Eggen, *Defining a Proper Role for Public Nuisance Law in Municipal Suits Against Gun Sellers: Beyond Rhetoric and Expedience*, 52 S.C. L. REV. 287, 297 (2001). But the position must be defensible in a court of law; the city's public nuisance claim against the bathhouses was weak and only partially successful. See *infra* text accompanying notes 38-39. In fact, the bathhouses were *not* closed (except briefly) by the courts.

bathhouses to close for fifteen days,³⁸ ultimately they were permitted to remain open under two sets of progressively more restrictive rules.³⁹ The story has been retold as ending with the court ordering the baths closed, in part because the city won the most important battle: the right to decide what counted as high-risk sex. But recall that Silverman's definition made no effort to distinguish between risk levels and left out sex between any pairing other than two males.

Although the bathhouse owners achieved only a limited legal victory—Pyrrhic by any measure, since the San Francisco bathhouses eventually closed under the unworkable constraints imposed⁴⁰—courts are typically even more deferential to governmental actions defined as public health measures. While the ability to second-guess public health decisions may be, in the words of one Australian observer “an extremely American process,”⁴¹ it is rarely invoked. Scott Burris, an accomplished law and public health scholar, makes the point matter-of-factly in his contribution, which surveys the bathhouse litigation from 1984 through 1995.⁴² While only eight such cases were reported during that period, Burris notes that the routine victories achieved by the public health community (seven of eight cases were winners, at least in substantial part) over the establishments they sought to close likely discouraged other potential litigants who closed without even trying to fight.⁴³

Judicial deference to public health officials, particularly in the case of epidemics (real or asserted) is not new to the bathhouse controversy. A staple case of any public health law course is the U.S. Supreme Court's decision in the century-old *Jacobson v. Massachusetts*,⁴⁴ in which a city ordinance requiring all adults to be vaccinated against smallpox because of

38. Disman, *supra* note 3, at 110.

39. *Id.* at 112-15. The first ruling “focused . . . on pragmatic ways to prohibit high-risk sex in the businesses.” *Id.* at 112. About a month later, a modified injunction placed final authority for defining high-risk sex with the director of the public health department. *Id.* at 114-15.

40. *See id.* at 116.

41. *Id.* at 113 (quoting Dennis Altman). The statement reflects a foreigner's incredulity at the American focus on individual rights. While such rhetoric frames judicial decisions that weigh policies by the public health authority against personal liberties, in practice courts are quite biased in favor of public health officials. *See infra* notes 48-58 and accompanying text.

42. Scott Burris, *Legal Aspects of Regulating Bathhouses: Cases From 1984 to 1995*, in GAY BATHHOUSES, *supra* note 1, at 131.

43. *Id.* at 134.

44. 197 U.S. 11 (1905).

increased prevalence of the disease was upheld against a liberty-based challenge.⁴⁵ While the court's deference to the public health authority's discretion was unsurprising—and is still good law⁴⁶—what is perhaps shocking to a rights-schooled reader is the Court's endorsement of this statement from a then-recent New York court decision, also involving smallpox vaccination: "A common belief, like common knowledge, does not require evidence to establish its existence, but may be acted upon without proof by the legislature and the courts. . . . [F]or what the people believe is for the common welfare must be accepted as tending to promote [it], whether it does . . . or not."⁴⁷

Although few would likely support such an abdication today, statements endorsing broad discretion for those charged with protecting public health continue to be articulated, and the heat generated by the AIDS crisis precluded a more balanced judicial approach. Consider this language from one of Burris's cited bathhouse cases, again from New York: "It is not for the courts to determine which scientific view is correct in ruling upon whether the police power has been properly exercised. 'The judicial function is exhausted with the discovery that the relation between means and end is not wholly vain and fanciful'"⁴⁸ As Burris points out, part of this mortifying deference has to do with "practical and doctrinal limitations on the role of courts."⁴⁹ One of the most significant achievements of his contribution is the nuanced connection he draws between this modesty—which surely has some logic to recommend it, in view of the serious consequences of erring on the side of keeping the bathhouses open—and judicial attitudes about sex and the status of sexual outliers. Define the constitutional right implicated as limited to *private* sexual conduct, and bathhouse sex as *public* conduct, and the patrons disappear as rights holders. Even private peep shows have been defined, without analysis, as public.⁵⁰ A more fully articulated approach, such as one that might be derived from looking at the physical, social, and environmental aspects of the bathhouses—as suggested by Woods and

45. *Id.* at 37-39.

46. *See, e.g.,* Stenberg v. Carhart, 530 U.S. 914, 970-72 (2000) (Kennedy, J., dissenting) (citing *Jacobson* in late-term abortion case for proposition that legislative determinations of public health policy are determinative absent some indication of improper motive). As Burris asserts throughout his essay the balance of power remains with public health.

47. *Jacobson*, 197 U.S. at 35 (quoting *Viemeister v. White*, 72 N.E. 97, 97 (N.Y. 1904)).

48. *City of New York v. New Saint Mark's Baths*, 130 Misc. 2d 911, 917 (N.Y. Sup. Ct. 1986) (quoting *Williams v. Mayor of Baltimore*, 289 U.S. 36, 42 (1933)).

49. Burris, *supra* note 42, at 138.

50. *Id.* at 144.

Binson in *A Theoretical Approach to Bathhouse Environments*⁵¹—is beyond the ken of most courts. They do not want to consider the possibility that bathhouse sex may be neither fully public nor private, nor do they want to second-guess the public health community's decisions about risk.

Why, though? Courts routinely make all kinds of difficult decisions and have brazenly gotten involved in everything from the management of prisons to the details of school busing.⁵² But straying from their comfort zone—legal analysis—requires a Herculean effort that courts take on only rarely and often reluctantly. Whether sex is public or private may be garbed as a question amenable to the kind of multi-factor balancing that makes courts comfortable, but the question taps into a deeper vein of disquiet. As Burris notes, the confluence of a public health crisis, societal squeamishness about sex—especially “non vanilla” sex—and the alloy created by fusing the stigma of gay identity to perceived public harm operates to keep courts mostly on the sidelines.⁵³

Rare exceptions prove the rule. In *Jew Ho v. Williamson*,⁵⁴ the attempt to quarantine a section of San Francisco to prevent the spread of bubonic plague was so clearly an act of discrimination that the public health charade was removed. The boundaries of the quarantine area zigzagged,⁵⁵ and it was enforced in a way likely to spread—rather than contain—the plague. Most significantly, it was enforced only against Chinese Americans.⁵⁶ Even here, the court was unwilling to second-guess the public health authority's finding that there was indeed plague (despite evidence to the contrary); it was only public health's inability to explain such counterproductive policies that could not be overlooked.⁵⁷ Usually, the cases are harder, so courts find it easier to defer to the public health authority's decisions. For example, the resurgent tuberculosis epidemic in the early 1990s led a trial judge in New Jersey to rule (in an unusually thoughtful and careful decision) that a man with infectious tuberculosis could be involuntarily confined if he refused to take medication that would

51. Diane Binson & William J. Woods, *A Theoretical Approach to Bathhouse Environments*, in *GAY BATHHOUSES*, *supra* note 1, at 23.

52. See Barbara E. Armacost, *Affirmative Duties, Systemic Harms, and the Due Process Clause*, 94 MICH. L. REV. 982, 1006-07 (1996) (citing these and other examples of cases where courts are criticized for operating beyond their institutional competence to resolve problems that are polycentric rather than binary).

53. Burris, *supra* note 42, at 137-39.

54. 103 F. 10 (N.D. Cal. 1900).

55. See *id.* at 23.

56. *Id.* at 23-24.

57. *Id.* at 26.

eliminate the risk to those with whom he came into contact.⁵⁸

Again, the bathhouse issue was and remains difficult. It remains impossible—even today—to know definitively what policy toward bathhouses will yield the best public health outcomes. With HIV now a chronic but manageable disease—at least in the United States and for people who have access to the best treatment—a more careful analysis may at last be possible. And it is in everyone's interest to engage in this more challenging project. Seen as a group, patrons have their health interests at stake in the best policy; bathhouse owners have a financial stake in the outcome and cannot risk bringing the power of the public health community down on them; and the public health community has its already depleted moral authority to defend. Legal coercion is possible, but widely and correctly viewed as a last resort. Where public health authorities can get “buy-in” from all constituents, they can obviate expensive and ham-handed measures and can help repair the trust that governmental policies (not always, but sometimes, those of public health officials) have damaged—especially in minority communities.⁵⁹

One size probably will not fit all. On-site HIV testing appears to have promise as one compromise measure. Based on a successful HIV testing program in Seattle, Freya Spielberg and her co-authors offer useful suggestions for achieving better design for HIV testing at bathhouses.⁶⁰ The program faced, and largely overcame, obstacles involving: owner and patron reluctance (patrons eventually saw the availability of testing as a convenience, and owners became convinced the idea was sound); space limitation and training problems; and patrons' frequent failure to return to pick up their results (an issue largely mitigated by the advent of tests that provide “while you wait” results).⁶¹ As the high number of HIV positive persons unaware of their status attests, any opportunity for such testing should not be squandered. But is the Seattle experience the authors describe transferable to other places?

The book's final essay raises hard questions like these without

58. *City of Newark v. J.S.*, 652 A.2d 265 (N.J. Super. Ct. Law Div. 1993).

59. The most infamous example of public health's own mistreatment of minority groups is the CDC-supported Tuskegee study of the course of syphilis infection in African-Americans, who were neither told of the study nor offered antibiotics from the early 1930s until 1972. See Allan M. Brandt, *Racism and Research: The Case of the Tuskegee Syphilis Study*, 8 HASTINGS CENTER REP. 21-29 (1978), reprinted in LAWRENCE O. GOSTIN, PUBLIC HEALTH LAW AND ETHICS: A READER 312-19 (2002).

60. Spielberg, *supra* note 18, at 203.

61. *Id.* at 207, 209 (patron reluctance); *id.* at 208 (space/training limitations); *id.* at 211 (failure to pick up results); *id.* at 215-16 (new testing).

answering them. The article is a needed counterweight to the relentless—and mostly justified—criticism of the public health community that lashes the rest of the volume together.

Standard wisdom has it that groups historically mistreated by governmental policy and its often-unfair application are the least likely to heed public health messages. In the HIV context, the demimonde of black men living on the “down low”—participating in a subculture marked by sex with other men, but also by having girlfriends or wives who are unaware of their partners’ conduct and the risk it creates for them⁶²—is often used as Exhibit A in the effort to illustrate the difficulty of reaching marginalized groups with public health messages. So one might expect that a study of two bathhouses, one frequented by young, mostly white men (Bathhouse “A”), and the other, by a more ethnically diverse mix with most of the patrons either African-American or Latino (Bathhouse “B”), would show greater condom use (a marker for trust in public health) among the first group.

In fact, just the opposite turned out to be the case. Bathhouse “A” patrons were likely to engage in even the riskiest behavior without using condoms, while those frequenting Bathhouse “B” “tended to state that they always use condoms for anal sex and none said that they . . . never used them.”⁶³ Other factors may help to explain this unexpected result: Bathhouse “A” residents were likelier to have used drugs such as ecstasy or crystal methamphetamine that can either increase sexual drive or cloud judgment; they tended to be younger, with more “beautiful” bodies, perhaps suggesting a perception of immortality.⁶⁴ The description of Bathhouse “B”—while it did mention alcohol use among some patrons—focused more on meeting “regular guys” and on watching erotic videos.⁶⁵ Oral sex was more prevalent than anal sex.⁶⁶ Interestingly, many patrons of Bathhouse B did fit the “down low” description, yet they avoided at least the riskiest conduct.⁶⁷

Despite the obvious limitations of such a small study, these findings do

62. See Benoit Denizet-Lewis, *Double Lives on the Down Low*, N.Y. TIMES, Aug. 3, 2003, at 6-28; see also JACOB LEVENSON, THE SECRET EPIDEMIC 54-55 (2004).

63. Mutchler, *supra* note 11, at 234.

64. *Id.* at 232-33.

65. *Id.* at 233-35.

66. *Id.* at 235.

67. The authors make the important point that oral sex, while posing a low risk for the transmission of HIV, does create a high risk of transmitting other STDs. Therefore, female partners of men on the “down low” are still in peril. *Id.* at 238.

suggest the need for further questioning of our assumptions about the best intervention and prevention policies. Their findings are a sobering warning against policies not steeped in careful attention to facts “on the ground.” Such a warning should be applied not only to bathhouse policy, but to public health issues generally.

Of course, sound public health policy is elusive. The needle-exchange example shows that even simple questions can receive the wrong answer. The bathhouse issue is more complex, so both sides can offer plausible arguments for their positions. As the public tried to absorb the unfolding horror of the AIDS crisis, rational arguments did not stand a chance of receiving a fair hearing, and it was inevitable that the advocates of prohibition would prevail. At times, the reader of *Gay Bathhouses* feels a sense of pessimism about the likelihood that good policy is even possible. That conclusion is perhaps too gloomy. Although time and reflection do not make the complexities of the issue disappear, the overall impression left by *Gay Bathhouses* is that good policy choices can eventually emerge.

CASE STUDY

QUESTION:

How can the current state of medical malpractice insurance be improved?

RESPONSES:

- 343 Effective Legal Reform and the Malpractice Insurance Crisis**
Richard E. Anderson, M.D.
- 357 Debunking Medical Malpractice Myths: Unraveling the False Premises Behind “Tort Reform”**
Geoff Boehm, J.D.
- 371 Health Care in Crisis: The Need for Medical Liability Reform**
Donald J. Palmisano, M.D., J.D.
- 385 Turning from Damage Caps to Information Disclosure: An Alternative to Tort Reform**
Kathryn Zeiler, M.S., J.D., Ph.D.

Effective Legal Reform and the Malpractice Insurance Crisis

Richard E. Anderson, M.D.*

This Case Study is built around two fundamental questions: First, is there really a malpractice insurance crisis in the United States today? Second, what is the best way to improve the medical liability system? While there is much ongoing debate, this Case Study argues that the answers to both questions are clear. I first review the nature, breadth, and source of the current crisis and then examine ways to ameliorate the problems in both the short and long-term. There is clear evidence that current problems are the result of a dramatic increase in the cost of litigation and that certain legal reforms would significantly alleviate the crisis.

I. THE MALPRACTICE INSURANCE CRISIS

A crisis is defined as “an unstable or crucial time or state of affairs in which a decisive change is impending; especially one with the distinct possibility of a highly undesirable outcome.”¹ The American Medical Association (AMA) has found this definition to be an apt description of the medical malpractice insurance situation in an increasing number of U.S. states:

America’s patients are losing access to care because the nation’s out-of-control legal system is forcing physicians in some areas of the country to retire early, relocate or give up performing high-risk medical procedures. There are now 20 states in a full-blown medical liability crisis—up from 12 in 2002. In crisis states, patients continue to lose access to care. In some states, obstetricians and rural family physicians no longer deliver babies. Meanwhile, high-risk specialists no longer provide trauma care or perform complicated surgical procedures.²

* Fellow of the American College of Physicians, Chairman and Chief Executive Officer of The Doctors Company.

1. MERRIAM WEBSTER’S COLLEGIATE DICTIONARY 275 (10th ed. 1995).

2. Am. Med. Ass’n, *America’s Medical Liability Crisis: A National View* (2004), at

The actions of protesting doctors—from selectively withholding medical services to marching on state capitols to demand legal reform—have also made it clear that we are in the midst of a crisis.³ In turn, multiple state legislatures,⁴ the United States Congress,⁵ and the media⁶ have turned their attention to medical malpractice, frequently concluding that increasing insurance rates represent an urgent concern that must be addressed. Important legislative action has been taken in states as diverse as Texas, Florida, and Idaho.⁷ Many other states are actively debating the issue, but no legislation has resulted.⁸ The fact that so many legislatures are simultaneously and independently discussing the malpractice insurance crisis attests to its urgency.

The current state of medical malpractice insurance has been precipitated by a sharp rise in the cost of malpractice claims—both due to the increasing volume of malpractice litigation and to the growing size of awards. This increase in the cost of claims has resulted in a dramatic rise in the cost of malpractice premiums. In 2002, malpractice insurance rates for physicians nationwide rose approximately twenty percent,⁹ but this average figure obscures a very wide range. States like California that enjoy effective legal reforms have seen rates increase only a few percent per year in this interval,¹⁰ while states lacking such reforms have seen increases in excess of

<http://www.ama-assn.org/ama/noindex/category/11871.html>.

3. Tom Ramstack, *Doctors' Careers on Life Support*, WASH. TIMES, Sept. 1, 2003 at A1.

4. AM. TORT REFORM ASS'N, TORT REFORM RECORD 2-3 (July 2004), http://www.atra.org/files.cgi/7802_Record6-04.pdf

5. Help Efficient, Accessible, Low-Cost, Timely Health Care (HEALTH) Act of 2004, H.R. 4280, 108th Cong. (2004); Pregnancy and Trauma Care Access Protection Act of 2004, S. 2207, 108th Cong. (2004).

6. See, e.g., Jessica Bruder, *New Jersey Starts Fund for Malpractice Costs*, N.Y. TIMES, June 8, 2004, at B5. Research searches using the phrase "malpractice crisis" return a multitude of articles, reports, and other resources. See, e.g., U. S. GEN. ACCOUNTING OFFICE, MEDICAL MALPRACTICE INSURANCE: MULTIPLE FACTORS HAVE CONTRIBUTED TO INCREASED PREMIUM RATES (June 2003), <http://www.gao.gov/new.items/d03702.pdf>; John E. Lemoine, Controlling the Medical Malpractice Crisis: A Letter from the CEO (Jan. 7, 2004), at <http://www.lammico.com/letter/addendum-detail.asp?addendum=17>; Casualty Actuarial Soc'y, The Medical Malpractice Crises: The Causes and Remedy, at <http://www.casact.org/media/medmal.htm> (last visited Oct. 12, 2004).

7. See AM. TORT REFORM ASS'N, *supra* note 4, at 2-3.

8. See *id.*

9. 2002 Rate Survey Finds Malpractice Premiums Are Soaring..., MED. LIABILITY MONITOR, Oct. 2002, at I-IV.

10. In fact, since 1976, when the Medical Injury Compensation Reform Act (MICRA) of 1975, chs. 1-2, 1975 Cal. Stat. 2d Ex. Sess. 3949, went into effect, California's rates have

one hundred percent for specialists in high-risk areas of medicine.¹¹ As a result, high-risk physicians in states lacking legal reforms face annual malpractice insurance premiums in excess of \$100,000 and in some cases in excess of \$200,000 per year, per doctor.¹² In the states most directly affected by rising premiums—for example, Mississippi, West Virginia, Nevada, and Pennsylvania—some physicians have found themselves uninsurable at any price or have turned to state-run plans, which are even more expensive than coverage available in the marketplace, as the insurer of last resort.¹³

A. The Underlying Problem: Malpractice Litigation

1. Frequency of Litigation

Even the now commonplace phrase “high-risk specialists” is indicative of this crisis. We used to speak of high-risk patients, referring to individuals with higher than normal risk of unfavorable outcomes, such as neurosurgical patients with spinal cord tumors. Now, we refer to entire medical specialties as high-risk, meaning that they face a much higher than normal risk of litigation. In fact, neurosurgeons practicing in the United States today face, on average, a malpractice claim every two years.¹⁴ For obstetricians, orthopedists, general surgeons, emergency room doctors, and other high-risk specialists, the figure is one claim every three years.¹⁵

More than three quarters of all such claims close without any payment to the plaintiff, but they are extremely costly to defend, averaging nearly \$23,000 per claim.¹⁶ If a case must go all the way through a jury trial before

increased by less than three percent per year. See RICHARD E. ANDERSON, *MEDICAL MALPRACTICE: A PHYSICIAN'S SOURCEBOOK* 214 (2004); see also *infra* notes 40-53 and accompanying text (discussing MICRA).

11. *Trends in 2002 Rates for Physicians' Medical Professional Liability Insurance*, MED. LIABILITY MONITOR, Oct. 2002 (special report), at 1-16.

12. *E.g., id.* at 4.

13. NGA CTR. FOR BEST PRACTICES, *ADDRESSING THE MEDICAL MALPRACTICE INSURANCE CRISIS* 3-4 (2002).

14. Richard E. Anderson, *Defending the Practice of Medicine*, 164 *ARCHIVES INTERNAL MED.* 1173, 1174 (2004).

15. *Id.*

16. *Harming Patient Access to Care: Implications of Excessive Litigation: Hearing Before the Subcomm. on Health of the House Comm. on Energy & Commerce*, 107th Cong. 84 (2002) [hereinafter *Hearing*] (statement of Richard E. Anderson, M.D., F.A.C.P., Chairman/CEO of The Doctors Co., for the Physician Insurers Ass'n of Am.).

a defense verdict, the average expense exceeds \$85,000.¹⁷ These costly victories are important drivers of medical malpractice premium rates.

There are more than 125,000 pending malpractice claims against America's 700,000 licensed physicians today.¹⁸ Thus, if you are reading this Case Study on a weekday, roughly six hundred more malpractice claims will be filed today. This large number of claims is even more striking when you consider that many licensed physicians are in research, academia, the military, or are retired and are thus not at risk of being sued.

2. Fallacy of the Bad Doctor

Faced with this onslaught of litigation, physicians feel that they are under siege.¹⁹ There might be less widespread concern about malpractice claims if they were primarily brought against negligent doctors. The frequency data cited above make clear that virtually all physicians face the prospect of litigation, though most are ultimately vindicated. There is little victory in vindication, however, given the costs, long duration of malpractice claims,²⁰ and the personal attacks on professional identity that are at the core of the malpractice allegations.

In any given year, two percent of claims are responsible for about half of the compensation provided to plaintiffs,²¹ leading some to argue that removing the two percent of doctors responsible for these large claims would eliminate the crisis. However, the two percent of physicians who have to make these payments differ every year. Were this fact not true, other doctors would not risk practicing with them, nor tolerate their negative impact on the profession, and insurance companies certainly would not offer them coverage. In truth, the problem with our current medical liability system is not the presence of a few bad doctors, it is that

17. *Id.* at 86.

18. See Health Care Liab. Alliance, Health Care Lawsuits, Claim Payments on Upswing (Apr. 27, 1995) (on file with author). The 125,000 figure is based on data from 1995 because the number of pending claims has not been tracked since then. The figure of 125,000 represents a conservative estimate of current suits, since the number of physicians practicing has increased significantly, WAYS AND MEANS COMM., HOUSE OF REPS., GREEN BOOK 2003 app. C, C27-28 (2003), and the frequency of litigation has certainly not decreased significantly since 1995, The Doctors Co., Annual Claims Per Mature Internal Medical Equivalent Doctor 1976-2002 (on file with author).

19. See ANDERSON, *supra* note 10, at ix-xv.

20. On average, "it takes 5 ½ years for an insurer to close a malpractice claim after the date of the incident." *Hearing, supra* note 16, at 87.

21. See ANDERSON, *supra* note 10, at 210.

every year a large number of physicians face meritless claims.²²

Why are the doctors involved in large claims different every year? The Harvard Medical Practice Study gives us the answer: There is no relationship between the presence or absence of medical negligence and the outcome of malpractice litigation. The only variable that predicts the outcome of claims is the degree of injury. A severely injured plaintiff is likely to be compensated in court whether or not the doctor was at fault.²³

3. Increasing Size of Claim Awards

While the volume of malpractice litigation alone is sufficient to qualify as a crisis, the cost of the average claim is rising at unprecedented rates. Between 1997 and 2000, the median malpractice award doubled to one million dollars.²⁴ The average (modal) jury verdict in malpractice trials was 3.5 million dollars in 2000.²⁵ In states without legal reforms, the outer limit of liability has skyrocketed to amounts never before seen in medical negligence cases.²⁶ Just under one billion dollars in medical malpractice compensation was paid out in New York and Pennsylvania (combined) in 2000,²⁷ and the total cost of medical malpractice litigation now exceeds twenty-four billion dollars annually and continues to grow.²⁸

22. Steve Ellman, *ABA Blasts Fla. Ballot Measure Limiting Attorney Fees*, MIAMI DAILY BUS. REV. (Oct. 18, 2004); see also Medical Malpractice Lawyers, Medical Malpractice Lawsuit, at <http://www.medmalattorney.us.com/lawsuit.html> (last visited Dec. 4, 2004). While the costs incurred by the medical profession are widespread, few injured patients benefit from the payments. It is worth noting that contingency fee lawyers take home up to forty percent of the awards won by plaintiffs.

23. See Troyen A. Brennan et al., *Relation Between Negligent Adverse Events and the Outcomes of Medical-Malpractice Litigation*, 335 NEW ENG. J. MED. 1963, 1963-67 (1996) (comparing outcomes based on independent assessments of the merit of plaintiffs' claims).

24. Press Release, Jury Verdict Research, Medical Malpractice Verdict and Settlement Study Released (Mar. 22, 2002) (on file with author).

25. Jury Verdict Research, Ins. Info. Inst., Trends in Malpractice Insurance: Behind the Chaos (on file with author).

26. A jury in Texas awarded a \$268 million verdict in a case involving the death of a single patient. *Model's Jury Award Top Verdict*, LAWYERS WEEKLY USA, Jan. 8, 2001, at 1. Pennsylvania has had multiple awards in excess of \$50 million. Christopher Guadagnino, *Malpractice Awards Surge in PA*, PHYSICIAN'S NEWS DIG., Jan. 2001, <http://www.physiciansnews.com/cover/101.html>.

27. PA. MED. SOC'Y, MEDICAL MALPRACTICE SEVERITY, 2000 (2002).

28. TILLINGHAST-TOWERS PERRIN, U.S. TORT COSTS: 2003 UPDATE 13 (2003).

B. Impact on Malpractice Insurance Companies

The rising cost of claims has meant that malpractice insurers will have paid close to \$1.60 for every dollar of premium collected between 2001 and 2003.²⁹ The cost of claims represents nearly eighty percent of an insurer's expenses³⁰ and the nonpartisan United States General Accounting Office (GAO) affirms that "losses on medical malpractice claims appear to be the primary driver of increased premium rates in the long term. Such losses are by far the largest component of insurer costs, and in the long run, premium rates are set at a level designed to cover anticipated costs."³¹

At the same time, falling interest rates between 2000 and 2002 lowered investment returns on premiums and reserves, reducing the subsidization of rates. The decline in investment income, however, accounted for only 7.2% of premium increases according to the GAO, underscoring the magnitude of rising claims costs.³² Moreover, even in a better economic environment, investment income can only be expected, at best, to bridge a small gap between insurance rates and expenses.

II. SOLUTIONS TO THE CRISIS

Some of the factors that have produced this litigation crisis are cultural and can be changed only over long periods of time. One of these is monetary desensitization: From awards on games shows³³ and the salaries of sports figures³⁴ and corporate executives³⁵ to attorneys' fees³⁶ and the

29. See CONNING RESEARCH & CONSULTING INC., MEDICAL MALPRACTICE: ANATOMY OF A CRISIS 2003, at 22 (2003); see also JAMES D. HURLEY, A NEW CRISIS FOR THE MED MAL MARKET?, 4 EMPHASIS (Tillinghast) 2 (2002), http://www.towersperrin.com/tillinghast/publications/publications/emphasis/Emphasis_2002_4/Hurley.pdf. This statistic reflects all medical malpractice insurers, including physician-owned medical practice insurers (termed mutuals and reciprocals). These insurers cover more than sixty percent of the country's doctors, see *Hearing, supra* note 14, at 86, and do not face shareholder pressure for profit.

30. U. S. GEN. ACCOUNTING OFFICE, *supra* note 6, at 16.

31. *Id.* at 43.

32. *Id.* at 27.

33. For example, *Who Wants To Be a Millionaire?* has replaced the *\$64,000 Question*.

34. The shortstop for a last place team signed a ten-year \$252 million contract. See Murray Chass, *Rodriguez Strikes It Rich in Texas*, N.Y. TIMES, Dec. 12, 2000, at D1.

35. In recent years, we have seen billionaires created after two or three years of hard work in the technology industry. MicroSolutions and Broadcast.com founder, Mark Cuban, and eBay founder, Pierre Omidyar, are just two examples. *The Forbes 400, The Richest People in America*, FORBES, Oct. 11, 2004 (Special Issue), at 186, 254.

federal budget,³⁷ we have seen dramatic increases in the amounts of money at stake in society. Hearing about people earning hundreds of millions of dollars or deals totaling billions of dollars is now surprisingly commonplace. In this environment, it is not surprising that a jury that would have awarded one million dollars a few years ago now responds with a ten million dollar verdict, even though actual economic damages in medical malpractice claims have not changed to nearly that extent. This monetary desensitization, in general, and the size of jury verdicts, in particular, has greatly exceeded the inflation rate for the economy as a whole.³⁸ More broadly, Phillip Howard, founder of the legal reform group Common Good, notes:

Fear of litigation has undermined our freedom to make sensible decisions. Doctors, teachers, ministers, even little league coaches, find their daily decisions hampered by legal fear. Our system of justice, long America's greatest pride, is now considered a tool for extortion, not balance. What's missing is the essential idea of law. Law is supposed to set the boundaries of legal action, so that people know where they stand. Law should make us feel comfortable doing what's reasonable and nervous doing what's wrong. Today Americans are nervous doing almost anything.³⁹

Changing this mindset will be difficult and will take a considerable amount of time to happen. Fortunately, we do not have to wait for such a cultural shift to occur on its own; there are several strategies available that will help create meaningful change in the short-term.

A. MICRA and Effective Tort Reform

There is more than a quarter century of experience and an abundance of evidence that the four principal reforms embodied in California's

36. Attorneys regularly now demand billion dollar fees in areas such as tobacco, asbestos, and other mass tort litigation. See, e.g., CTR. FOR LEGAL POL'Y AT THE MANHATTAN INST., TRIAL LAWYERS, INC.: A REPORT ON THE LAWSUIT INDUSTRY IN AMERICA 2003, at 6-7, 10-11 (2003); Lester Brickman, Want To Be a Billionaire? Sue a Tobacco Company, at <http://www.calahouston.org/billion.html> (last visited Oct. 13, 2004).

37. The federal budget is now expressed in trillions of dollars.

38. The average annual inflation rate in the United States between 1990 and 2000 was 2.7%. See North America: Economy: Inflation, at http://www.nationmaster.com/graph-T/eco_inf_199/NAM (last visited Oct. 13, 2004).

39. *Why We Have Come Together*, Common Good, <http://cgood.org/about> (last visited July 9, 2004).

Medical Injury Compensation Reform Act (MICRA) statutes prevent the kind of malpractice insurance crisis we are experiencing today.⁴⁰

MICRA was passed by the California legislature in 1975 under circumstances similar to those described in current headlines. A tidal wave of malpractice litigation in the state drove up insurance rates by several hundred percent, but eventually most insurers in California concluded that the practice of medicine was not an insurable risk and simply refused to provide coverage under any circumstances.⁴¹ Local doctors went on strike, and physicians marched on Sacramento. The legislature responded with MICRA, and California has had a stable insurance environment ever since.

There are four major components to MICRA: First, it provides for a \$250,000 cap on non-economic damages.⁴² This provision is the single most important provision of MICRA. It is critical to note that there is no limit on total awards for actual damages, but capping awards for pain and suffering removes the potential for medical malpractice plaintiffs to be awarded incalculable windfalls. Second, MICRA allows defendants to introduce into evidence additional sources of compensation for injury that have already been paid; this is known as collateral source reform.⁴³ For example, if an injured patient has already had lost wages or medical costs covered by disability or medical insurance, the recovery need not be duplicated. Third, MICRA provides for periodic payments, allowing damage awards to be paid over the time frame they are intended to cover.⁴⁴ This sensible reform permits the insurance system to pay large awards without facing insolvency by taking advantage of the time value of money and assures funds will be available for the patient when needed. Finally, MICRA limits contingency fees by using a sliding scale.⁴⁵ For example, an attorney may keep forty percent of the first \$50,000 of an award, but is limited to \$221,000 (plus expenses) of a one million dollar judgment, meaning an additional \$179,000 actually reaches the injured patient as compared to a state with a straight forty percent contingency fee. Not only is this provision of direct benefit to the injured patient, but it also makes it more difficult

40. Medical Injury Compensation Reform Act (MICRA) of 1975, chs. 1-2, 1975 Cal. Stat. 2d Ex. Sess. 3949.

41. Approximately eighty percent of the malpractice claims filed in California during the twentieth century (up to 1975) were filed between 1970 and 1975. Barry Keene, *California's Medical Malpractice Crisis*, HEALTH CARE LIABILITY ALLIANCE 1 (2003).

42. CAL. CIV. CODE ANN. §3333.2 (West 1997 & Supp. 2004).

43. *Id.* § 3333.1.

44. CAL. CIV. PROC. CODE ANN. § 667.7 (West 1997 & Supp. 2004).

45. CAL. BUS. & PROF. CODE ANN. § 6146 (West 1997 & Supp. 2004).

for attorneys to finance large numbers of non-meritorious cases with the few that they win.⁴⁶

MICRA has reduced California's malpractice premiums by forty percent in constant dollars since 1975. Uncorrected for inflation, this statistic translates into increases in insurance premiums of less than three percent per year,⁴⁷ less than one-third the rate at which premiums have risen nationally.⁴⁸

It is reliably estimated by entities as diverse as the U.S. Congressional Budget Office,⁴⁹ the U.S. Department of Health and Human Services,⁵⁰ Milliman and Robertson,⁵¹ the Florida Governor's Select Task Force on Healthcare Professional Liability Insurance,⁵² and the American Academy of Actuaries⁵³ that passage of reforms similar to MICRA in states currently lacking such statutes would result in premium savings of twenty-five to thirty percent annually.

Not only is there convincing evidence that these reforms are effective when enacted, we have, unfortunately, compelling evidence of the damage that occurs when these reforms are withdrawn. The state of Ohio enacted MICRA-like statutes in 1975.⁵⁴ Malpractice insurance rates in the state fell steadily from 1975 until the law was challenged in 1982, and the Ohio Supreme Court found the statutes to be unconstitutional.⁵⁵ Thereafter, malpractice insurance rates resumed their climb.⁵⁶ Not surprisingly, Ohio is

46. More than three quarters of claims close without payment. *See supra* note 16 and accompanying text.

47. ANDERSON, *supra* note 10, at 214.

48. *See* U.S. DEP'T OF HEALTH & HUMAN SERVS., CONFRONTING THE NEW HEALTH CARE CRISIS: IMPROVING HEALTH CARE QUALITY AND LOWERING COSTS BY FIXING OUR MEDICAL LIABILITY SYSTEM 19 (July 2002).

49. CONG. BUDGET OFFICE, COST ESTIMATE FOR H.R. 4600 HEALTH ACT OF 2002, at 1 (Sept. 25, 2002) (ordered reported by the House Committee on Energy and Commerce).

50. *Supra* note 48, at 18.

51. R.S. BIONDI & K. QUINTILIAN, MILLIMAN & ROBERTSON, MEDICAL LIABILITY MUTUAL INSURANCE COMPANY PROJECTED EFFECT ON NEW YORK PROFESSIONAL LIABILITY COSTS OF CAPPING NONECONOMIC DAMAGES 2-4 (1995).

52. UNIV. OF CENTRAL FLA. GOVERNOR'S SELECT TASK FORCE ON HEALTHCARE PROFESSIONAL LIABILITY INSURANCE (2003) [hereinafter GOVERNOR'S TASK FORCE].

53. AM. ACAD. OF ACTUARIES, MEDICAL MALPRACTICE TORT REFORM: LESSONS FROM THE STATES 1-4 (Issue Brief, Fall 1996).

54. OHIO REV. CODE ANN. § 2307.43 (Anderson 1994) (repealed 1997).

55. *See* *Morris v. Savoy*, 576 N.E.2d 765 (Ohio 1991).

56. *Medical Malpractice Reform in California, Ohio and New York*, CONTINGENCIES (Am. Acad. of Actuaries), Sept./Oct. 1995, at 22.

one of the states the AMA has declared to be in “crisis” and is again debating the need for legal reforms.

Similarly, Oregon capped non-economic damages in 1987.⁵⁷ In 1998, the Oregon Supreme Court nullified the law.⁵⁸ By 2001, the cost of malpractice claims in the state had increased from a base \$15 million in 1998 to \$60 million, an increase of 400%, and has continued to rise since.⁵⁹

Moreover, just as the California experience has illustrated the effectiveness of MICRA, the experiences of other states have shown us how much less effective other types of reforms have been. For example, New York, Texas, and Florida have all at various times passed more limited reform measures that predictably did not affect the malpractice crisis.⁶⁰ In every case, legal reform opponents were able to substitute these measures for MICRA-based statutes knowing that they would be less effective. Those who would block necessary modification in the law will argue that tort reform sometimes fails to reduce malpractice premiums.⁶¹ Invariably, these critics cite the experiences of states that have passed peripheral or minor reforms rather than the fundamental protections embodied in MICRA. In 1996, Texas passed a package of reforms that included none of the MICRA provisions and, as could have been easily predicted, resulted in no change in malpractice insurance rates. This failure merely proves that minor reforms will often prove ineffective. In 2003, the state passed a \$250,000 limit on non-economic damages, and premium rates have already stabilized and started to head downward.⁶²

57. OR. REV. STAT. § 31.710 (2003); *see also* Press Release, Am. Med. Ass’n, AMA Joins Oregon Doctors To Protect Patient Access to Care (Aug. 10, 2004), <http://www.ama-assn.org/ama/pub/category/13939.html>.

58. *Lakin v. Senco Prods., Inc.*, 987 P.2d 463 (Or. 1997).

59. James T. Dorigan, Jr., Northwest Physicians Mutual Ins. Co., Oregon Medical Malpractice 1996-2001, Presentation to Lane County Medical Society (Sept. 10, 2002) (on file with author).

60. *See* ANDERSON, *supra* note 10, at 215; *see also* FLA. STAT. ANN. § 95.11(4)(b) (Harrison 1996 & Supp. 2000); N.Y. C.P.L.R. 214(a), 3021-a, 5031 (McKinney 1992 & Supp. 2004); TEX. CIV. PRAC. & REM. CODE ANN. §§ 33.011, 33.015, 41.008 (Vernon 1997 & Supp. 2004-2005); TEX. REV. CIV. STAT. ANN. art. 4590, §§ 10.01, 14.01 (Vernon 1997 & Supp. 2004-2005).

61. *Medical Malpractice Reform in California, Ohio and New York*, *supra* note 56, at 23.

62. Damon Adams, *Texas Tort Reform Vote Signals Lower Liability Rates*, AMNEWS, Oct. 6, 2003, <http://www.ama-assn.org/amednews/2003/10/06/gv111006.htm>.

CONCLUSION

America's physicians face an unprecedented tide of litigation.⁶³ The direct costs of this crisis exceed twenty-four billion dollars per year, but the indirect costs are much higher: The U.S. Department of Health and Human Services conservatively estimates that the cost of defensive medicine may approach \$100 billion per year.⁶⁴ The Pew Charitable Trusts project on medical liability in Pennsylvania reported that nearly forty percent of the doctors surveyed were dissatisfied with the practice of medicine.⁶⁵ These doctors are more likely to engage in "riskier prescribing practices . . . to leave clinical practice or relocate, disrupting continuity of care."⁶⁶ In particular, "[p]hysicians dissatisfied with liability risks and costs may also take specific steps to reduce their exposure, such as restricting scope of practice, avoiding high-risk patients, and engaging in 'defensive medicine.'"⁶⁷ More than ninety percent of specialists said that "the malpractice system limits doctors' ability to provide the highest-quality medical care."⁶⁸

Our medical system has been described as being on the verge of "meltdown,"⁶⁹ the AMA has declared that twenty states face medical liability crisis,⁷⁰ and physicians have started to talk about a coming "medical apocalypse."⁷¹ In many cases, legal standards of care have replaced medical standards, and the practice of defensive medicine has become the norm.⁷²

The most serious and immediate effect of the malpractice crisis is its impact on access to care. The Florida Governor's Select Task Force on Healthcare Professional Liability Insurance concluded:

The concern over litigation and the cost and lack of medical malpractice insurance have caused doctors to discontinue high-risk procedures, turn away high-risk patients, close practices, and move out of the state. In

63. See *supra* note 18 and accompanying text.

64. See U.S. DEP'T OF HEALTH & HUMAN SERVS., *supra* note 48, at 7.

65. Michelle M. Mello et al., *Caring for Patients in a Malpractice Crisis: Physician Satisfaction and Quality of Care*, 23 HEALTH AFF. 42, 45 (2004).

66. *Id.* at 43.

67. *Id.*

68. *Id.* at 49.

69. Earl R. Washburn, *The Coming Medical Apocalypse*, PHYSICIAN EXEC., Jan.-Feb. 1999, at 34, 36.

70. See Am. Med. Ass'n, *supra* note 2.

71. See Washburn, *supra* note 69, at 34.

72. See Anderson, *supra* note 14, at 1177.

some communities, doctors have ceased or discontinued delivering babies and discontinued hospital care.⁷³

The U.S. Department of Health and Human Services Agency for Healthcare Research and Quality found that in 2000 the number of physicians per capita was twelve percent higher in states with caps on non-economic damages than in states lacking these reforms.⁷⁴ The General Accounting Office has found localized health care access problems in five states experiencing rapid increases in malpractice insurance premiums,⁷⁵ and there are innumerable specific instances of this effect.⁷⁶

In sharp contradistinction, analysis of the effect of MICRA on health care access in California found that the enactment of MICRA was important to ensuring that high-cost and low-income groups have access to health care.⁷⁷ Moreover, MICRA played an important role in lowering the cost of health care in California.⁷⁸ Finally, the resulting reduction in "malpractice pressure" is expected to result in a greater number of physicians practicing in the state.⁷⁹

For more than twenty-five years, the nation has accumulated direct experience with the effect of tort reform on medical malpractice insurance premiums and access to health care. The four major reforms embodied in MICRA, including, most importantly, a \$250,000 limitation on non-economic damages, promote a stable insurance market, preserve access to care, and still provide full compensation for actual damages. We also know that lesser reforms are ineffective and divert attention from the necessary enactment of substantive legislation needed to effect real change.

Once this tort hemorrhaging has been stanchied, we need to look ahead to more profound reform. Phillip Howard has proposed specialized

73. See GOVERNOR'S TASK FORCE, *supra* note 52, at vi.

74. FRED J. H. & WILLIAM E., U.S. DEP'T OF HEALTH & HUMAN SERVS., THE IMPACT OF STATE LAWS LIMITING MALPRACTICE AWARDS ON THE GEOGRAPHIC DISTRIBUTION OF PHYSICIANS 1 (July 2003).

75. U. S. GEN. ACCOUNTING OFFICE, MEDICAL MALPRACTICE: IMPLICATIONS OF RISING PREMIUMS ON ACCESS TO HEALTH CARE (Aug. 2003). Examples of typical local health care access issues include decreased availability of Pap smears, reluctance to test HIV vaccines, and the absence of even a single neurosurgeon in large areas of West Virginia.

76. HEALTH CARE LIAB. ALLIANCE, FACT SHEET: THE HEALTH CARE LIABILITY SYSTEM BARS ACCESS TO HEALTH CARE (1997) (on file with author).

77. WILLIAM G. HAMM, LECC, AN ANALYSIS OF HARVEY ROSENFELD'S REPORT: "CALIFORNIA'S MICRA" 12 (1997).

78. *Id.* at 18.

79. *Id.* at 23.

health courts, staffed by specially trained judges with the power to hire neutral experts.⁸⁰ The goal would be to advance patient safety and increase the reliability and predictability of legal rulings on the provision of health care. Though such a proposal seems a long way from today's "shame and blame" courts, similar systems are already in place for such specialized areas as taxes, worker's compensation, and vaccine liability.⁸¹ There can be little doubt that our flawed system of medical liability is in crisis. Solutions that will provide immediate relief are available, and more profound long-term change is also needed. The alternative is simply unacceptable.

80. Philip K. Howard, Op-Ed, *A Case for Medical Justice*, PHILA. INQUIRER, May 16, 2004, <http://cgood.org/learn-reading-cgpubs-opeds-26.html>.

81. *Id.* Senate Majority Leader Dr. William Frist has endorsed the concept. See Senator Bill Frist, Summary of Key Health Policy Proposals, Address to National Press Club (July 12, 2004), http://frist.senate.gov/index.cfm?FuseAction=Speeches.Detail&Speech_id=97&Month=7&Year=2004).

Debunking Medical Malpractice Myths: Unraveling the False Premises Behind “Tort Reform”

Geoff Boehm, J.D.*

Medical malpractice—negligence and recklessness by hospitals and physicians—injures hundreds of thousands of people each year. In 2000, the Institute of Medicine released a lengthy report, *To Err Is Human*, revealing that preventable medical errors result in up to 98,000 deaths in hospitals annually.¹ Unfortunately, lawmakers and others have focused too much on reducing liability for those preventable errors and too little on reducing their occurrence. As a result, a July 2004 study shows that over a decade in which two-thirds of states passed “tort reform” measures that limit or restrict medical malpractice lawsuits, there was no improvement in safety: The number of avoidable deaths in hospitals alone is now approximately 195,000 per year, not including obstetrics patients.² Despite these bleak statistics, when organizations like the American Medical Association (AMA) speak about a malpractice “crisis,” they are referring not to the people injured or killed by medical errors or the widespread failure to discipline negligent doctors (including repeat offenders), but rather to doctors’ increasing malpractice insurance premiums.³

I. THE UNFOUNDED RHETORIC OF TORT REFORM LOBBYISTS

Tort reform lobbyists seeking to limit the rights of victims of medical

* Legal Director, Center for Justice & Democracy.

1. INST. OF MEDICINE, *TO ERR IS HUMAN: BUILDING A SAFER HEALTH SYSTEM* (2000), <http://www.nap.edu/openbook/0309068371/html/> (reporting only on deaths in hospitals, therefore not including those in physicians’ offices).

2. HEALTHGRADES, *HEALTHGRADES QUALITY STUDY: PATIENT SAFETY IN AMERICAN HOSPITALS* 6 (2004), http://www.healthgrades.com/media/english/pdf/HG_Patient_Safety_Study_Final.pdf (“[E]xcluding obstetric patients, we calculated that . . . 575,000 preventable deaths occurred, as a direct result of the 2.5 million patient safety incidents that occurred in U.S. hospitals from 2000 through 2002.”).

3. See, e.g., Donald Palmisano, AMA Past-President, Letter to Editor, N.Y. TIMES, July 2, 2004, at A18.

malpractice through caps on damages often string together various concerns about health care in the United States that are unrelated to, or would not be addressed by, the reforms they seek. In particular, the insurance industry and other tort reform proponents rely on misinformation and largely anecdotal evidence that the civil justice system is “out of control” and needs to be scaled back.⁴ However, the facts reveal a different picture.

First, the number of medical malpractice cases being filed per capita has dropped over the last ten years, as have tort filings generally.⁵ Even in the states that the AMA has labeled “crisis states,”⁶ the number of cases per capita has been dropping.⁷ The vast majority of those injured by malpractice never file a claim seeking to hold the wrongdoers accountable. Even though medical malpractice kills some 195,000 hospital patients every year and injures many more, only about one in eight of those injured files a claim.⁸

Second, while the claim that medical malpractice cases tend to be “frivolous” is frequently heard,⁹ proponents of that claim have failed to

4. *E.g.*, Texans Against Lawsuit Abuse, *Lawsuit Abuse: Patients Are Paying the Price*, at <http://www.sickoflawsuits.org/news/patients.cfm> (last visited Sept. 23, 2004) (referring to “increasing number of lawsuits in the healthcare industry,” “runaway tort system,” and “lawsuit abuse”).

5. NAT’L CTR. FOR STATE COURTS, *EXAMINING THE WORK OF STATE COURTS*, 2002, at 28 (2003) [hereinafter NCSC, *EXAMINING STATE COURTS*, 2002]; Nat’l Ctr. for State Courts, *Medical Malpractice Filings per 100,000 Population in 11 and 17 States, 1993-2002* (July 19, 2004) (unpublished manuscript, on file with author) [hereinafter NCSC, *Medical Malpractice Filings*].

6. Am. Med. Ass’n, *America’s Medical Liability Crisis: A National View* (2004), at <http://www.ama-assn.org/ama/noindex/category/11871.html>.

7. *See* NCSC, *EXAMINING STATE COURTS*, 2002, *supra* note 5, at 28; NCSC, *Medical Malpractice Filings*, *supra* note 5. Ironically, the type of cases being filed ever more frequently are contract cases, which are much more likely to be filed by a business and are not affected by caps or any other “tort reforms.” *See* THOMAS H. COHEN ET AL., U.S. DEP’T OF JUSTICE, *CIVIL TRIAL CASES AND VERDICTS IN LARGE COUNTIES*, 2001, at 3 (2004); NAT’L CTR. FOR STATE COURTS, *EXAMINING THE WORK OF STATE COURTS*, 2003, at 23 (2004).

8. HARVARD MED. PRACTICE STUDY, *PATIENTS, DOCTORS AND LAWYERS: MEDICAL INJURY, MALPRACTICE LITIGATION, AND PATIENT COMPENSATION IN NEW YORK* (1990).

9. *See, e.g.*, Elizabeth Zuckerman, *Doctors Protest Rising Medical Liability Insurance Rates*, ASSOCIATED PRESS, Sept. 23, 2004 (“[T]he immediate past president of the AMA . . . faults what he said is a higher number of frivolous lawsuits.”). President George W. Bush apparently referred to “junk” or “frivolous” lawsuits in 224 different speeches between January 1 and November 8, 2004, and in 86 speeches in 2003. Search on Nexis, *Public Papers of the Presidents Database* (Nov. 8, 2004).

support it with strong empirical support.¹⁰ Politicians, insurance industry executives, and medical society lobbyists often support their claim that the system is filled with “frivolous” malpractice lawsuits by citing the statistic that patients only prevail in their medical malpractice lawsuits about twenty-seven percent of the time.¹¹ Yet, a 2004 report from the Federal Trade Commission and the U.S. Department of Justice found that doctors’ own lawsuits against employers and hospitals fare even worse: Doctor-plaintiffs win only fourteen percent of those verdicts.¹² The fact is that some types of cases are difficult to win, even when they are legitimate—that they will have low win percentages is not a reflection of frivolity.

Our civil justice system has various checks and balances to discourage frivolous suits and punish those who file them. Not only can sanctions be imposed on the lawyers responsible,¹³ but the contingency fee arrangement under which plaintiffs’ attorneys work—they only get paid and have their expenses reimbursed if they succeed in the case—also screens out baseless lawsuits. As far back as 1986, James Gattuso, then of the conservative Heritage Foundation, wrote an article for the *Wall Street Journal* entitled *Don’t Rush To Condemn Contingency Fees*. He argued that the contingency fee system ensures that injured persons who could not otherwise afford legal representation obtain access to the legal system and “helps screen [baseless lawsuits] out of the system.”¹⁴ Even insurance executives, when put under

10. See, e.g., Alisa Ulferts, *Hitch in Malpractice Deal? Bush*, ST. PETERSBURG TIMES, July 16, 2003, at 1B (“Florida Medical Association CEO Sandy Mortham said she wasn’t in a position to say whether frivolous lawsuits caused higher insurance rates, even though the FMA has blamed such lawsuits in news releases and statements on its Web site.”).

11. Lawrence Smarr, President of the Physician Insurers Association of America, has stated that a properly functioning system “would be a system where only cases with merit would be brought forward, where the trial lawyers would triage the cases so that they don’t lose 80 percent of the time when they go to court. . . . We have a legal system that encourages the filing of frivolous lawsuits.” THOMAS H. COHEN, U.S. DEP’T OF JUSTICE, MEDICAL MALPRACTICE TRIALS AND VERDICTS IN LARGE COUNTIES, 2001, at 1 (2004); see also, e.g., Donald J. Palmisano, President, AMA, Speech at National Press Club (July 9, 2003), <http://www.npr.org/programs/npc/2003/030709.dpalmisano.html>; *NewsHour with Jim Lehrer* (PBS television broadcast, Jan. 16, 2003).

12. FED. TRADE COMM’N & U.S. DEP’T OF JUSTICE, IMPROVING HEALTH CARE: A DOSE OF COMPETITION 38 (2004), http://www.usdoj.gov/atr/public/health_care/204694.htm (describing antitrust health care litigation).

13. See, e.g., FED. R. CIV. P. 11. State corollaries also provide for such sanctions.

14. James Gattuso, *Don’t Rush To Condemn Contingency Fees*, WALL ST. J., May 15, 1986, at 28.

oath, have admitted that frivolous suits are not a problem.¹⁵

It should also be noted that the issue of “frivolous lawsuits” is a red herring when caps are being considered. By limiting award amounts, caps target the most egregious cases of malpractice and the most severely injured patients—the very opposite of the “frivolous” or “junk” lawsuits that advocates for caps portray when they are trying to rile up the public or lawmakers to limit victims’ rights. Two recent studies have confirmed that caps on damages in medical malpractice cases, such as California’s draconian \$250,000 cap on non-economic damages, are most devastating to those who suffered the most heinous injuries, those killed by the defendants’ acts, and those who suffered the greatest loss to their quality of life.¹⁶

In addition to mischaracterizing the quantity and quality of medical malpractice suits, supporters of tort reform make unsupported assertions about the impact of medical malpractice litigation on the quality and availability of health care. Despite the claims of the AMA and state medical societies, the number of medical professionals is growing. Moreover, these organizations repeatedly aver that doctors are leaving the twenty “AMA crisis states,” and even the twenty-four “AMA problem states,” in droves because of litigation concerns, resulting in a lack of access to care. However, investigations of such claims by the U.S. General Accounting Office, various reporters, and state agencies have shown the claims to be false or widely exaggerated. To the extent there are access problems, many

15. E.g., *Hearing To Receive Testimony from Invited Parties Regarding Medical Malpractice Before the Fla. Senate Comm. on Judiciary*, 2003 Leg., C Sess. 56 (Fl. 2003) (testimony of Robert White, President, First Professional Ins. Co.) (“I don’t feel you can have a frivolous lawsuit in the State of Florida.”); see also Paige St. John, *Testimony Reveals Malpractice Myths*, FLA. TODAY, July 15, 2003, at 1; Ulferts, *supra* note 10.

16. NICHOLAS M. PACE ET AL., RAND INST. FOR CIVIL JUSTICE, CAPPING NON-ECONOMIC AWARDS IN MEDICAL MALPRACTICE TRIALS: CALIFORNIA JURY VERDICTS UNDER MICRA 32-33, 47, 48 (2004) [hereinafter RAND INST.]; David M. Studdert et al., *Are Damage Caps Regressive? A Study of Malpractice Jury Verdicts in California*, 23 HEALTH AFF. 54 (2004). California’s cap also has a disproportionate impact on children under a year old and females who are injured by medical malpractice. RAND INST., *supra*, at 32 (female); *id.* at 48 (infant). Victims of medical malpractice with the severest injuries—“brain damage, paralysis, or a variety of catastrophic losses”—had their recoveries capped most often. *Id.* at 47. Patients who suffered “a great loss to their quality of life” but who had smaller economic damages lost the highest percentage of their total awards. *Id.* Death cases, where the malpractice resulted in the patient’s death, are capped more frequently and have higher percentage reductions than injury cases. *Id.*

other explanations can be established.¹⁷

For example, it is true that some rural and impoverished urban areas do not have a sufficient supply of health care providers.¹⁸ But it is a fiction to tie that lack of access to malpractice litigation or jury awards, or to claim that a cap would make a difference. Such areas often have difficulty attracting or retaining other professionals as well.¹⁹ Moreover, this problem has existed for a long time, even before physicians considered malpractice insurance premiums problematic. In fact, the Council on Graduate Medical Education has stated, “The relative shortage of health professionals in rural areas of the United States is one of the few constants in any description of the United States medical care system.”²⁰ Rural health

17. *E.g.*, PUBLIC CITIZEN & ORE. STATE PUB. INT. RES. GROUP, OREGON’S INCREASED NUMBER OF DOCTORS: GOVERNMENT DATA REFUTES MEDICAL LOBBY CLAIMS 1-11 (2004), http://ospirg.org/reports/ORIncreasedNumberDoctors8_04.pdf; U.S. GEN. ACCOUNTING OFFICE, MEDICAL MALPRACTICE: IMPLICATIONS OF RISING PREMIUMS ON ACCESS TO HEALTH CARE 16-19 (2003), <http://www.gao.gov/new.items/d03836.pdf>; John M.R. Bull, *Doctors Can’t Prove Thinning Ranks: Medical Society Chief Admits Group Lacks Statistics To Show Physicians Are Leaving*, MORNING CALL (Allentown, Pa.), Apr. 23, 2004, at A1; Diane Hirth & Bill Cotterell, *Senate Experts Dispute “Myths”*, *Malpractice Debate*, BRADENTON HERALD (Fla.), July 15, 2003, at A11.

18. *See, e.g.*, U.S. Dep’t of Health & Human Servs., Medically Underserved Areas/Medically Underserved Populations, at <http://bphc.hrsa.gov/databases/newmua/> (last visited Sept. 28, 2004).

19. *See, e.g.*, S. Res. 22, 108th Cong. (2003) (“[R]ural school districts will have difficulty competing with large school districts in recruiting and retaining quality teachers.”); *Saving America’s Great Places: The Role of Tax Incentives in Preserving Rural Communities: Hearing Before the Senate Comm. on Fin.*, 108th Cong. 1, 5 (2004) (testimony of Peter K. Froelich, Coordinator, Great Plains Population Symposium Project) (noting that “rural communities are being silently destroyed by the out-migration of young people” and that out-migration is causing “[t]he loss of our highly educated young people”); Georgeanne Artz, *Rural Brain Drain: Is It Reality?*, CHOICES, Dec. 2003, <http://www.choicesmagazine.org/2003-4/2003-4-03.htm>; *see also* Jim Damici, *Growing the Information Technology Sector in Rural Areas*, MAINE IS TECHNOLOGY, Nov. 2003, at http://www.state.me.us/newsletter/nov2003/growing_the_information_technolo.htm (noting this difficulty as it pertains to information technology professionals); Clayton W. Faubion et al., *Rural/Urban Differences in Counselor Satisfaction and Extrinsic Job Factors*, J. REHABILITATION, Oct./Nov./Dec. 2001, at 1, http://www.findarticles.com/p/articles/mi_m0825/is_4_67/ai_81759712/pg_1 (noting this difficulty as it pertains to rehabilitation counselors); Nat’l Teacher Recruitment Clearinghouse, *Teacher Shortage Areas*, at http://www.rnt.org/channels/clearinghouse/becometeacher/121_teachershort.htm (last visited Sept. 28, 2004) (noting that the need for teachers is “greatest in urban and rural communities”).

20. COUNCIL ON GRADUATE MED. EDUC., U.S. DEP’T OF HEALTH & HUMAN SERVS.,

care shortages occur throughout the world, including places where there is nothing like the U.S. civil justice system in place.²¹

II. THE TRUTH ABOUT CAPS AND OTHER MEDICAL MALPRACTICE "REFORMS"

The increasing cost of health care in the United States and the high costs of medical malpractice insurance are legitimate and pressing concerns.²² Unfortunately, caps will do little to address these issues.

First and foremost, costs related to litigation are a miniscule portion of health care spending; according to the United States Congressional Budget Office (CBO), these malpractice costs are less than two percent of total spending.²³ CBO has, in fact, noted that "a cap on noneconomic damages and a ban on punitive damages . . . would lower health care costs by only about 0.4 percent to 0.5 percent, and the likely effect on health insurance premiums would be comparably small."²⁴

Tort reform advocates often claim that doctors practice "defensive medicine" because of fears of medical malpractice suits and that this practice, in turn, raises the cost of health care.²⁵ However, in 1994, the

PHYSICIAN DISTRIBUTION AND HEALTH CARE CHALLENGES IN RURAL AND INNER-CITY AREAS 11 (Feb. 1998), <http://www.cogme.gov/10.pdf>; see also, e.g., Gregg Broffman, *How Can Pediatric Care Be Provided in Underserved Areas? A View of Rural Pediatric Care*, 96 PEDIATRICS 816, 818 (Supp. 1995); Am. Acad. of Family Physicians, *Rural Practice, Keeping Physicians In* (2002), at <http://www.aafp.org/x16635.xml>; Am. Med. Student Ass'n, *Health Care Delivery: Rural vs. Urban Communities*, at <http://www.amsa.org/programs/gpit/ruralurban.cfm> (last visited Sept. 28, 2004).

21. See, e.g., CTR. FOR JUSTICE & DEMOCRACY (CJ&D), *RURAL ACCESS TO HEALTH CARE-A GLOBAL PROBLEM, NOT A LAWSUIT PROBLEM* (2004), at http://centerjd.org/free/mythbusters-free/MB_Rural-International.pdf.

22. See, e.g., U.S. CONG. BUDGET OFFICE, *LIMITING TORT LIABILITY FOR MEDICAL MALPRACTICE 1* (2004) ("On average, premiums for all physicians nationwide rose by 15 percent between 2000 and 2002—nearly twice as fast as total health care spending per person.").

23. *Id.* at 6. CBO's "malpractice costs" are based on premiums paid, which is presumably the cost passed on to health care spenders (patients, health insurers, etc.). Such premiums cover all damages paid, the costs of litigation, insurance overhead, and other related expenses. See *id.* at 1 n.3, 6; Tillinghast Towers-Perrin, *U.S. Tort Costs: 2003 Update* 16-17, app. 5 (2003), at http://www.towersperrin.com/tillinghast/publications/reports/2003_Tort_Costs_Update/Tort_Costs_Trends_2003_Update.pdf; Ams. for Ins. Reform, Tillinghast's "Tort Cost" Figures Vastly Overstate the Cost of the American Legal System (Jan. 6, 2004), at http://centerjd.org/air/pr/Tillinghast_Overstates.pdf.

24. *Id.* at 5-6.

25. E.g., President George W. Bush's Remarks in a Discussion on Health Care in

congressional Office of Technology Assessment (OTA) found that less than eight percent of all diagnostic procedures result primarily from liability concerns.²⁶ OTA found that most physicians who “would order aggressive diagnostic procedures . . . would do so primarily because they believe such procedures are medically indicated, not primarily because of concerns about liability.”²⁷ Thus, the effects of tort reform on defensive medicine “are likely to be small.”²⁸ The CBO has also reported that “some so-called defensive medicine may be motivated less by liability concerns than by the income it generates for physicians or by the positive (albeit small) benefits to patients. . . . CBO believes that savings from reducing defensive medicine would be very small.”²⁹

The insurance industry, the U.S. Chamber of Commerce, and corporate front groups such as the American Tort Reform Association³⁰ have spent many tens of millions of dollars in pursuit of immunity or limitations on liability from wrongdoing.³¹ Their efforts include promoting insurance companies’ legislative agenda to limit liability for doctors, hospitals, HMOs, nursing homes, and drug companies that cause injury. Moreover, federal and state lawmakers, regulators, doctors, and the general public are being told by medical and insurance lobbyists that doctors’ insurance rates are rising due to increasing claims by patients, rising jury verdicts, and exploding tort system costs in general, despite clear evidence to the contrary.³² Just as caps and other tort reforms do not

Muskegon, Michigan, 40 WEEKLY COMP. PRES. DOC. 1958 (Sept. 13, 2004) (“[T]he cost of health care is skyrocketing because of the defensive medicine being practiced by docs.”).

26. *Id.*

27. OFFICE OF TECH. ASSESSMENT, U.S. CONG., DEFENSIVE MEDICINE AND MEDICAL MALPRACTICE 1 (1994), <http://www.wws.princeton.edu/cgi-bin/byteserv.prl/~ota/disk1/1994/9405/9405.PDF>.

28. *Id.* at 18.

29. CBO, *supra* note 22, at 6.

30. ATRA is funded by the AMA, the tobacco industry, gun makers, and the insurance industry. *See, e.g.*, CARL DEAL & JOANNE DOROSHOW, CJ&D & PUBLIC CITIZEN, THE CALA FILES: THE SECRET CAMPAIGN OF BIG TOBACCO AND OTHER MAJOR INDUSTRIES TO TAKE AWAY YOUR RIGHTS (2000).

31. *Id.*; DAVID C. JOHNSON, COMMONWEAL INST., THE ATTACK ON TRIAL LAWYERS AND TORT LAW 1-20, 26-51 (2003), <http://www.commonwealinstitute.org/reports/tort/tortreport.html>; Am. Tort Reform Ass’n, 50 Representative Members, at <http://atra.org/about/members.php> (last visited Aug. 12, 2004).

32. *See, e.g.*, AMS. FOR INS. REFORM, MEDICAL MALPRACTICE INSURANCE: STABLE LOSSES/UNSTABLE RATES 2003 (2003), <http://insurance-reform.org/StableLosses2003F.pdf> [hereinafter AIR, STABLE LOSSES/UNSTABLE RATES]; NCSC, EXAMINING STATE COURTS 2002,

succeed in significantly reducing aggregate health care costs, they also fail to control individual insurance premiums.

Insurers state that to recoup money paid to patients, they must raise insurance rates or, in some cases, pull out of the market altogether. Since insurers say that jury verdicts are the cause of the current "crisis" in affordable malpractice insurance for doctors, they insist that the only way to bring down insurance rates is to limit an injured consumer's ability to sue in court.³³ However, historically, the cause of skyrocketing rates has little to do with the legal system.

Insurance companies make profits primarily from investment income. Insurance companies take in money in the form of premiums paid and then hold it for some length of time until they need to make a payout to, or on behalf of, a policyholder.³⁴ In the interim, the money being held, known as the "float," is invested and earns money for the insurance company. When the investment market is strong and/or interest rates high, the companies make a good profit by investing the float and may under-price policies in an effort to attract more premium dollars to invest—this scenario is termed a "soft market."³⁵ But when investment income falls because of a decline in the markets and/or drops in interest rates, insurance companies will raise their rates or cut back coverage. Such a "hard market" occurred in the mid-1970s, more severely in the mid-1980s, and again between 2002 and 2003. Insurance rates for doctors

supra note 5, at 23-24, 28; NCSC, Medical Malpractice Filings, *supra* note 5; U.S. DEP'T OF JUSTICE, CIVIL TRIAL CASES AND VERDICTS IN LARGE COUNTIES, 2001, at 9 (Apr. 2004); Press Release, Jury Verdict Research, Jury Verdict Research Releases Verdict Survey (Apr. 1, 2004),

http://www.juryverdictresearch.com/Press_Room/Press_releases/Verdict_study/verdict_study8.html; Press Release, Public Citizen, New 2002 Government Data Dispute Malpractice Lawsuit "Crisis" (July 7, 2003), <http://www.citizen.org/pressroom/release.cfm?ID=1480> [hereinafter Press Release, Public Citizen] (reporting that malpractice payouts declined as insurance premiums spiked and that 5.2% of doctors are responsible for 55% of malpractice payouts). *But see* Press Release, Ctr. for Justice & Democracy, Flawed Jury Data Masks Trends (2002), <http://centerjd.org/press/release/020322.pdf> (regarding Jury Verdict Research's over-inflation of award figures).

33. *See, e.g., Patient Access Crisis: The Role of Medical Litigation: Joint Hearing Before the Senate Comm. on the Judiciary and the Senate Comm. on Health, Educ., Labor & Pensions*, 108th Cong. (2003) (statement of Lawrence Smarr, President, Physician Ins. Ass'n of Am.), http://judiciary.senate.gov/testimony.cfm?id=600&wit_id=1594.

34. *See* AIR, STABLE LOSSES/UNSTABLE RATES, *supra* note 32, at 4-6.

35. *Id.* at 1-2, 4-6.

skyrocketed in each of the hard markets.³⁶

Thus, while insurers and other tort reform proponents blame malpractice litigation for the hard market premium increases, they are in fact consistently driven by the insurance companies' response to the broader economic cycle.³⁷ In fact, claims and payouts stayed flat or declined through each of the "crises" or hard markets.³⁸ With payouts flat, rising premiums have caused property-casualty insurers' profits to skyrocket. From 2002 to 2003, profits rose 997% and they continue to soar³⁹—reportedly doubling between the first quarters of 2003 and 2004.⁴⁰ Despite these striking statistics, successful lobbying by interest groups in response to increasing insurance rates for doctors has yielded a wave of legislative activity to restrict injured patients' rights to sue for medical malpractice.

Because insurers target the civil justice system, rather than the economic cycle that leads to periodic "crises," "tort reform" remedies—including caps—pushed by insurance companies and their advocates during each hard market failed to bring down rates.⁴¹ When confronted with a report showing that tort reform does not lead to reduced premiums, the American Insurance Association responded, "Insurers never promised that tort reform would achieve specific savings."⁴² Over the past year and a half, insurers continued to raise premiums, even in states where tort reforms were enacted, even though claims and payouts dropped⁴³ and the

36. *Id.* at 4-6.

37. See, e.g., Ralph Nader, *The Assault on Injured Victims' Rights*, 64 DENV. U. L. REV. 625, 628 (1988).

38. See AIR, STABLE LOSSES/UNSTABLE RATES, *supra* note 32, at 5.

39. See Press Release, Ins. Servs. Office, Inc., & Prop. Cas. Insurers Ass'n of Am., Sharp Increase in P/C Industry's Net Income Propels Surplus Upward in 2003 (Apr. 14, 2004), http://www.iso.com/press_releases/2004/04_14_04.html ("The property/casualty insurance industry's net income after taxes rose to \$29.9 billion in 2003—nearly ten times the industry's \$3 billion in net income in 2002.").

40. See *P/C Industry's Q1 Net Income Doubles*, INS. J., June 29, 2004, <http://www.insurancejournal.com/news/national/2004/06/29/43661.htm>.

41. See J. ROBERT HUNTER & JOANNE DOROSHOW, CJ&D, PREMIUM DECEIT: THE FAILURE OF "TORT REFORM" TO CUT INSURANCE PRICES (2002), <http://insurance-reform.org/PremiumDeceit.pdf>.

42. Press Release, Am. Ins. Ass'n, AIA Cites Fatal Flaws in Critic's Report on Tort Reform (Mar. 13, 2002), <http://www.aiadc.org/DocFrame.asp?DocID=7027>; see also Press Release, Ams. for Ins. Reform, Industry Insiders Admit – And History Shows: Tort Reform Will Not Lower Insurance Rates (June 2, 2003), <http://centerjd.org/air/pr/Quotes.pdf>.

43. A.M. Best, Medical Malpractice Total Industry (Premiums and Losses), 2002 & 2003

investment markets began to improve. It appears we are now entering a soft market: Premiums are beginning to drop or increase more slowly in all lines of insurance, including medical malpractice—in states with and without caps or other tort reforms.⁴⁴ While the soft market will bring some relief as premiums drop, if there is no significant increase in regulation of the insurance industry, we can expect that the next downturn in the economy and the market will bring back rising premiums and, predictably, renewed efforts to blame injured patients and seek ineffective and harmful tort reforms, as insurers once again raise their rates to make up for investment losses.

So if one puts aside the unfounded rhetoric that claims to connect a need for caps to rising insurance premiums and health care costs, to a supposedly growing number of frivolous lawsuits, and to alleged movement of doctors among the states, what then are the true motivators for tort reform proponents? First, tort reform efforts (including caps), are based on a mistrust of, or discomfort with, the American institution of civil trial by jury. This fundamental right of ordinary citizens and consumers to hold accountable those with power—including corporations, large institutions, professionals, and even government—is a fulcrum of our democracy. In fact, one reason that several state courts have struck down tort reform laws as unconstitutional is the way in which the laws limit the power of juries to decide cases.⁴⁵

(data set) (copy on file with author).

44. E-mail from J. Robert Hunter, Director of Insurance, Consumer Federation of America, to Joanne Doroshow, Executive Director, Center for Justice & Democracy (July 25, 2004) (on file with author) (reporting this finding based upon Council of Insurance Agents and Brokers' Commercial Property-Casualty Market Surveys); Letter from J. Robert Hunter, Director of Insurance, Consumer Federation of America, & Birny Birnbaum, Executive Director, Center for Economic Justice, on behalf of Americans for Insurance Reform (AIR), to Insurance Commissioners of fifty states and D.C. (May 11, 2004), http://insurance-reform.org/AIR_Ins_Comm_04.pdf [hereinafter Letter to Insurance Commissioners]; Press Release, Consumer Fed. of Am., Insurer Profits Shoot Up As "Hard Market" of Soaring Commercial Insurance Rates Comes to an End (Nov. 3, 2003), <http://www.consumerfed.org/110303hardmarket.pdf>.

45. *E.g.*, *Mahomes-Vinson v. United States*, 751 F. Supp. 913 (D. Kan. 1990) (holding that a \$1,000,000 overall damage cap and \$250,000 non-economic damage cap violated jury trial right); *Waggoner v. Presbyterian Med. Ctr.*, 647 F. Supp. 1102 (N.D. Tex. 1986) (holding that a \$500,000 cap on medical malpractice recoveries violates equal protection and open courts guarantees); *Smith v. Schulte*, 671 So. 2d 1334 (Ala. 1995) (*per curiam*) (holding that a \$1 million cap in wrongful death cases against health care providers violates both equal protection and the right to jury trial); *Henderson v. Ala. Power Co.*, 627 So. 2d

Judges, who have more intimate knowledge of the system than anyone, find such mistrust of juries inappropriate. A 2000 survey sent to one thousand trial judges, including every federal trial judge, revealed that:

- Judges have “a high level of day-to-day confidence in [the jury] system.”⁴⁶
- “Only 1 percent of the judges who responded gave the jury system low marks.”⁴⁷
- “[N]ine of every 10 trial judges, those who work closest

878 (Ala. 1993) (holding that a \$250,000 punitive-damage cap violates the right to jury trial); *Moore v. Mobile Infirmary Ass'n*, 592 So. 2d 156, 158 (Ala. 1991) (holding that a \$400,000 economic damage cap in medical malpractice cases violates jury trial and equal protection guarantees); *Smith v. Dep't of Ins.*, 507 So. 2d 1080, 1089 (Fla. 1987) (per curiam) (holding that a \$450,000 cap on non-economic damages recoverable in actions for personal injury violates open courts provision); *Best v. Taylor Machine Works*, 689 N.E.2d 1057 (Ill. 1997) (holding that a \$500,000 cap on non-economic damages was a legislative remittitur, in violation of the separation of powers doctrine, and constituted impermissible special legislation as did abolition of joint and several liability and discovery statutes which mandate the unlimited disclosure of plaintiffs' medical information and records); *Wright v. Cent. Du Page Hosp. Ass'n*, 347 N.E.2d 736 (Ill. 1976) (holding a \$500,000 cap unconstitutional as a denial of equal protection); *Brannigan v. Usitalo*, 587 A.2d 1232, 1237 (N.H. 1991) (holding that a \$875,000 limitation on non-economic damages recoverable in actions for personal injury violates equal protection); *Carson v. Mauer*, 424 A.2d 825, 836-38 (N.H. 1980) (holding that abrogation of the collateral source rule and the \$250,000 non-economic damage cap in medical malpractice cases violate equal protection); *Arneson v. Olson*, 270 N.W.2d 125, 135-36 (N.D. 1979) (holding that the \$300,000 limit on damages recoverable in medical malpractice actions violates state and federal equal protection guarantees); *State ex rel. Ohio Acad. of Trial Lawyers v. Sheward*, 715 N.E.2d 1062 (Ohio 1999) (holding that a \$250,000 non-economic damages cap, a \$250,000 punitive damages cap, a certificate of merit, and modification of the collateral source rule violate separation of powers); *Lakin v. Senco Prods., Inc.*, 987 P.2d 463 (Ore. 1999) (holding that a \$500,000 cap on non-economic damages in personal injury and wrongful death actions violates the right to a jury trial); *Knowles v. United States*, 544 N.W.2d 183 (S.D. 1996) (holding that a \$1 million medical malpractice compensatory damage cap violates substantive due process); *Lucas v. United States*, 757 S.W.2d 687, 690-92 (Tex. 1988) (holding that a \$500,000 cap for damages in medical malpractice actions violates the open courts guarantee); *Condemarin v. Univ. Hosp.*, 775 P.2d 349, 364, 366 (Utah 1989) (holding that a \$100,000 medical malpractice liability limit for state hospitals violates the right to jury trial).

46. Allen Pusey, *Judges Rule in Favor of Juries: Surveys by Morning News*, *SMU Law School Find Overwhelming Support for Citizens' Role in Court System*, DALLAS MORNING NEWS, May 7, 2000, at 1J.

47. *Id.*

with the nation's jury system, think the system needs only minor tinkering, at best."⁴⁸

- "Overwhelmingly . . . state and federal judges said they have great faith in juries to solve complicated issues."⁴⁹
- "[N]ine of 10 judges responding said jurors show considerable understanding of legal issues involved in the cases they hear."⁵⁰

Statistics also show that juries are generally conservative and reasonable, and their decisions rarely differ from what a judge would decide.⁵¹

III. RECOMMENDATIONS

Our civil justice system exists to provide those who have been wronged a forum to seek truth and compensation, even to the dismay of those who may have acted negligently, recklessly, or worse. Caps not only limit the liability of wrongdoers, take away the fundamental power of juries to decide adequate compensation, and leave the most severely injured victims without sufficient means of redress, but they do not even address the increasing costs of health care or medical malpractice insurance.

An important solution to avoiding future spikes in premiums is stronger regulation of the insurance industry. Unlike caps and other tort reforms, insurance industry regulation would lower premiums charged to doctors, hospitals, and other policyholders, while protecting the rights of patients and consumers. Given the soaring profits of insurance companies,⁵² such regulation is unlikely to put them in financial harm.

48. *Id.*

49. *Id.*

50. *Id.*; see also CJ&D, JUDGES SHOW EXTRAORDINARY SUPPORT FOR THE CIVIL JURY SYSTEM, at <http://centerjd.org/private/mythbuster/JudgeSurvey.pdf>.

51. E.g., Marc Galanter, *Real World Torts: An Antidote to Anecdote*, 55 MD. L. REV. 1093, 1110-12 & tbl.2 (1996) (citing Mark I. Taragin et al., *The Influence of Standard of Care and Severity of Injury on the Resolution of Medical Malpractice Claims*, 117 ANNALS INTERNAL MED. 780, 782 (1992)).

52. Profits Are the Highest Ever, at <http://centerjd.org/free/mythbusters-free/InsProfitsHighestEver.pdf> (last visited Nov. 29, 2004); Press Release, Ins. Servs. Office, Inc. & Property Cas. Ins. Ass'n of Am., Property/Casualty First-Half Income and Surplus Rose on Strong Underwriting Results and Investment Gains (Oct. 18, 2004), http://www.iso.com/press_releases/2004/10_18_04.html; Press Release, Ins. Servs. Office, Inc. & Property Cas. Ins. Ass'n of Am., *supra* note 39.

State insurance regulators should take the following steps, as suggested by Americans for Insurance Reform—a coalition of over one hundred consumer and public interest groups and a project of the Center for Justice & Democracy—in a recent letter sent to all state insurance commissioners:

(1) Undertake a review of rate levels to determine if rates are excessive in any line of insurance; . . . (2) Initiate an investigation into anti-competitive behavior of insurance companies in making statements and other acts to hold off competition; . . . (3) If any insurer files a rate request in excess of current inflation for that line of insurance, a rate hearing should be called; . . . (4) [B]egin the process of careful analysis as to what led to this most recent cycle, and your department's role in it by allowing rates to fluctuate between excessive (such as now at the end of the hard market) and inadequate (such as right before the turn in the market from soft to hard); . . . (5) Alert your legislature to the end of the hard market and advise them that there is no need to rush into legislative fixes, such as legal limits on victims' rights; . . . (6) Review successes from other states in averting the same sort of price spikes you may have endured over the last two years. Clearly, insurance rate regulation is one thing that has helped tremendously to prevent large rate increases in some states. Nowhere has this been more evident than in California, a state that in 1988 passed the strongest insurance reform law in the country.⁵³

No one denies that there is a broad array of very serious health care issues facing the United States right now—patient safety, rising costs, availability and affordability of health insurance, and, in some places, rapidly rising malpractice premiums (although they are easing as we enter a soft market). But even with these problems, caps are not a solution. Lawmakers and regulators should stop the insurance industry from price-gouging their policyholders, even while the industry's profits rocket upwards. Moreover, doctors would better serve themselves and their patients by directing their anger and efforts regarding rising premiums toward the questionable practices of the insurance industry and the subset of doctors who repeatedly commit malpractice without facing adequate discipline.⁵⁴ Seeking to take away patients' rights is not the answer.

53. Letter to Insurance Commissioners, *supra* note 44.

54. See, e.g., Press Release, Public Citizen, *supra* note 32 (noting that 5.2% of doctors are responsible for 55% of malpractice payouts).

Health Care in Crisis: The Need for Medical Liability Reform

Donald J. Palmisano, M.D., J.D.*

Health care in the United States is currently in a state of crisis, and the need for reform is significant. The American Medical Association (AMA) was established in 1847 in large part to help safeguard and improve medical care and patient safety.¹ Unfortunately, 150 years later, the health care system is seriously jeopardized by the detrimental effects of this nation's broken medical liability system.

Indeed, medical negligence lawsuits are as old as the AMA. At the same time the AMA was taking shape, pioneering physicians were discovering new treatments for previously untreatable conditions—for example, doctors developed methods to heal compound fractures that did not require amputation.² Yet these advancements produced a surprising result: Trial lawyers began using the example of “a limb [that] had healed to a shortened, deformed, or frozen position” as the basis for medical negligence lawsuits.³ As a result of these lawsuits, “some of the best physicians in the country stopped taking such cases.”⁴

Today, lawsuits against skilled physicians are yielding largely the same result: Experienced obstetricians no longer deliver babies; highly-trained neurosurgeons no longer perform life-saving brain and trauma surgery; and orthopedic surgeons no longer perform complex procedures.⁵ Ironically, as physicians grow increasingly skilled at treating the most complex conditions, personal injury lawyers target those same high-risk specialists.⁶ Indeed, the AMA has found that the number of U.S. states with

* Immediate Past President, American Medical Association.

1. AM. MED. ASS'N, CARING FOR THE COUNTRY 14-15 (1997).

2. James Mohr, *American Medical Malpractice Litigation in Historical Perspective*, 283 JAMA 1733 (2000).

3. *Id.*

4. *Id.*

5. Am. Med. Ass'n, *America's Medical Liability Crisis: A National View* (2004), at <http://www.ama-assn.org/ama/noindex/category/11871.html>.

6. Mohr, *supra* note 2.

"crisis" situations has increased from twelve to twenty since it began its most recent national medical liability crisis in 2002.⁷

The developments prompting the AMA's concerns have not gone unnoticed, and the resulting policy debates have been contentious. Several state legislatures have gone into extra sessions to try to resolve the crisis.⁸ In Congress, the fight to address the medical liability crisis has been particularly divisive: The House of Representatives has passed medical liability reforms multiple times,⁹ but none has passed the Senate.

The bitterness of this dispute can be traced to personal injury lawyers' desire to maintain the status quo of a civil justice system where multi-million dollar jury awards benefit a very few, but have negative ripple effects that affect many. The average jury award in 2002 reached \$6.2 million in medical negligence cases.¹⁰ Between 1996 and 2002, the average liability judgment increased 234%, and by 2001-2002, fifty-two percent of

7. Am. Med. Ass'n, *supra* note 5. Those twenty states are Arkansas, Connecticut, Florida, Georgia, Illinois, Kentucky, Massachusetts, Mississippi, Missouri, New Jersey, Nevada, New York, North Carolina, Ohio, Oregon, Pennsylvania, Texas, Washington, West Virginia, and Wyoming. As a result, the AMA has made medical liability reform its top legislative priority. *Id.* In determining whether a state is "in crisis," "showing problem signs," or "currently okay," *id.*, the AMA considers a wide variety of factors and available sources. The primary factor in the AMA analysis is the degree to which patients have lost access to medical care. For example, the AMA is concerned with newspaper and other anecdotal reports showing that a growing number of physicians no longer provide crucial medical services such as delivering babies and providing trauma care. The AMA also considers each state's legislative, legal, and judicial climates; the affordability and availability of professional liability insurance; and the trends of jury awards and settlements. Other factors include the frequency and severity of lawsuits, the quality and presence of a state's medical liability laws, and the likelihood of reforms being enacted and/or constitutionally upheld. See, e.g., Am. Med. Ass'n, Statement for the Record of the American Medical Association to the Senate Health, Education, Labor and Pensions Committee and the Senate Judiciary Committee RE: Patient Access Crisis: The Role of Medical Litigation (Feb. 11, 2003) [hereinafter Am. Med. Ass'n Statement], <http://www.ama-assn.org/ama/pub/category/12990.html>.

8. See, e.g., Steve Stanek, *Illinois Legislature Fails Again at Tort Reform*, HEALTH CARE NEWS (Sept. 1, 2004), <http://www.heartland.org/Article.cfm?artId=15515>.

9. See HEALTH Act of 2004, H.R. 4280, 108th Cong. (2004); Pregnancy and Trauma Care Access Protection Act of 2004, S. 2207, 108th Cong. (2004); Healthy Mothers and Healthy Babies Access to Care Act of 2003, S.2061, 108th Cong. (2003); Help Efficient, Accessible, Low-cost, Timely Healthcare (HEALTH) Act of 2003, H.R. 5, 108th Cong. (2003); Patients First Act of 2003, S. 11, 108th Cong. (2003).

10. JURY VERDICT RESEARCH, CURRENT AWARD TRENDS IN PERSONAL INJURY 18 (43d ed. 2004).

all awards for medical negligence cases were for one million dollars or more.¹¹ Physicians and patients seek reform because these excesses have caused significant disruption and skyrocketing costs to the health care system.

This Case Study argues that California's Medical Injury Compensation Reform Act (MICRA) of 1975¹² is a model of the type of reform needed to guide deliberations and action in Congress and in the states without reform.

I. HEALTH CARE IN CRISIS

In medicine, it is necessary to diagnose the problem before one can correctly treat the patient; the same holds true in the medical liability reform debate. When people think about the medical liability crisis, they may think first of the staggering jury verdicts leveled against defendant physicians.¹³ Indeed, current trends in jury awards illustrate why the medical liability crisis has taken such deep hold: The median medical liability award in medical liability cases jumped 114% from 1996 to 2002, topping one million dollars.¹⁴

It is also important to note the significant costs that trials inflict on physicians, even when they are not found liable, as is often the case. Nearly seventy percent of medical liability claims in 2002 were closed without payment to the plaintiff.¹⁵ In fact, plaintiffs lost the majority of their cases that went to a jury: Of the 4.9% of claims decided by jury verdict, the defendant won 82.4% of the time.¹⁶ However, physicians who prevail at trial still have large fees—on average, more than \$77,000 per claim—to pay for their defenses.¹⁷ Yet, as significant as these costs are, the most dramatic consequences of the medical liability crisis are not the direct effects on physicians, but the indirect effects on patients and the health care system as a whole.

11. *Id.* at 18, 43.

12. Medical Injury Compensation Reform Act (MICRA) of 1975, chs. 1-2, 1975 Cal. Stat. 2d Ex. Sess. 3949.

13. *E.g.*, Walter Olsen, *Curing Health Care; Delivering Justice*, WALL ST. J., Feb. 27, 2003, at A12 ("Most juries, it seems, decide such [medical negligence] cases in favor of the defense. But those that find for the plaintiff return awards that not infrequently top \$10 million.").

14. JURY VERDICT RESEARCH, *supra* note 10, at 18.

15. PHYSICIAN INSURERS ASS'N OF AM., PIAA CLAIM TREND ANALYSIS exhibits 1-2 (2003).

16. *Id.* exhibits 1-2, 6a.

17. *Id.* exhibits 6a-4. In cases where the claim was dropped or dismissed, costs to defendants averaged almost \$16,307. *Id.* exhibits 6b-4.

The importance of these indirect effects is reflected in the criteria that the AMA uses to determine whether a state is in a "state of crisis" as a result of its medical liability environment. While the AMA considers a wide variety of factors, the most important of these is the magnitude of patients losing access to care.¹⁸ The largely indiscriminate nature of the system—where anyone can file a lawsuit for any reason regardless of whether there is evidence that negligence occurred—has engendered a fear of liability in physicians that is harmful to individual patients and to the health care system as a whole. Fear of liability influences both the specialties that physicians pursue, as well as the ways in which they practice medicine. Medical residents, for example, appear to be growing increasingly concerned about liability issues.¹⁹ Sixty-two percent of medical residents reported that liability issues were their top concern in 2003—surpassing any other concern, and representing an enormous increase from 2001, when only fifteen percent of residents said liability was a concern.²⁰ The AMA is concerned that medical residents' growing concerns may cause them to avoid choosing high-risk specialties or practicing in a crisis state.

These fears extend to our nation's medical students as well. Approximately half of the respondents to a recent AMA survey indicated that the current medical liability environment was a factor in their specialty choice.²¹ There are many reasons medical students and residents choose their future specialty, but it is a troubling sign that our nation's vicious litigation system may exacerbate a potential shortage of high-risk specialists. In addition, thirty-nine percent said the medical liability environment was a factor in their decisions about whether they would like to complete residency training in a given state.²² Finally, sixty-one percent of students reported that they are extremely concerned that the current medical liability environment is decreasing physicians' ability to provide quality medical care.²³ These fears become no less salient once physicians

18. See Am Med. Ass'n, *supra* note 7. We use the term "magnitude" to indicate that we consider not only the number of patients that are affected, but also the extent to which they are affected.

19. Daniel Merenstein, *Winners and Losers*, 291 JAMA 15 (2004).

20. MERITT, HAWKINS & ASSOC., SUMMARY REPORT: 2003 SURVEY OF FINAL YEAR MED. RESIDENTS 5 (2003).

21. DIV. OF MKT. RESEARCH & ANALYSIS, AM. MED. ASS'N, AMA SURVEY: MEDICAL STUDENTS' OPINIONS OF THE CURRENT MEDICAL LIABILITY ENVIRONMENT 1 (Nov. 2003). Forty-eight percent of students in their third or fourth year of medical school indicated that the liability situation was a factor in their specialty choice. *Id.*

22. *Id.*

23. *Id.*

start practicing. On the one hand, liability fears can discourage innovation in medical practice: Fifty-nine percent of physicians believe that the fear of liability discourages open discussion and thinking about ways to reduce health care errors.²⁴ On the other hand, it can encourage the performance of unnecessary and costly tests.²⁵ This practice of “defensive medicine” takes many forms, including ordering tests and performing procedures that may not be clinically indicated; referring patients to emergency departments, safety net hospitals, and academic health centers; declining to take calls in the emergency department and declining elective referrals from emergency departments and safety net clinics, especially for uninsured patients. All of these forms of “defensive medicine” are driven by liability concerns.²⁶ Defensive medicine is one of the most difficult components of the medical liability debate to quantify, but it is perhaps one of the most costly—the costs of defensive medicine are estimated to be between \$70 billion and \$126 billion per year.²⁷

The costs of the liability crisis affect the U.S. health care system in a number of ways. Most disturbingly, as physicians’ liability insurance premiums increase dramatically, physicians restrict services, retire early, or relocate to another geographic area where the liability system is more stable. For example, forty-five percent of hospitals reported that the professional liability crisis has resulted in the loss of physicians and/or reduced coverage in emergency departments.²⁸ In turn, patients may be forced to wait longer to see a specialist (such as to receive a mammogram) or travel longer distances to receive care (such as when a pregnant woman in a rural community loses her doctor); cases of resulting patient deaths have been reported.²⁹

24. HARRISINTERACTIVE INC., COMMON GOOD, COMMON GOOD FEAR OF LITIGATION STUDY: THE IMPACT ON MEDICINE 65 (2002), <http://cgood.org/assets/attachments/57.pdf>.

25. Richard Anderson, Commentary, *Billions for Defense: The Pervasive Nature of Defensive Medicine*, 159 ARCH. INTERNAL MED., 2399, 2399 (1999) (“Malpractice litigation is the primary cause of defensive medicine.”).

26. ROBERT BERENSON ET AL., MEDICAL MALPRACTICE LIABILITY CRISIS MEETS MARKETS: STRESS IN UNEXPECTED PLACES (Ctr. for Studying Health System Change, Issue Brief No. 68, Sept. 2003), <http://www.hschange.org/CONTENT/605/?words=malpractice>.

27. OFFICE OF THE ASSISTANT SEC’Y FOR PLANNING & EVALUATION, U.S. DEP’T OF HEALTH & HUMAN SERVS., ADDRESSING THE NEW HEALTH CARE CRISIS: REFORMING THE MEDICAL LITIGATION SYSTEM TO IMPROVE THE QUALITY OF HEALTH CARE 11 (2003).

28. AM. HOSP. ASS’N, PROFESSIONAL LIABILITY INSURANCE: A GROWING CRISIS: RESULTS OF THE AHA SURVEY OF HOSPITALS ON PROFESSIONAL LIABILITY EXPERIENCE 2 (2003).

29. See Health Coalition on Liab. & Access, Who Is Being Hurt: Patient Profiles, at <http://www.hcla.org/who.html> (last visited Oct. 18, 2004).

II. CAUSES OF THE CRISIS

Perhaps medical malpractice claims would be less problematic if such claims were the result of real negligence on the part of physicians and others in the medical community. Yet, the data indicate otherwise: One study found that “a substantial majority of medical negligence claims filed are not based on actual provider carelessness.”³⁰ In fact, the study found that negligence had occurred in only one-sixth of the filed claims³¹ and that “in its initial filing stage the tort system is even more error-prone than the medical care system.”³² Another study, conducted in 1996, found that the only significant predictor of payment to medical liability plaintiffs in the form of a jury verdict or a settlement was disability and *not* the presence of an adverse event due to negligence.³³ In other words, the severity of a patient’s disability determined the jury award, not the actions of the physician.³⁴ These data suggest that it is not physician negligence, but the zealotry of personal injury attorneys, that is prompting the medical liability crisis.

Some have offered alternative explanations for the crisis to avoid criticism of plaintiffs and their attorneys, but these explanations do not hold up under scrutiny. Some claim that physicians are victims of insurance companies that made bad business decisions and are now trying to recoup their losses.³⁵ However, investment yields of medical liability insurers have been stable and positive since 1998.³⁶ Moreover, a report by

30. PAUL C. WIELER ET AL., A MEASURE OF MALPRACTICE: MEDICAL INJURY, MALPRACTICE LITIGATION & PATIENT COMPENSATION 140 (1993).

31. *Id.* at 139.

32. *Id.* at 140.

33. Troyen A. Brennan et al., *Relation Between Negligent Adverse Events and the Outcomes of Medical-Malpractice Litigation*, 335 NEW ENG. J. MED. 1963 (1996).

34. *Id.* at 1965.

35. JACKSON WILLIAMS, PUBLIC CITIZEN, MEDICAL MALPRACTICE BRIEFING BOOK: CHALLENGING THE MISLEADING CLAIMS OF THE DOCTORS’ LOBBY 24 (2004), <http://www.citizen.org/documents/MedMalBriefingBook08-09-04.pdf>. Opponents of medical liability reform often suggest that the insurance industry is to blame for indiscriminately increasing physicians’ premiums. *See, e.g.*, Press Release, Ams. for Ins. Reform, Insurers Continue To Price-Gouge Doctors Despite Dropping Medical Malpractice Payouts (Oct. 12, 2004), <http://www.insurance-reform.org/pr/041012.html>.

36. Those returns have ranged from 4.5%-5.4% and include income from interest, dividends, and real estate income. *See* AM BEST, BEST’S AGGREGATES & AVERAGES - PROPERTY/CASUALTY, QUANTITATIVE ANALYSIS REPORT, MEDICAL MALPRACTICE PREDOMINATING 335 (2003).

the U.S. General Accounting Office sheds light on the cause of recent escalation in physicians' medical liability insurance premiums and found that "[i]ncreased losses on claims are the primary contributor to higher medical malpractice premium rates"³⁷ and "[i]nsurers are not charging and profiting from excessively high premium rates."³⁸ The facts simply do not justify placing blame on the insurance industry for an out-of-control legal system.

III. ENDING THE CRISIS

Numerous studies of the medical liability crisis in states that have implemented reforms have revealed the value of such efforts. To begin, reforms have been linked with an overall decrease in medical expenditures: "[M]alpractice reforms that directly reduce provider liability pressure lead to reductions of 5 to 9 percent in medical expenditures without substantial effects on mortality or medical complications."³⁹ Importantly, reforms are also credited with reducing physicians' premiums. For example, one study found that in states with direct reforms, including caps on non-economic damages, premiums declined by 8.4% within three years.⁴⁰ According to another report, capping medical liability awards reduced premiums for general surgeons by an average of thirteen percent in the year following enactment of the reform and by an average of thirty-four percent over the long term.⁴¹ Premiums for general practitioners and obstetricians were affected similarly.⁴²

Comparative data provide support. If we consider similar major metropolitan markets and the premiums charged to physicians, we observe vast differences between states which limit non-economic damages,⁴³ such

37. U.S. GEN. ACCOUNTING OFFICE, MED. MALPRACTICE INSURANCE: MULTIPLE FACTORS HAVE CONTRIBUTED TO INCREASED PREMIUM RATES 15 (2003), <http://www.gao.gov/new.items/d03702.pdf>.

38. *Id.* at 32.

39. DANIEL P. KESSLER & MARK B. MCCLELLAN, DO DOCTORS PRACTICE DEFENSIVE MEDICINE? 2 (Nat'l Bur. of Econ. Analysis, Working Paper No. 5466, 1996), <http://www.nber.org/papers/w5466.pdf>.

40. Daniel P. Kessler & Mark B. McClellan, *The Effects of Malpractice Pressure and Liability Reforms on Physicians' Perceptions of Medical Care*, 60 LAW & CONTEMP. PROBS. 81, 98 (1997).

41. Stephen Zuckerman et al., *Effects of Tort Reforms and Other Factors on Medical Malpractice Insurance Premiums*, 27 INQUIRY 167 (1990).

42. *Id.*

43. Twenty-two states currently have some type of a cap on non-economic damages, and

as California, and states which do not provide such limits:

Table 1: Professional Liability Insurance:
Manual rates (in U.S. dollars) for \$1M/\$3M policies⁴⁴

	2000	2001	2002	2003
Ob-gyn				
Florida (Miami-Dade)	147,621	166,368	201,376	249,196
Illinois (Chicago)	78,880	88,928	102,640	139,696
Pennsylvania (Philadelphia)	37,556	45,938	100,045	134,335
Ohio (Cleveland)	56,166	72,541	100,691	119,482
California (Los Angeles)	52,874	52,874	54,563	60,259
General Surgery				
Florida (Miami-Dade)	110,068	124,046	174,268	226,542
Illinois (Chicago)	52,364	59,016	68,080	92,576
Pennsylvania (Philadelphia)	33,684	35,793	82,157	108,038
Ohio (Cleveland)	39,676	51,274	70,948	84,056
California (Los Angeles)	32,507	32,507	36,740	45,421
Internal Medicine				
Florida (Miami-Dade)	32,744	38,378	56,153	65,697
Illinois (Chicago)	19,604	22,060	26,404	35,756
Pennsylvania (Philadelphia)	7,390	7,853	18,429	24,546
Ohio (Cleveland)	12,192	15,828	21,375	25,013
California (Los Angeles)	10,097	10,097	11,164	12,493

six states have a cap on total damages. For a full discussion and comparison of different state laws, see AM. MED. ASS'N, MEDICAL LIABILITY REFORM—NOW! (2004), <http://www.ama-assn.org/amal/pub/upload/mm/450/mlrnowjune112004.pdf>.

44. While California has a “hard” damages cap, the other states do not. The October issues of *Medical Liability Monitor* for the years 2000 through 2003 provide these manual rates for professional liability insurance. *Medical Liability Monitor*, an independent Chicago-based publication, completed comprehensive rate reports of insurers in all fifty states. This table does not include all the rates reported for the geographic areas selected above, nor the premiums paid by physicians in other areas of the country, which may be higher or lower. These rates reflect the manual rates for one of the state’s marketshare leaders. The MLM notes that these rates do not reflect credits, surcharges, or other factors that may reduce or increase the actual rates charged to physicians. The AMA alone is responsible for the accuracy of the above information taken from the MLM and believes the rates listed above are a reasonable benchmark to demonstrate professional liability insurance trends for select specialties in certain geographic areas.

The AMA supports California's reforms, as set forth in MICRA, as a model for federal and state legislation: MICRA has successfully moderated physicians' professional liability insurance premium increases, while preserving patients' access to the courts.⁴⁵ This is not to say MICRA is the only legislative solution, but its efficacy is now time-tested.

With its \$250,000 cap on non-economic damages (it does not limit economic damages), joint and several liability reform, a sliding-scale contingency fee schedule, and other reforms, MICRA has resulted in stable and moderate increases in premiums in California: Between 1976 and 2002, premiums in California rose 235%, while premiums in the rest of the United States rose 750%.⁴⁶ According to Phil Hinderberger of Norcal Mutual, a major California insurer, before MICRA was passed "California physicians paid almost 25 percent of all medical liability premiums paid in the [United States] at a time when they represented only about 10 percent of all practicing physicians in the [United States]. Today, California physicians pay about [ten] percent of all medical liability premiums paid in the [United States] which represents a fair share."⁴⁷ Because of MICRA, premiums for specialists in Los Angeles are substantially less than for specialists in metropolitan areas in states without reforms such as Florida, Illinois, and Nevada.⁴⁸ Moreover, in California, claims are settled in one-third less time than in states without caps on non-economic damages⁴⁹—not only decreasing the cost of litigation, but also resulting in injured patients being compensated far faster. An important element of MICRA's success is that it has been upheld by the California State Supreme Court.⁵⁰ Other states have not been so lucky. Illinois, Ohio, Oregon, and Washington have had reforms overturned by the courts,⁵¹ while the state

45. AMA policy is decided by its House of Delegates, which has determined that MICRA-type reforms should be the basis for federal legislative support. AMA policy also supports a state's right to determine whether other types of medical liability reforms may be more appropriate for that state.

46. NAT'L ASS'N OF INS. COMM'RS, PROFITABILITY BY LINE BY STATE, 1976-2002, at 116-17 (2002).

47. Posting of Phil Hinderberger, phil-hinderberger@norcalmutual.org, to asmac-l@unity.ama-assn.org (Jan. 20, 2003) (copy on file with author).

48. See *supra* note 44 and accompanying table.

49. *Harming Patient Access to Care: The Impact of Excessive Litigation: Hearing Before the Subcomm. on Health of the Comm. on Energy and Commerce*, 107th Cong. 88 (2002) (statement of Richard E. Anderson, Chairman of the Doctors' Co. for the Physician Ins. Ass'n of Am.).

50. *Hoffman v. United States*, 767 F.2d 1431 (9th Cir. 1985); *Fein v. Permanente Med. Group*, 695 P.2d 665 (Cal. 1985).

51. *Best v. Taylor Mach. Works*, 689 N.E. 2d 1057 (Ill. 1997); *State ex rel. Ohio Acad. of*

constitutions in Arizona, Kentucky, and Pennsylvania prohibit caps on non-economic damages.⁵²

Another important element of MICRA compared to other states that have enacted a cap is the quality of the cap. For example, a state with a “hard” cap on non-economic damages should not be compared to a state with a “soft” cap on non-economic damages. A hard cap, like the \$250,000 cap found in California’s MICRA is not subject to exceptions, does not adjust over time, and applies irrespective of the number of defendants or plaintiffs. By contrast, a soft cap may be subject to numerous exceptions; increase annually with inflation, other economic indicators, or based on a set schedule; or apply individually to every defendant or plaintiff, thereby allowing several caps for a single claim. Missouri illustrates the problems presented by soft caps: The cap in Missouri increases with inflation. Originally set at \$350,000 in 1986, the cap reached \$565,000 as of February 1, 2004.⁵³ Missouri’s cap was also considerably weakened by the courts in a 2002 decision, *Scott v. SSM Healthcare*, in which the court held that the cap can be applied separately for each act of medical liability.⁵⁴ Therefore, if there are two separate and distinct “occurrences” of liability that contribute to a single injury the court can apply a separate cap for each occurrence even if they are applied to a single defendant. Where there are exceptions to the caps, there is not the same predictability afforded to physicians and insurers under MICRA.

Indeed, while the need for reform is clear, achieving it has not been as easy as one might hope. Florida has only been able to pass untested reforms, including a \$500,000 cap on non-economic damages that is subject to broad exceptions which will certainly be the subject of judicial interpretation for years to come.⁵⁵ In Iowa and Missouri, after bitter debate, the legislatures finally passed reforms,⁵⁶ including hard caps on non-economic damages, but the governors vetoed them.⁵⁷ State legislatures in

Trial Lawyers v. Sheward, 715 N.E. 2d 1062 (Ohio 1999); *Lakin v. Senco Prods., Inc.*, 987 P.2d 463 (Or. 1999); *Sofie v. Fibreboard Corp.*, 771 P.2d 711 (Wash. 1989).

52. ARIZ. CONST. art. II, § 31; KY. CONST. § 54; PA. CONST. art. III, § 18.

53. MO. REV. STAT. § 538.210 (2004).

54. *Scott v. SSM Healthcare*, 70 S.W.3d 560, 569, 571 (Mo. Ct. App. 2002); *see also Cook v. Newman*, 142 S.W.3d 880 (Mo. Ct. App. 2004).

55. FLA. STAT. ch. 766.118 (2004).

56. H.F. 2440, 80th Gen. Assem., Reg. Sess. (Iowa 2004); H.B. 1304, 92d Gen. Assem., Reg. Sess. (Mo. 2004).

57. *See Iowa Med. Soc’y, Governor Vetoes Tort Reform Bill*, IMS ADVOCATE, May 21, 2004, at 1; MISSOURI STATE MED. ASS’N, 2004 LEGISLATIVE REV. 1 (2004), <http://msma.org/docs/Legislative%20Review.pdf>.

Pennsylvania,⁵⁸ Massachusetts,⁵⁹ North Carolina,⁶⁰ Virginia,⁶¹ Connecticut,⁶² and Washington⁶³—to name a few—were unable to enact proven reforms in 2004.

MICRA-type legislation has also been pursued on the federal level. However, the battles in Congress have largely mirrored those of the states and have been characterized by intense partisanship. Multiple acts have passed in the Republican-dominated House of Representatives, but have repeatedly stalled in the Senate.⁶⁴

Despite the obvious challenges, there have been some signs of promising change: Patients and policy makers worked together in Texas in late 2003 to enact reforms that have lowered liability insurance premiums.⁶⁵ In September 2003, Texas voters cemented the reforms with enactment of Proposition 12, a “constitutional amendment concerning civil lawsuits against doctors and health care providers, and other actions, authorizing the legislature to determine limitations on non-economic damages.”⁶⁶ Reforms recently enacted in West Virginia and Mississippi have potential, but their future will not truly be known until the laws pass likely

58. Tanya Albert, *Pennsylvania Physicians' Efforts on Liability Reform Hit a Roadblock: In Wyoming, Legislators Discuss Tort Reform in a July Special Session*, AMNEWS, July 26, 2004, at <http://www.ama-assn.org/amednews/2004/07/26/gvsb0726.htm>.

59. Tom Walsh, *Lawmakers Pass Limited Medical Liability Bill, MMS Vows To Continue Push for More Adequate Reform* (July 28, 2004), at <http://www.massmed.org/vitalsigns/aug04/top1.html>.

60. Dan Kane, *Special Interests Get Their Way: The 2004 Legislative Session, Set to End Today, Shows that Well-Financed Interests Get Legislators' Attention*, NEWS & OBSERVER (Raleigh, N.C.), July 18, 2004, at A1.

61. Virginia S.B. 601 did not include a cap on non-economic damages, leading the Medical Society of Virginia to make passing a cap on of its top priorities for 2005. Med. Soc'y of Va., *MSV's 2005 Legislative Agenda*, at <http://www.msv.org/public/articles/index.cfm?cat=225>).

62. Conn. State Med. Soc'y, *Important Update on Medical Liability Insurance Reform: Governor Vetoes HB 5669* (May 11, 2004), at <http://csms.org/content/showpage.asp?page=lu6>.

63. Carol M. Ostrom, *Initiative Would Seek Malpractice-Suit Caps*, SEATTLE TIMES, June 15, 2004, at B2.

64. See *supra* note 9.

65. TEX. CIV. PRAC. & REM. § 74.301 (2004); Senator John Cornyn, *Address to Senate on One-Year Anniversary of Prop. 12 Passage* (Sept. 13, 2004), <http://www.cornyn.senate.gov/record.cfm?id=226028&ref=home>.

66. Tex. Sec'y of State, *Proposed Constitutional Amendments* (Sept. 13, 2003), <http://www.sos.state.tx.us/elections/voter/2003sepconsamend.shtml>.

constitutional challenges.⁶⁷

Several states have enacted a number of reforms over the years that may be viable enhancements to MICRA-type reforms, such as pre-trial screening panels, arbitration, mediation, alternative dispute resolution, binding arbitration, and private judging. An alternative judicial system for medical liability cases has also been studied. While these reforms do not diminish the need for MICRA reforms at the state and federal level, if properly structured in collaboration with MICRA, they may further help curb skyrocketing medical liability premiums. Realistically these reforms could only be implemented at the state level and should be initiated as target pilot projects in select states to determine their efficacy.

CONCLUSION

We are all frustrated by the inability of policy makers to enact proven reforms. Physicians are frustrated because they are being forced to give up providing care for their patients due to the excesses of the legal system and liability insurance costs.⁶⁸ Patients are frustrated because they are losing access to care, frustrated when they are forced to find a new doctor, frustrated when they are forced to drive longer distances, and frustrated when they incur additional costs. Patients also are keenly aware of the impact of lawsuits on health care costs: Over seventy percent agree that medical liability litigation is driving up health care costs⁶⁹ and favor a law that would guarantee full payment for lost wages and medical expenses, but would limit non-economic damages.⁷⁰

That the system is out of balance is more than evident to anyone willing to look. Without action based on proven reforms and demonstrable data, the crisis will continue to spread. There are available solutions, but

67. See MISS. CODE ANN. § 11-1-60 (2004); W. VA. CODE § 55-7B-8 (2004). The AMA will closely watch the experience of West Virginia following its reforms because while the state has a base cap of \$250,000 on non-economic damages—with certain exceptions that could increase the cap to \$500,000 depending on the severity of the injury—the law also provides for annual adjustments up to \$375,000 (and \$750,000 depending on the injury severity).

68. DIV. OF MKT. RESEARCH & ANALYSIS, AM. MED. ASS'N, NATIONAL PHYSICIAN SURVEY ON PROFESSIONAL MEDICAL LIABILITY 3 (Apr. 2003).

69. HEALTH COALITION ON LIAB. & ACCESS, THE MEDICAL LIABILITY CRISIS: A FEDERAL PROBLEM THAT REQUIRES A FEDERAL SOLUTION (2003), <http://www.hcla.org/factsheets/2003-221-federalissue.pdf>.

70. Press Release, Health Coalition on Liab. & Access, Americans Say Health Care Access Threatened by Liability Crisis (Apr. 7, 2004), <http://www.hcla.org/pressreleases/2004-HCLAPollRelease.pdf>.

they will require policy makers willing to stand up for patients and change the status quo in the crisis states. California's MICRA provides a prime example of the type of reforms that are necessary if we are to fix the medical liability crisis that currently pervades the United States health care system. We must be relentless in our quest to fix our broken system. Failure to do so will cause irreparable harm not only to physicians, but also to the patients who depend on their care.

Turning from Damage Caps to Information Disclosure: An Alternative to Tort Reform

Kathryn Zeiler, M.S., J.D., Ph.D.*

With the 2004 U.S. presidential election close at hand, George W. Bush and his Administration resurrected a previously-killed federal proposal to cap medical malpractice damage awards.¹ The Bush Administration once again claimed that the United States is experiencing a medical malpractice insurance crisis and that frivolous medical malpractice lawsuits are the cause of this crisis.² According to the current Administration, large jury awards lead to significant increases in medical malpractice insurance premiums, driving physicians from the practice of medicine.³ Indeed, an array of policymakers continue to argue that

* Associate Professor of Law and Co-director of the Law and Economics Workshop Series at Georgetown University Law Center. The author thanks Victor Fleisher and Mitu Gulati for helpful comments and Joshua Ellis and Denise Shiu for excellent research assistance.

1. In March 2003, the House passed the Help Efficient, Accessible, Low Cost, Timely Healthcare (HEALTH) Act of 2003, H.R. 5, 108th Cong. (2003). The House version of the proposed legislation caps punitive damages at twice the economic damages or at \$250,000, whichever is greater, and limits attorney's fees in contingency cases. In July 2003, the Patients First Act of 2003, S. 11, 108th Cong. (2003) was proposed and defeated. If passed, the legislation would have placed an award cap of \$250,000 on non-economic damages and limited attorney's fees in contingency cases.

2. In a recent speech Vice President Richard Cheney argued that

Medical liability litigation is a serious problem in almost every state in the land, and it's not getting any better. Frivolous lawsuits are clogging the courts, and delaying justice for those with real problems. . . . We must protect the rights of those with real grievances, and we have to fix the medical liability problem at its source—the frivolous lawsuits that are filed solely with the hope of winning massive verdicts. That is why President Bush has set forth some responsible, practical reforms to put doctors and patients back in charge of healthcare in America The President has proposed a reasonable federal cap of \$250,000 on non-economic damage awards

Vice President Richard Cheney, Address to Dana Conference Center, Medical College of Ohio (July 19, 2004), <http://www.georgewbush.com/HealthCare/Read.aspx?ID=3006>.

3. *Id.* (quoting Cheney as arguing that “huge payoffs for personal injury trial lawyers”

damage caps will quell sharply increasing medical malpractice premiums, despite the fact that empirical evidence regarding the impact of damage caps on premiums is inconclusive.⁴

This Case Study argues that imposing statutory caps on medical malpractice damages is not an effective method of remedying the medical malpractice insurance crisis; therefore, policymakers should consider alternatives to damage caps. In particular, evidence suggests that implementing mandatory disclosure of the contract terms between managed care organization (MCOs) and physicians for the provision of services to enrollees reduces medical malpractice insurance premiums.

Part I of this Case Study reviews the controversy regarding the efficacy of damage caps in remedying medical malpractice insurance crises and discusses the state of empirical research investigating the effects of caps. Part II argues that a particular alternative—forcing disclosure of contract terms between MCOs and physicians—might more effectively reduce premiums. Policymakers interested in regulating medical malpractice insurance premiums should consider implementing MCO-physician contract disclosure requirements as a means to their desired end.

I. A POPULAR “SOLUTION”: MEDICAL MALPRACTICE DAMAGE CAPS

Several policymakers have proposed imposing medical malpractice damage caps to solve the current perceived medical malpractice insurance crisis.⁵ The proposals have reinvigorated a long-standing debate as to whether damage caps, in fact, significantly reduce medical malpractice premiums. Section A provides a short summary of the debate. In Section B, I discuss an important component of the analysis that has been largely missing from the debate: the effects of caps on treatment choices. Finally, in Section C, I briefly analyze the body of empirical research designed to study the relationship between caps, litigation, and medical malpractice insurance premiums.

and “massive increases in medical liability insurance premiums, for doctors across the country” have reached crisis proportions.) In the same address, Cheney claimed that imposing caps on damages would remedy the problem of increasing medical malpractice premiums. *Id.*

4. For a summary of the empirical literature studying the effects of damage caps on medical malpractice insurance premiums, see Kathryn Zeiler, *An Empirical Study of the Effects of State Regulations on Medical Malpractice Litigation Decisions* (July 2004) (unpublished manuscript, on file with author).

5. See *infra* Section I.A.

A. *The Controversy over Damage Caps*

Politicians and industry players claim that implementing medical malpractice damage caps will help end the medical malpractice insurance crisis. Recently, the media reported that “[damage] caps are being pushed nationally by Republicans including President Bush, who argue that excessive jury awards are largely responsible for escalating malpractice premiums.”⁶ The National Association of Mutual Insurance Companies (NAIC) supports medical malpractice caps on damages, arguing that caps would limit runaway jury awards of non-economic and punitive damages.⁷ Insurers also argue that caps reduce uncertainty, making it easier for them to set insurance premiums.⁸

On the other hand, opponents of damage caps argue that caps will not solve the medical malpractice crisis and that the cost of caps outweighs any potential benefits (if, indeed, they create benefits at all).⁹ Some commentators claim that caps are unconstitutional because they infringe on injured patients’ rights to trials by jury, to open courts, and to equal protection.¹⁰ Others note the potentially perverse effects of damage caps; for example, some claim that if caps are imposed, fewer legitimate medical malpractice cases might be filed because the costs of pursuing each claim might exceed expected awarded damages.¹¹ In addition, empirical evidence suggests that caps might lead to *larger* jury awards in some cases

6. John Wagner, *Doctors Wooed in Malpractice Insurance Fight*, WASH. POST, July 26, 2004, at B1.

7. Nat’l Ass’n of Mutual Ins. Cos., *Medical Malpractice Liability Reform*, at <http://www.namic.org/fedkey/04MedMal.asp> (last visited Nov. 10, 2004).

8. See W. Kip Viscusi, *Tort Reform and Insurance Markets*, 7 RISK MGMT. & INS. REV. 9, 20 (2004).

9. See *infra* notes 10-13 and accompanying text.

10. Ashley Stewart, Note, *Texas’ House Bill Four’s Noneconomic Damage Caps Impose the Burden of Supporting the Medical Industry Solely upon the Most Severely Injured and Therefore Most in Need of Compensation*, 57 SMU L. REV. 497, 503 (2004) (arguing also that damage caps “cause harm by preventing the most deserving victims from gaining compensation for their injuries”).

11. See ERIC NORDMAN ET AL., *MEDICAL MALPRACTICE INSURANCE REPORT: A STUDY OF MARKET CONDITIONS AND POTENTIAL SOLUTIONS TO THE RECENT CRISIS* 47 (2004), http://www.naic.org/models_papers/papers/MMP-OP-04-EL.pdf (draft report presented to the NAIC’s Property and Casualty Committee July 14, 2004) (arguing that “[s]ince the costs of researching and arguing a medical malpractice case can be very large, awards available once caps are introduced may not, in some cases, cover even the costs associated with pursuing a claim.”); see also Rachel Zimmerman, *As Malpractice Caps Spread, Lawyers Turn Away Some Cases*, WALL ST. J., Oct. 8, 2004, at A1.

because jurors might perceive the cap as the correct amount to award to all injured plaintiffs.¹² Similar perceptions by negotiating parties can also skew settlement outcomes in unexpected ways.¹³

Importantly, neither proponents nor opponents of caps have considered how caps might affect treatment choices made by physicians and managed care organizations and how these choices influence patient injury rates. In the following Section, I extend the boundaries of the debate by arguing that caps, at least theoretically, affect treatment choices, which in turn impact injury rates and medical malpractice claim rates.

B. The Missing Component: The Influence of Damage Caps on Treatment Choices¹⁴

While proponents of damage caps frequently argue that excessive litigation increases the practice of defensive medicine by physicians,¹⁵ critics of caps might similarly assert that limits on damages may also adversely affect treatment decisions.¹⁶ An examination of how caps influence the behavior of a wider array of health care market actors reveals that they can produce perverse incentives at the treatment decision stage.

12. See Jennifer K. Robbennolt & Christina A. Studebaker, *Anchoring in the Courtroom: The Effects of Caps on Punitive Damages*, 23 LAW & HUM. BEHAV. 353 (1999). Of course, these results might be important only in regimes in which juries are informed of statutory damage caps prior to deliberating about damages. Currently, only courts in Massachusetts are required to instruct the jury that, if it finds the defendant liable, it may not (in most cases) award more than the statutory limit for non-economic damages. See MASS. GEN. LAWS ANN. ch. 231, § 60H (West Supp. 1995). West Virginia allows the court to instruct the jury in this manner. See W. VA. CODE § 55-7B-8 (1994).

13. See Greg Pogarsky & Linda Babcock, *Damage Caps, Motivated Anchoring, and Bargaining Impasse*, 30 J. LEGAL STUD. 143 (2001).

14. For a complete analysis of the effects of damage caps on treatment choices, see Zeiler, *supra* note 4 (presenting theoretical predictions regarding how damage caps affect medical malpractice claim rates).

15. See U.S. DEP'T OF HEALTH & HUMAN SERVS., CONFRONTING THE NEW HEALTH CARE CRISIS: IMPROVING HEALTH CARE QUALITY AND LOWERING COSTS BY FIXING OUR MEDICAL LIABILITY SYSTEM 19 (July 2002) (arguing that "[t]he excesses of the litigation system are an important contributor to 'defensive medicine'"). Defensive medicine refers to the practice of providing patients with an inefficient amount of medical care to avoid exposure to liability for medical practice. For example, a physician might order an excessive number of diagnostic tests to be sure that she meets the legal standard of care when treating a particular patient. For a discussion of defensive medicine, see CHARLES E. PHELPS, HEALTH ECONOMICS 442-45 (3d ed. 2003).

16. To the best of my knowledge, this has not been argued in the past by those who oppose caps on damages.

Consider, first, how damage caps might influence the decision of an injured patient (or her attorney who likely is employed by the patient on a contingency-fee basis) regarding whether to file a medical malpractice claim against her physician. Assume that a patient will file a claim only if expected damages exceed litigation costs.¹⁷ Estimates of expected damages depend on two variables: the anticipated damage award and the probability that the patient will succeed in recovering this amount from the physician. The probability of success (whether by court award or through settlement), in turn, depends on the likelihood that the physician provided non-compliant medical care. All other things being equal, as the likelihood that the physician provided non-compliant treatment increases, the probability of recovering damages increases, as does the probability that an injured patient will file a claim.¹⁸

The next step in the analysis is to consider how caps affect the probability that a physician will provide non-compliant treatment. In theory, when deciding whether to provide costly compliant treatment, the physician (in conjunction with the patient's MCO) weighs the costs and the benefits of providing such care.¹⁹ Costs refer to all the expenses incurred in providing compliant care; the benefits include the reduction in exposure to liability for medical malpractice. Damage caps reduce the exposure to liability; therefore, the imposition of caps makes it optimal, in some cases, for physicians (or MCOs) to face potential liability for medical malpractice rather than provide costly treatment that complies with the legal standard of care. Recent research does, in fact, indicate that

17. In other words, assume injured patients act perfectly rationally when deciding whether to sue for medical malpractice. Of course, in some cases, these decisions may be driven by factors other than the expected monetary costs and benefits of filing a claim (e.g., emotions, revenge and strategic behavior). However, since patients must convince lawyers to take on these cases in exchange for a cut of the pie, it is unlikely that filing decisions are driven significantly by emotional factors. In addition, physicians tend not to cave easily to patients' demands based on nonmeritorious claims because they highly value their reputations (and would risk sanctions). See Linda Oberman, *IG Asks Why More Hospitals Don't Report Adverse Actions*, AM. MED. NEWS, Feb. 13, 1995, at 4 (claiming that reputation effects, in part, drive physician reluctance to settle medical malpractice cases).

18. To obtain this result we need only assume that the court is better at verifying whether the physician provided negligent treatment than it would be if it flipped a fair coin. This assumption does not seem unreasonable.

19. For a detailed analysis of the physician's treatment choice, see Kathryn Zeiler, *Medical Malpractice and Contract Disclosure: An Equilibrium Model of the Effects of Legal Rules on Behavior in Health Care Markets 16-18* (Apr. 2004) (unpublished manuscript, on file with author).

physicians react to different sorts of financial incentives in this way.²⁰

To summarize, if damage caps reduce exposure to liability, physicians (and MCOs), on average, may be less likely to provide compliant treatment. This will result in an increase in patient injuries, and in turn, an increase in the number of injured patients who file claims for medical malpractice.²¹

This increase in the claims rate, coupled with the potential decrease in the average damage award (and settlements) due to the cap, is likely to yield indeterminacy: Because of these competing forces the influence of caps on ex ante calculations of expected damages from medical malpractice claims (and therefore medical malpractice insurance premiums) will depend on other variables, such as the cost of treatment relative to expected damages, the probabilities of injuries given compliant and non-compliant treatment, and the amount of the cap.²² Therefore, the claim that caps will decrease medical malpractice insurance premiums is arguably shortsighted because it does not account for the influence of caps on influence treatment choices.

20. While the effects of tort reform on treatment choices have not been studied empirically to date, some have investigated the effects of financial incentives on treatment choices and find that physicians do respond to financial incentives. See, e.g., Thomas S. Crane, *The Problem of Physician Self-Referral Under the Medicare and Medicaid Antikickback Statute*, 268 JAMA 85, 86 (1992) (citing government studies indicating that physicians respond to financial incentives in their treatment practices); David Hemenway et al., *Physicians' Responses to Financial Incentives: Evidence from a For-Profit Ambulatory Care Center*, 322 NEW ENG. J. MED. 1059, 1062 (1990) (showing that physicians react to bonus arrangements that reward them for ordering laboratory tests by significantly increasing the number of tests they order). In addition, studies have found that physicians who report that their contracts with MCOs include incentives to reduce referrals were "more likely than others to have felt pressure to limit referrals in a manner that compromised care." See Kevin Grumbach et al., *Primary Care Physicians' Experience of Financial Incentives in Managed-Care Systems*, 339 NEW ENG. J. MED. 1516 (1998).

21. It is important to note that physician exposure to liability likely will influence MCO behavior. If the costs of practicing medicine increase due to increased exposure liability, then physicians will demand more in compensation from MCOs. MCOs can influence physician treatment choices directly by approving or denying reimbursement for treatments and indirectly through financial incentives written into MCO-physician contracts. Therefore, changes in physicians' exposure to liability will influence MCO-physician contracts and MCO decisions regarding whether to approve particular treatments.

22. See Zeiler, *supra* note 4, for a complete characterization of the equilibria under various conditions. Despite the indeterminacy, "unless the cap is so restrictive that total damages fall below litigation costs, caps are likely to cause an increase in ex ante expected damages." *Id.* at 13.

With a more complete understanding of how damage caps might influence the choices of health care market actors, we are positioned to evaluate the empirical results, produced using field data, to study the effects of caps on medical malpractice insurance and litigation behavior. The following Section summarizes the state of the empirical literature and argues that the cumulative findings do not allow us to draw conclusions regarding how damage caps influence medical malpractice insurance markets or litigation behavior.

C. Empirical Evidence

Several researchers have employed field data to investigate whether caps significantly influence medical malpractice insurance premiums and losses incurred by insurers.²³ A review of this empirical literature reveals two general themes. First, the empirical results generally are mixed. Second, given the difficulties in directly measuring the influence of caps, reliance on the results of most studies is controversial.

Results vary significantly depending on the data employed, the specifications of the empirical models, and the time periods studied. For example, Professor Frank Sloan investigated the influence of damage caps on premiums paid by physicians in three specific fields.²⁴ The study incorporated data for the years 1974-1978. Using regression analysis, he found that damage caps significantly affected neither premiums nor annual percentage change in premiums for any of the three fields tested. Professor Kip Viscusi and his colleagues focused mainly on the effects of the second generation of tort reforms to be implemented by state legislators.²⁵ Using 1988 aggregated premiums by state, they considered the change in premiums from 1985 to 1987. The analysis controlled for differences in state regulation of insurers. The authors considered limits on non-economic damages and limits on punitive damages and, like Frank, found that limits on non-economic damages did not significantly affect premiums.

Conversely, Stephen Zuckerman of the Urban Institute and his

23. See Zeiler, *supra* note 4 (reviewing the empirical literature).

24. Frank A. Sloan, *State Responses to the Malpractice Insurance "Crisis" of the 1970s: An Empirical Assessment*, 9 J. HEALTH POL. POL'Y & L. 629, 637-643 (1985).

25. See, e.g., W. K. Viscusi et al., *The Effect of 1980s Tort Reform Legislation on General Liability and Medical Malpractice Insurance*, 6 J. RISK & UNCERTAINTY 165, 186 (1993) (analyzing the effects of reforms such as modifications of joint and several liability, limits on liability and establishments of immunities, limits on noneconomic and punitive damages and provisions for structured and periodic payments of damage awards).

colleagues found that damage caps significantly reduced medical malpractice insurance premiums.²⁶ The study uses data covering a thirteen-year period—1974 through 1986—and including data for most states. Likewise, Professor Vasanthakumar Bhat examined the influence of damage caps on several indicators including claim rates, severity of claims, and premiums.²⁷ Using data on the payment rate per physician of each state for the period 1991-1995, Bhat found that caps on economic and non-economic damages, taken together, significantly decreased premiums. However, similar to the results of some others, Bhat found that caps on non-economic damages, considered alone, had no effect on premium levels.

These mixed empirical results are most likely due to the difficulties in measuring the influence of caps on medical malpractice insurance and litigation.²⁸ These difficulties arise for a variety of reasons. First, isolating the effects of caps in the presence of other sorts of tort reform is complicated. Second, the uncertainty generated by legal challenges claiming, for example, that reforms are unconstitutional makes it difficult to measure the direct effect of caps on insurance and litigation.²⁹ Third, the availability of data is limited and the data that is available presents challenges in the design of empirical studies. For instance, data on actual losses paid aggregated by state are generally unavailable, although some researchers have obtained data of this sort directly from insurers.³⁰ Furthermore, using proxies for losses paid (e.g., losses incurred) presents additional concerns, including the danger that accounting adjustments might substantially reduce the correlation between losses paid and the proxy.³¹ Specifically, the managers of insurance companies have incentives

26. Stephen Zuckerman et al., *Effects of Tort Reforms and Other Factors on Medical Malpractice Insurance Premiums*, 27 INQUIRY 167 (1990).

27. VASANTHAKUMAR N. BHAT, MEDICAL MALPRACTICE: A COMPREHENSIVE ANALYSIS (2001).

28. CONG. BUDGET OFFICE, THE EFFECTS OF TORT REFORM: EVIDENCE FROM THE STATES (June 2004) (discussing the difficulties in evaluating the results of the empirical studies on the factors discussed in this Section of the Case Study).

29. See Heidi Li Feldman, *Harm and Money: Against the Insurance Theory of Tort Compensation*, 75 TEX. L. REV. 1567, 1568 n.4 (1997) (discussing cases in which tort reforms have been held unconstitutional on theories of violation of equal protection, violation of right to court access and violations of rights to due process).

30. See Albert Yoon, *Damage Caps and Civil Litigation: An Empirical Study of Medical Malpractice Litigation in the South*, 3 AM. L. & ECON. REV. 199, 209 (2001).

31. See Zeiler, *supra* note 4 (analyzing data on losses incurred gathered by the National Association of Insurance Commissioners). Of 550 observations of state-level losses incurred

to manipulate the reserves to manage the bottom lines of their companies. Thus, if we observe lower incurred losses in regimes which cap damages, this may not reflect the true effects of the caps, but rather the effects of earnings manipulations that might be unrelated to manager expectations regarding future losses. In fact, managers might use the fact that damage caps are in place to justify decreasing reserves. Finally, as cogently explained by Professor Albert Yoon, employing simple regressions to study the effects of damage caps on medical malpractice insurance premiums can be problematic if the implementation of caps is endogenous to perceived market conditions related to premiums. In other words, if caps are implemented in response to rising medical malpractice insurance premiums, then it becomes difficult to measure the influence of the caps on premiums.³² This problem is referred to as endogeneity³³

In a recent study, Yoon demonstrated empirical modeling techniques that can be used in the face of potential endogeneity.³⁴ By using a difference-in-difference approach, Yoon was able to account for conditions present before and after the implementation and repeal of damage caps. Employing this technique and others to control for various additional modeling concerns, Yoon found that caps decreased the average relative recovery by medical malpractice claimants. This study offers an important step toward determining the actual effects of caps on insurance and litigation. The results, however, do not allow us to make claims about whether total losses paid out to claimants increase or decrease when caps are imposed because the results do not provide insights into how caps

(fifty states over eleven years), thirteen observations are negative. *Id.* This suggests that adjustments to reserves might swamp losses incurred and reduce the correlation between losses incurred and losses paid. *See also Patient Access Crisis: The Role of Medical Litigation*, House Comm. on the Judiciary, 108th Cong. 2-3 (2003) (statement of Jay Angoff, Counsel, Roger G. Brown & Associates) (reporting testimony describing insurance company manager incentives to inflate or understate estimates of losses incurred).

32. *See* Yoon, *supra* note 30, at 202 (arguing that if “the policy is a codification of underlying conditions . . . that actually caused the policy to be implemented in the first place,” then determining the causal connection between the implementation of caps and indicators, such as losses, becomes difficult).

33. For an explanation of the problems resulting from endogeneity, *see* STEPHEN J. SCHMIDT, *ECONOMETRICS* 263-81 (2004). *See also* HOWELL E. JACKSON ET AL., *ANALYTICAL METHODS FOR LAWYERS* 565-566 (2003) (illustrating endogeneity, which they refer to as “two-way causation,” by pointing out that it is difficult to understand the influence of increasing the number of police on crime rates because more police tend to be sent to particular areas: namely, those with high crime rates).

34. *See* Yoon, *supra* note 32, at 203.

influence the number of patient injuries and the number of claims filed. If caps result in an increase in the number of claims filed, then despite the fact that average recoveries decrease, caps could increase the total losses paid.

Despite the fervent push to implement damage caps as a solution to the medical malpractice insurance crisis,³⁵ more research clearly is needed to determine how caps and other sorts of tort reform actually affect behavior in health care markets. Not only is more empirical research necessary, but also it is important that the empirical research be grounded in sound theoretical models of the effects of tort reform on behavior in health care markets. In the meantime, turning our attention to other possible remedies might prove useful. Part II presents an alternative remedy yet to be addressed by policymakers.

II. AN ALTERNATIVE REMEDY: MANDATORY CONTRACT DISCLOSURE

As of 2001, twenty-one states required MCOs to disclose to their enrollees or prospective enrollees the terms of their contracts with physicians.³⁶ While the goal of forcing contract disclosure is simply to provide information to consumers during the health plan selection process,³⁷ evidence suggests that disclosure of contract terms might result in lower medical malpractice insurance premiums.

The relationship between contract disclosure and medical malpractice insurance premiums is not intuitive. To understand the relationship, one must consider how the revelation of MCO-physician contract terms influences two types of decisions: litigation decisions made by injured patients and contract decisions made by MCOs.³⁸

First, consider how contract disclosures affect whether an injured patient pursues a medical malpractice claim against her physician. As discussed above, patients considering whether to file a medical malpractice

35. See *supra* Section I.A.

36. For a list of states that force disclosure of contracts between MCOs and physicians, see Zeiler, *supra* note 4.

37. Cf. William M. Sage, *Regulating Through Information: Disclosure Laws and American Health Care*, 99 COLUM. L. REV. 1701, 1825 (1999).

38. For a more complete game-theoretic analysis of the influence of mandatory contract disclosure rules on MCO-physician contracts, compliant treatment rates and medical malpractice litigation rates, see Zeiler, *supra* note 19; see also, Kathryn Zeiler, *Medical Malpractice and Contract Disclosure: An Equilibrium Model of the Effects of Legal Rules on Behavior in Health Care Markets* (Apr. 2004) (Working Paper presented to the Am. L. & Econ. Ass'n Annual Meeting, 2004).

claim frequently must do so under conditions of imperfect information. The injured patient is not always able to observe whether her injury was truly caused by negligent behavior on the part of her physician.³⁹ Injured patients (and their attorneys) benefit from information that helps to resolve this uncertainty when deciding whether to pursue costly litigation.⁴⁰

MCO-physician contract terms are just this sort of information. In theory, injured patients should be able to update their prior beliefs about whether the physician acted negligently by considering the contract terms. For example, if the patient observes that the MCO and the physician agreed to a traditional fee-for-service arrangement (i.e., the physician is reimbursed a fee by the MCO for each particular medical service provided), then the injured patient might be more likely to believe that expensive compliant treatment was provided than if the MCO and physician agreed to a capitated arrangement (i.e., the MCO pays the physician a fixed dollar amount per patient per month and the physician pays for overruns out of his own pocket).⁴¹

Still, how does a change in the way potential litigants make decisions about whether to file claims lead to lower medical malpractice insurance premiums? The next step in the analysis is to consider how behavior at the litigation stage affects contract choices. MCOs design physician contracts to provide incentives for physicians to choose treatments that maximize MCO profits. In regimes that force contract disclosure, MCOs must consider not only how the contract terms shape physician treatment decisions, but also how the contract terms will influence litigation decisions by injured patients. By observing disclosed contract terms, patients are able to update their beliefs about the likelihood that they

39. Whether the patient is able to infer that the injury was caused by the negligent actions of the physician or MCO depends on two probabilities: the probability that non-negligent treatment results in injury and the probability that negligent treatment results in injury. If these probabilities fall somewhere between zero and one, but are not equal to zero or one, then the patient will be uncertain about whether the injury resulted from negligent treatment. These probabilities, of course, will differ from case to case and will depend on the nature of the treatment, the characteristics of the patient and other such factors. See Zeiler, *supra* note 14.

40. This general concept is not novel: When principals are not able to observe behavior, they often turn to other sources of information, such as the number of hours the agent worked or whether the agent seemed intoxicated. See EDWARD P. LAZEAR, PERSONNEL ECONOMICS (1995).

41. See Zeiler, *supra* note 19, at 21-29 (predicting the manner in which MCOs will employ various contract types to influence physician behavior given a particular legal regime).

received compliant treatment. In fact, MCOs benefit from disclosing because they can use disclosures to signal the provision of compliant treatment and potentially reduce the number of claims filed.⁴² In other words, when injured patients receive the signal that compliant treatment was provided, they are less likely to file a costly medical malpractice claim because the likelihood of succeeding is low. Therefore, when patients can observe contract terms it is more likely that the MCO, when comparing the cost of compliant treatment to the expected damage award, will find it optimal to employ particular contract terms to encourage the physician to provide compliant treatment. This is because, in those cases, the cost of providing the level of compliant treatment necessary to ensure that very few medical malpractice claims are filed is less than the reduction in exposure to liability that results from the increase in the provision of compliant treatment.⁴³ As a result, in regimes that mandate contract disclosure, MCOs are more likely to use contract terms that encourage physicians to provide compliant treatment and patients are less likely to file medical malpractice claims.⁴⁴

Initial empirical tests of the theoretical predictions regarding how mandatory contract disclosure rules affect medical malpractice insurance premiums support claims that disclosure rules decrease ex ante expected damages arising from medical malpractice claims. A study using data on medical malpractice insurance premiums per physician in the fifty U.S.

42. That we do not observe MCOs voluntarily disclosing physician contract terms is most likely due to the fact that disclosure of this information is costly. Not only is the disclosure itself costly to produce, but an MCO might lose its competitive advantage if it discloses information about innovative contract terms that create efficiencies not enjoyed by competing MCOs. See Zeiler, *supra* note 4, at 28-29. In addition, that we do not observe consumers demanding disclosure of contractual arrangements might be due to market failures. See Bruce C. Greenwald & Joseph E. Stiglitz, *Externalities in Economies with Imperfect Information and Incomplete Markets*, 101 Q. J. ECON. 229 (1986). Market failures abound in health care insurance markets. In particular, given that a substantial number of consumers obtain their health insurance through their employers, most consumers of health insurance are not involved in the bargaining process. See KAISER FAMILY FOUND., EMPLOYER HEALTH BENEFITS 2004 ANNUAL SURVEY (2004), <http://www.kff.org/insurance/7148/summary/index.cfm> (reporting that “[e]mployer-sponsored health insurance reaches more than three out of every five nonelderly Americans”). In addition, employers’ interests are not necessarily aligned with the interests of their employees.

43. See Zeiler, *supra* note 19, at 21-31 for a detailed explanation of this result.

44. For a complete analysis of how mandatory contract disclosure rules lead to more compliant treatment and less litigation, see Zeiler, *supra* note 19, at 29-31 for a detailed explanation of this result.

states for the period 1991-2001 provides some support for the prediction that mandatory disclosure rules decrease ex ante expected damages from medical malpractice claims.⁴⁵ The empirical results indicate that medical malpractice insurance premiums are lower in states that force disclosure of contract terms.

Mandatory disclosure has some potential drawbacks as well. For example, by forcing MCOs to disclose information about physician contracts they are, in essence, forced to reveal trade secrets.⁴⁶ Innovative physician contracts arguably afford MCOs the opportunity to obtain an advantage over competitors. This benefit provides an incentive for MCOs to design creative, efficient physician contracts, an endeavor advantageous not only for the MCO but also for enrollees who enjoy lower prices and/or higher quality. Forcing MCOs to disclose information about these contracts might diminish the incentive to expend resources to develop innovative physician contracts. It is important to weigh these disadvantages against the benefits gained or consider ways to work around them before implementing such policies.

CONCLUSION

This Case Study focuses on the “best way” to address or improve the current state of malpractice insurance. Given the complexity of the industry, the solution likely will be complex itself, as it must address

45. See Zeiler, *supra* note 4, at 19-22, 24-26 (reporting regression results indicating that, under reasonable specifications, mandatory contract disclosure leads to lower medical malpractice insurance premiums). The effect of mandatory disclosure rules, however, becomes statistically insignificant when assuming (1) that a lag exists between the time statutes are passed and insurance rates reflect the new rule and (2) that current year premiums depend on prior year premiums. *Id.* At 24-26. While these empirical results provide some support for the theoretical prediction that mandatory disclosure rules lead to lower medical malpractice insurance premiums, it is important to note that these are preliminary findings. Further investigation is required before we can recommend policy prescriptions.

46. HMOs have argued that required disclosure of physician contracts is unfair because the contracts are trade secrets. They claim that contracts are the result of much time and effort spent negotiating with physicians, and forced disclosure will allow competitors to unfairly take advantage of the end product without contributing to the costs. See *Wilmington Star-News v. New Hanover Reg'l Med. Ctr.*, 480 S.E.2d 53, 56 (N.C. Ct. App. 1997) (discussing whether pricing information in HMO contracts constitutes a trade secret). Forcing disclosure of physician incentives might create an economic disincentive to expend resources constructing innovative incentive arrangements. It is important to consider this when evaluating whether mandatory disclosure is socially optimal.

information asymmetries, agency problems, the negative effects of adverse selection, and various other market imperfections resulting from the structure of health care markets. Likely, no one remedy will be a panacea. In addition, seemingly intuitive remedies often produce unintended, perverse effects.

The main point of this Case Study is to argue that damage caps, while a seemingly intuitive fix, might not be the cure-all touted by politicians and industry actors. Deeper analyses of the effects of caps reveal that they might affect health care markets in ways that make matters worse. In addition, given the nature of the inquiry and inherent methodological problems, we cannot draw strong conclusions from the body of empirical studies that investigate the effect of caps on medical malpractice insurance premiums.

Given these difficulties, we should focus on alternatives to damage caps. One such alternative—mandating disclosure of MCO-physician contract terms—appears promising. An analysis considering how the market will react to the mandate indicates that forcing disclosure will lead to lower medical malpractice insurance premiums. Preliminary empirical evidence suggests that it is worthwhile to explore this remedy further.

Patchwork remedies and politically-driven policies likely will not ameliorate the negative consequences of health care market imperfections. If we have any hope of structuring and regulating health care markets so as to reduce the probability of experiencing various sorts of crises, we must step back and take a comprehensive look at how market actors will adjust to regulations and how various regulations interact with one another.

ANNIVERSARY ESSAYS

INTRODUCTION:

- 401 Fifth Anniversary Essay Collection**
Jed Adam Gross

RESPONSES:

- 403 Gene Patents: The Need for Bioethics Scrutiny and Legal Change**
Lori B. Andrews, J.D. and Jordan Paradise, J.D.
- 413 World Health Law: Toward a New Conception of Global Health Governance for the 21st Century**
Lawrence O. Gostin, J.D., LL.D.
- 425 The Challenge of Assuring Continued Post-Trial Access to Beneficial Treatment**
Christine Grady, R.N., Ph.D.
- 437 The Most Important Health Care Legislation of the Millennium (So Far): The Medicare Modernization Act**
Timothy Stoltzfus Jost, J.D.
- 451 The Swinging Pendulum: The Supreme Court Reverses Course on ERISA and Managed Care**
Aaron S. Kesselheim, M.D., J.D. and Troyen A. Brennan, M.D., J.D., M.P.H.

- 465 **Inequality, Infections, and Community-Based Health Care**
 Evan Lyon, M.D. and Paul Farmer, M.D., Ph.D.
- 475 **Medicare Reform and Social Insurance: The Clashes of 2003 and
 Their Potential Fallout**
 Theodore R. Marmor, Ph.D. and Jacob S. Hacker, Ph.D.
- 491 **The Pedagogical Significance of the Bush Stem Cell Policy: A
 Window into Bioethical Regulation in the United States**
 O. Carter Snead, J.D.

Fifth Anniversary Essay Collection

Jed Adam Gross

This issue of the *Yale Journal of Health Policy, Law and Ethics* marks our fifth year of publication. Our anniversary is a time to celebrate, but also a time for reflection. The past half-decade has brought dramatic developments in medicine and health policy—from the debate over embryonic stem cell research to the intensification of global public health challenges to a major overhaul of Medicare. At the same time, we continue to wrestle with stark health disparities, enduring ethical dilemmas, and complex legal precedents.

To commemorate our fifth anniversary, we invited a number of distinguished scholars to survey the most important recent developments in health policy, law, and ethics and to project which health issues will be of particular importance in the years ahead. Their responses were all that we could have hoped for: an engaging catalogue of wrong turns and new beginnings, opportunities missed and opportunities seized, perils and possibilities, sources of consternation, aspiration, and inspiration. We hope you find this collection both informative and thought-provoking.

Gene Patents: The Need for Bioethics Scrutiny and Legal Change

Lori B. Andrews, J.D.* and Jordan Paradise, J.D.†

In May 2004, the European Patent Office dealt a serious blow to gene patents by revoking Myriad Genetics's controversial patent on the BRCA1 gene.¹ That patent covered any method of diagnosing a predisposition for breast or ovarian cancer that used the BRCA1 gene sequence.² Elsewhere, gene patents are also being challenged in courtrooms,³ legislatures,⁴ and in the arena of public opinion. Numerous international organizations, such as the Council of Europe's Committee on Legal Affairs and Human Rights and UNESCO, view genes as belonging to the common heritage of mankind.⁵ Intense opposition to gene patents is also coming from

* Distinguished Professor of Law, Chicago-Kent College of Law and Director, Institute for Science, Law and Technology, Illinois Institute of Technology. The research for this Essay was supported by grant #DE-FG02-02ER63460, from the Office of Biological and Environmental Research, the Office of Science, U.S. Department of Energy, and The Robert Wood Johnson Foundation Investigator Awards in Health Policy Research Program.

† Legal Fellow with the Institute for Science, Law and Technology at the Illinois Institute of Technology.

1. Press Release, Eur. Patent Office, "Myriad/Breast Cancer" Patent Revoked After Public Hearing (May 18, 2004), http://www.european-patent-office.org/news/pressrel/2004_05_18_e.htm. This type of legal challenge is called an "opposition" to a granted patent under European patent law and allows third parties to challenge a patent's validity within nine months after it is granted. See European Patent Convention, art. 99 (1998), <http://www.european-patent-office.org/legal/epc/>. The revocation was made by the Opposition Division, a panel of three patent examiners and one legal expert, applying current law as set forth in the European Patent Convention. Decision Revoking the European Patent (Art. 102(1), (3) EPC) (Eur. Pat. Office May 17, 2004) (revoking European Patent No. 0699754) [hereinafter Revocation Decision].

2. E.U. Patent No. EP0699754 (issued Mar. 6, 1996).

3. See *Greenberg v. Miami Children's Hosp.*, 264 F. Supp. 2d 1064 (S.D. Fla. 2003).

4. See Genomic Research and Diagnostic Accessibility Act of 2002, H.R. 3967, 107th Cong. (2002).

5. *Protection of the Human Genome by the Council of Europe*, Council of Eur., Comm. on Legal Affairs & Human Rights Doc. 9002 (Mar. 19, 2001), <http://assembly.coe.int/Documents/WorkingDocs/doc01/EDOC9002.htm>; *Universal Declaration on the Human*

researchers,⁶ politicians,⁷ organized religions,⁸ indigenous groups,⁹ patient groups,¹⁰ and medical professional organizations.¹¹ Patents covering human genetic material raise a variety of issues related to legal appropriateness, scientific and medical research, and access to health care, as well as issues regarding privacy, autonomy, religious freedom, and reproductive liberty. While there are reasons to celebrate many new developments in medicine and bioethics, patents for human genetic material are an example of a bad policy that needs to be corrected. Gene patents raise bioethical concerns because they can impede access to appropriate health care and violate individual rights.

I. THE UNCOMFORTABLE FIT BETWEEN GENES AND PATENTS

Over two centuries ago, the framers of the U.S. Constitution realized that it was important to create incentives for technological innovation.¹² In return for a patent, the inventor must show the invention satisfies a number of requirements, including a sufficient written description, as well as utility, novelty, and nonobviousness.¹³ Yet not all inventions are patentable. For example, products of nature are not patentable.¹⁴

Genome and Human Rights, UNESCO Gen. Conference (Nov. 11, 1997), adopted by G.A. Res. 152, U.N. GAOR, 53d Sess., U.N. Doc. A/RES/53/152 (1999).

6. Declan Butler & Sally Goodman, *French Researchers Take a Stand Against the Cancer Gene Patent*, 413 NATURE 95, 95 (2001).

7. See, e.g., Paul Willcocks, *Canadian Premiers Wade into Gene Patenting Debate*, REUTERS, Aug. 3, 2001.

8. See, e.g., Fred B. Charatan, *U.S. Religious Groups Oppose Gene Patents*, 310 BRIT. MED. J. 1351, 1351 (1995); see also Southern Baptist Convention, Resolution on the Patenting of Animal and Human Genes (June 1995), <http://www.sbc.net/resolutions/amResolution.asp?ID=570>.

9. See, e.g., Letter from Debra Harry, Indigenous Peoples Council on Biocolonialism, to Commissioner of Patents & Trademarks (Mar. 21, 2000), <http://www.uspto.gov/web/offices/com/sol/comments/utilguide/ipcb.pdf>.

10. See, e.g., *Greenberg v. Miami Children's Hosp.*, 264 F. Supp. 2d 1064 (S.D. Fla. 2003).

11. See Ass'n for Molecular Pathology, Clinical Practice Committee, AMP Position on Patenting of Genetic Tests (Nov. 22, 1999), <http://www.ampweb.org/PRC/prc-tests.htm>.

12. The Constitution gives Congress the power to "promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." U.S. CONST. art. I, § 8, cl. 8.

13. See 35 U.S.C. §§ 100-05 (2000).

14. *Diamond v. Chakrabarty*, 447 U.S. 303, 309 (1980) (citing *Funk Brothers Seed Co. v. Kalo Inoculant Co.*, 333 U.S. 127, 130 (1948)).

How is it then that genes are patentable? Applicants who seek human gene patents assert that they have isolated and purified a gene or genetic material, thereby producing something new—a product whose non-coding regions have been eliminated, but which still performs the same function as a naturally-occurring gene.¹⁵ While some courts have held isolated and purified products of nature to be patentable,¹⁶ the useful properties of a gene—such as its ability to bind to another complementary strand of DNA for diagnosis or its ability to code for a particular protein—are not ones that the scientist has invented, but rather are natural, inherent properties of genes themselves.¹⁷ Often gene patent holders lay claim to gene segments that actually occur in nature and exist within the bodies of human beings.¹⁸ In fact, one Australian company has acquired global patent protection over non-coding regions of the human genome, amassing millions of dollars in licensing deals with drug companies and universities for the right to use this information in research and drug development.¹⁹

The patent system is generally designed to incentivize research and innovation, but there are many other incentives for the discovery of

15. Sheldon Krinsky, *The Profit of Scientific Discovery and Its Normative Implications*, 75 CHI.-KENT L. REV. 15, 26 (1999).

16. *E.g.*, *Parke-Davis & Co. v. H. K. Mulford & Co.*, 196 F. 496 (2d Cir. 1912) (upholding a patent on adrenaline, a natural hormone that was found in animal glands). In *Parke-Davis*, the patent applicant identified, isolated, and purified the active ingredient—adrenaline—creating a product that did not exist in nature in that precise form and that could be used for medical treatment. The U.S. Supreme Court's subsequent *Chakrabarty* decision that allowed a patent on genetically-modified bacteria dealt with a new invention—a genetically engineered life form invented by combining genes in ways that did not occur in nature. *Diamond v. Chakrabarty*, 447 U.S. 303 (1980).

17. See Am. Coll. of Medical Genetics, Position Statement on Gene Patents and Accessibility of Gene Testing (Aug. 2, 1999), at <http://www.acmg.net/resources/policies/pol-015.asp>; Coll. of Am. Pathologists, Gene Patents Detrimental To Care, Training, Research, at http://www.cap.org/apps/docs/advocacy/advocacy_issues/Issue_Genepat.html (last visited Nov. 21, 2004).

18. See U.S. Patent No. 5,679,635 (issued Oct. 21, 1997) (claiming the genetic sequence of the aspartoacylase gene and protein).

19. Carina Dennis, *Geneticists Question Fees for Use of Patented 'Junk' DNA*, 423 NATURE 105, 105 (2003). That Australian company has entered into licensing agreements with a number of large biotech corporations for between \$250,000 and \$1 million each. *Malcolm in the Middle*, BIOIT WORLD, Aug. 13, 2003, <http://www.bio-itworld.com/archive/081303/firstbase.html>. In the United States, two patents covering these non-coding regions are U.S. Patent No. 5,851,762 (issued Dec. 22, 1998) and U.S. Patent No. 5,612,179 (issued Mar. 18, 1997).

genetic sequences. Molecular biologists were attempting to identify genes long before patents were awarded for genetic material. When biologists began the Human Genome Project, they had no idea they would be able to patent genes;²⁰ they had other reasons to search for genes, namely medical interests and the potential for academic advancement and status.²¹

The discovery of genes does not require the same commercial incentives as drug development. The development of drugs is undertaken primarily with private funds (for which investors expect a commercial return),²² while the discovery of genes has been undertaken with vast quantities of public funds. For example, national governments and non-profit institutions spent over \$1.8 billion of taxpayers' money on genomics in 2000.²³ Myriad, the U.S. genetics company that first patented BRCA1, used over five million dollars from a government agency when researching the patent²⁴ and utilized sequence data from public databases. Thus, if gene patents continue, the public will pay twice—first for the research and second for the high royalty costs that many patent holders require for subsequent use of their patented gene in a product.

Unlike drug development, gene discovery does not require expensive clinical trials and approval from the Food and Drug Administration. Testing for mutations in a disease gene can begin almost immediately after the gene has been identified.²⁵ Thus, the need to provide financial compensation to a gene-discoverer through gene patent royalties is not as great as the need to compensate the developer of a drug that must undergo costly clinical trials, especially since only a small number of drugs actually become commercially-viable products.

Moreover, there are fewer drawbacks to granting a patent on a drug or a medical device than granting a patent on a gene. For instance, other

20. See Leslie Roberts, *Controversial from the Start*, 291 SCIENCE 1182, 1185-86 (2001).

21. See *id.* at 1182.

22. PHARM. RESEARCH & MFRS. OF AM., PHARMACEUTICAL INDUSTRY PROFILE 2004, at 7 (2004), <http://www.phrma.org/publications/publications//2004-03-31.937.pdf>. The high cost of bringing a drug to market includes the salaries for research and development scientists, the expense of animal research and human clinical trials, and the cost of obtaining FDA approval. See Joseph DiMasi et al., *The Price of Innovation: New Estimates of Drug Development Costs*, 22 J. HEALTH ECON. 151, 152 (2004).

23. Stanford in Washington, Genomics Research Funding 1998-2000, at <http://www.stanford.edu/class/siw198q/websites/genomics/Gov&nonprofitTotal.htm> (last modified Dec. 7, 2001).

24. Bryn Williams-Jones, *History of a Gene Patent: Tracing the Development and Application of Commercial BRCA Testing*, 10 HEALTH L.J. 123, 131 (2002).

25. Jon F. Merz et al., *Diagnostic Testing Fails the Test*, 415 NATURE 577, 577 (2002).

researchers can create alternatives to drugs and devices. In contrast, there are no alternatives to use of the patented human genes for genetic diagnosis and gene therapy.²⁶

II. GENE PATENTS CREATE PROBLEMS FOR ACCESS TO APPROPRIATE HEALTH CARE

Under patent law, the patent holder has the right, for twenty years from the date of the application filing, to prevent any other individual or institution from making, using, offering to sell, or selling the invention.²⁷ The patent holder can choose to license the patented invention to others, can choose to use the patented invention exclusively itself, or can choose to prevent any use of the patented invention by itself or by others. In the gene patent area, the exclusive rights of the patent holder can raise the costs of genetic services, diminish the quality of genetic tests and treatments, and interfere with access to health care.

In some cases, gene patent holders will only let their own laboratories use the test for the patented gene. Exclusive licensing of a gene patent can itself interfere with the development of diagnostics. Various mutations in the same gene can cause a particular disease, but companies that do not let anyone else test for “their” gene make it more difficult for the discovery of other significant mutations in that gene. In countries where the Alzheimer’s gene and hemochromatosis gene were not patented, researchers were able to discover previously unknown mutations.²⁸ These additional mutations are often critical tools for diagnosing individuals who would not otherwise be diagnosed by the patented gene or diagnostic test.

The possibility of inappropriate diagnostics was part of the concern that prompted the French challenge to the Myriad patent.²⁹ Myriad forbid French doctors from undertaking BRCA1 testing and required the tests to be sent to Myriad’s lab.³⁰ But the sequencing technique by Myriad Genetics fails to detect ten to twenty percent of expected mutations in BRCA1.³¹

26. Aude Lecrubier, *Patents and Public Health*, 3 EMBO REP. 1120, 1120 (2002).

27. 35 U.S.C. § 154(a)(1) & (2) (2000).

28. Andrea Knox, *Companies Holding Patents to Disease-Related Genes Limiting Access*, PHILA. INQUIRER, Feb. 13, 2000, at A1 (quoting Jon Merz).

29. Press Release, Institut Curie, Assistance Publique-Hôpitaux de Paris, & Institut Gustave-Roussy, *Against Myriad Genetics’s Monopoly on Tests for Predisposition to Breast and Ovarian Cancer Associated with the BRCA1 Gene 5* (Sept. 26, 2002) (on file with authors) [hereinafter Press Release, Institut Curie].

30. *See id.* at 4.

31. *Id.* at 5. For an example of a mutation that Myriad missed, see Sophie Gad et al.,

Thus, gene patenting runs the risk of directly harming a patient by failing to make available a medical diagnostic procedure that can detect a disease in her genetic make-up. Recent NIH-proposed guidelines recommend wide licensing of patented inventions to nonprofit researchers and public health agencies in order to remedy this problem, stressing that exclusive licensing agreements have “‘detrimental short-term and long-term effects on both the quantity and quality of health care.’”³²

A gene patent allows its holder to charge whatever price it wants. For example, prior to the patent opposition mentioned above, Myriad required that all BRCA1 and BRCA2 diagnostic testing be performed by their Utah laboratory at a cost of \$2,975 per test,³³ three times the amount French laboratories charged.³⁴

Gene patents can interfere with clinical adoption of genetic tests, potentially compromising the quality of testing by limiting the development of higher quality and lower-cost alternative testing methods.³⁵ A survey of seventy-two genetic-testing laboratories found that twenty-five percent of the laboratories have been deterred from offering a test due to the enforcement of a patent or license.³⁶ For example, beginning in 1998, SmithKline Beecham Clinical Laboratories sent letters to labs ordering them to stop performing or developing tests for the hemochromatosis (HFE) gene.³⁷ The patent holder asked for an up-front fee of \$25,000 from academic laboratories and as much as \$250,000 from commercial laboratories, plus a fee of twenty dollars per test.³⁸ As a result, thirty percent of labs that received the letter discontinued testing or ceased development of HFE testing services.³⁹

A patent holder might forbid anyone from using the genetic sequence it has patented, even if the patent holder does not itself offer a diagnostic

Identification of a Large Rearrangement of the BRCA1 Gene Using Colour Bar Code on Combed DNA in an American Breast/Ovarian Cancer Family Previously Studied by Direct Sequencing, 38 J. MED. GENETICS 388, 388 (2001).

32. David Malakoff, *NIH Roils Academe with Advice on Licensing DNA Patents*, 303 SCIENCE 1757, 1758 (2004) (quoting NIH draft guidelines).

33. Andrew Pollack, *Patent on Test for Cancer Is Revoked by Europe*, N.Y. TIMES, May 19, 2004, at C3.

34. See Press Release, Institut Curie, *supra* note 29, at 6.

35. See Merz et al., *supra* note 25, at 578.

36. MILDRED K. CHO, PREPARING FOR THE MILLENNIUM: LABORATORY MEDICINE IN THE 21ST CENTURY 47-58 (2d ed. 1998) (monograph by Bayer Corp.).

37. See Merz et al., *supra* note 25, at 578.

38. *Id.*

39. *Id.* at 577-78.

test using that sequence.⁴⁰ This practice could become more prevalent as more pharmacogenomic discoveries are made and inventors sit on their patent rights, prohibiting patients from receiving testing for genetic disease and interfering with the doctor-patient relationship. Most drugs only work on a certain percentage of patients who use them.⁴¹ Genetic testing can help distinguish those patients for whom a drug will work from those for whom it will not. But such tests will also limit the market for drugs. For example, one pharmaceutical company has filed for a patent on a genetic test to determine the effectiveness of its asthma drug, yet does not plan to develop the test or let anyone else develop it.⁴² Patent law in Europe, unlike in the United States, provides protections against such actions by requiring that the inventor actually “work” (i.e., use or develop) the invention; if the inventor does not “work” the invention, the inventor may be compelled to license the invention to another entity.⁴³

III. SOME GENE PATENTS VIOLATE INDIVIDUAL RIGHTS

A. Informed Consent Issues

In many different settings in the United States over the past thirty years, blood, tissue, and other bodily fluid samples have been collected from individuals and used in genetic research without the person’s consent or knowledge.⁴⁴ If a lucrative gene was found, it was patented. Once a gene is identified and patented, its availability is often severely restricted, even to the people who provided tissue samples and funding for the genetic research.⁴⁵ In one case, the court held that individuals who provided tissue and monetary support to a researcher for the discovery of a particular disease gene could maintain a claim of unjust enrichment against both the researcher and the hospital that patented the gene and charged a fee for

40. Cf. Gaia Vince, *Gene Patents “Inhibit Innovation,”* NEW SCIENTIST.COM, July 23, 2002, at www.newscientist.com/news/print.jsp?id=ns99992580.

41. See Allen D. Roses, *Pharmacogenomics and the Practice of Medicine*, 405 NATURE 857 (2000).

42. Geeta Anand, *Big Drug Makers Try To Postpone Custom Regimens*, WALL ST. J., June 18, 2001, at B1.

43. Yee Wah Chin, *Unilateral Technology Suppression: Appropriate Antitrust and Patent Law Remedies*, 66 ANTITRUST L.J. 441, 450 (1998).

44. Jeffery R. Botkin, *Informed Consent for the Collection of Biological Samples in Household Survey*, in CELLS AND SURVEYS: SHOULD BIOLOGICAL MEASURES BE INCLUDED IN SOCIAL SCIENCE RESEARCH? 276, 276-77 (Caleb E. Finch et al. eds., 2001).

45. See *Greenberg v. Miami Children’s Hosp.*, 264 F. Supp. 2d 1064 (S.D. Fla. 2003).

use of the genetic sequence in testing.⁴⁶ But the court also held that the tissue sources had no right to be informed about the potential commercialization of their tissue before they provided tissue to the researcher.⁴⁷ This could lead to the anomalous situation where a person's tissue could be used for commercial purposes without her knowledge or consent in ways that violate her personal or religious beliefs, and her only legal remedy would be monetary compensation after the offending act took place.

This is not a trivial concern. Many religion denominations oppose gene patents.⁴⁸ Certain religious and ethnic groups have concerns about the use of their tissue for research. In pending litigation, the Havasupai tribe of Arizona is suing researchers for unauthorized use of their genetic samples.⁴⁹ The group consented to give blood samples to a particular researcher for diabetes research.⁵⁰ They allege that without their consent, their samples were sent to other researchers around the country for research, which they had not approved, including research that might lead to discrimination against them as a group (such as schizophrenia research) and research that could contradict their religious beliefs (such as research on the purported origins and migrations of the group).⁵¹

In Europe, concern about informed consent of patients whose tissue is used in developing a gene patent is so important that it is mentioned in European patent provisions. Directive 98/44/EC of the European Parliament and Council of the European Union, created as a means to ensure uniformity in intellectual property rights as applied to biotechnological inventions throughout the European Union, states that where "an invention is based on biological material of human origin or if it uses such material . . . the person from whose body the material is taken must have had an opportunity of expressing free and informed consent thereto, in accordance with national law."⁵² If such a policy were in force in

46. *Id.* In extreme instances, the biobank that unjustly enriched itself might be required to disgorge all of its profits to the tissue sources. See *Univ. of Colo. Found. v. Am. Cyanamid Co.*, 153 F. Supp. 2d 1231 (D. Colo. 2001) (requiring disgorgement of patent royalties in an unjust enrichment context).

47. *Greenberg*, 264 F. Supp. 2d at 1070.

48. Southern Baptist Convention, *supra* note 8.

49. Larry Hendricks, *Havasupai Tribe Files \$50M Suit Against ASU*, ARIZ. SUN, Mar. 16, 2004, at A1. The case was filed in Coconino County Superior Court on March 12, 2004 (Case No. S-0300-CV-20040146).

50. Hendricks, *supra* note 49.

51. See *id.*

52. Council and Parliament Directive 98/44/EC of 6 July 1998 on the Legal Protection

the United States, it would protect individuals whose blood samples were used without their consent in genetics research and served as the basis for patent applications.

B. Reproductive Liberty Issues

Since a gene patent holder has the power to forbid all use of that specific gene or mutation for the lifetime of the patent, the patent holder can limit its use entirely in certain situations, such as by forbidding prenatal diagnosis for that particular gene. The company that holds patents on mutations in the BRCA1 and BRCA2 genes has indicated that it will use its control to forbid prenatal testing for breast cancer, perhaps due to the controversial potential for selective abortion.⁵³ However, such a stance interferes with a woman's reproductive liberty, a right guaranteed by the U.S. Constitution.⁵⁴ Because the issuance of a patent is a state action, even when issued to a private party, it might be deemed to infringe on reproductive rights where it limits the availability of genetic testing needed for a woman to make an informed decision. In *Lifchez v. Hartigan*, a federal judge struck down an embryo research ban as unconstitutional because it interfered with a woman's right to use innovative prenatal screening.⁵⁵ The judge said, "The cluster of constitutional choices that includes the right to abort a fetus within the first trimester must also include the right to submit to a procedure designed to give information about that fetus which can then lead to a decision to abort."⁵⁶

IV. TOWARD A NEW POLICY HORIZON

There is growing interest in the U.S. Congress in dealing with the problems created by patents on genetic sequences.⁵⁷ There are several potential policies that could be adopted. Genes could be declared

of Biotechnological Inventions, 1998 O.J. (L 213) 14, recital 26, http://europa.eu.int/eur-lex/pri/en/oj/dat/1998/l_213/l_21319980730en00130021.pdf.

53. LORI ANDREWS & DOROTHY NELKIN, *BODY BAZAAR: THE MARKET FOR HUMAN TISSUE IN THE BIOTECHNOLOGY AGE* 44 (2001).

54. *See Eisenstadt v. Baird*, 405 U.S. 438 (1972); *Griswold v. Connecticut*, 381 U.S. 479 (1965).

55. *Lifchez v. Hartigan*, 735 F. Supp. 1361, 1377 (N.D. Ill. 1990).

56. *Id.*

57. *See Consolidated Appropriations Act, 2004*, Pub. L. No. 108-199 § 634, 118 Stat. 3, 101 (2004) ("[N]one of the funds appropriated or otherwise made available by this act may be used to issue patents on claims directed to or encompassing a human organism.").

unpatentable subject matter. Another potential remedy is to allow doctors to perform diagnostic testing on patients without deeming the procedures to be infringement of the relevant gene patent. For example, Congress enacted a statutory provision exempting licensed medical physicians from infringement for use of a patented medical or surgical procedure.⁵⁸ Enacting a similar amendment for gene patents would permit doctors and laboratories to use patented gene sequences in diagnostic tests without having to pay a royalty or obtain a license.⁵⁹ Alternatively, the government could impose compulsory licensing for all uses of gene patents. Under this system, patent holders would have to grant licenses to researchers and physicians to use a patented genetic sequence in return for a reasonable fee to the patent holder.⁶⁰

CONCLUSION

Gene patents create problems for health care, medical research, and individual rights. While it might be appropriate to award patent rights to a genetic diagnostic kit or a genetic therapy, it is not appropriate to award protection over an isolated sequence or a clone of a gene. Prohibiting the patenting of genetic sequences is not inimical to patent law. Rather, it would be permissible in the United States and around the world under the public health exceptions in the World Trade Organization's TRIPS Agreement.⁶¹ It is crucial for high quality health care and individual autonomy that the United States reexamine its gene patent policy.

58. 35 U.S.C. § 287(c)(1) (2000) (noting that infringement actions "shall not apply against the medical practitioner or against a related health care entity with respect to such medical activity").

59. See Genomic Research and Diagnostic Accessibility Act of 2002, H.R. 3967, 107th Cong. (2002).

60. For a more comprehensive discussion of this proposal, see Lori B. Andrews, *The Gene Patent Dilemma: Balancing Commercial Incentives with Health Needs*, 2 HOUS. J. HEALTH L. & POL'Y 65, 103 (2002); see also Jon F. Merz, *Disease Gene Patents: Overcoming Ethical Constraints on Clinical Laboratory Medicine*, 45 CLINICAL CHEMISTRY 324, 328-29 (1999).

61. Agreement on Trade-Related Aspects of Intellectual Property Rights, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, LEGAL INSTRUMENTS—RESULTS OF THE URUGUAY ROUND vol. 31, 33 I.L.M. 81, art. 27 (1994).

World Health Law: Toward a New Conception of Global Health Governance for the 21st Century

Lawrence O. Gostin, J.D., LL.D.*

The international community joined together during the late twentieth century to form a world trade system. Although imperfect, the world trade system contains adjudicable and enforceable norms designed to facilitate global economic activity. Human health is at least as important as trade in terms of its effects on the wellbeing of populations. Moreover, health hazards—biological, chemical, and radionuclear—have profound global implications. Whether these threats' origins are natural, accidental, or intentional, the harms, as well as the response, transcend national frontiers and warrant a transnational response. Despite their high importance, the International Health Regulations (IHR) are antiquated, limited in scope, and burdened by inflexible assumptions and entrenched power structures.¹ This essay examines problems of obsolescence, narrow reach, and rigidity associated with the IHR, and proposes a new conception for world health law in the 21st Century.

ANTIQUATED GLOBAL HEALTH GOVERNANCE: THE HISTORICAL ORIGINS OF THE IHR

The origins of the IHR, the only global rules governing the international spread of infectious diseases, date back to the first International Sanitary Conference, held in Paris in 1851 to address the European cholera epidemics. During the latter half of the nineteenth

* John Carroll Research Professor of Law, Georgetown University, Professor of Public Health, Johns Hopkins University, and Director of the Center for Law & the Public's Health at Johns Hopkins and Georgetown Universities. Professor Gostin is working with the WHO on the IHR revision process. He also directs the CDC Collaborating Center on Law and the Public's Health. The views in this Essay do not necessarily reflect those of the WHO or the CDC.

1. Lawrence O. Gostin, *International Infectious Disease Law: Revision of the World Health Organization's International Health Regulations*, 291 JAMA 2623 (2004) [hereinafter Gostin, *IHR*].

century, ten sanitary conferences were held and eight conventions were negotiated (most did not come into force) to address the trans-boundary effects of infectious diseases. The International Sanitary Convention dealing with cholera was adopted in Venice in 1892, followed by another Convention dealing with plague in 1897.² In 1903, the International Sanitary Convention replaced the conventions of 1892 and 1897.³

At the turn of the twentieth century, the international community established regional and international institutions to enforce these conventions. American states set up the International Sanitary Bureau (ISB) in 1902, which became the Pan American Sanitary Bureau (PASB), a precursor to the Pan American Health Organization (PAHO).⁴ European States developed their own multilateral institution in 1907, L'Office International d'Hygiène Publique (OIHP).⁵ The Health Organization of the League of Nations (HOLN) was formed between the two world wars in 1923.⁶ Article XXIII of the League of Nations Covenant meekly stated that members would "endeavor to take steps in matters of international concern for the prevention and control of disease." The ISB, OIHP, and HOLN were separate institutions, without harmonization of goals or practices.

The United Nations was established after the horrors of World War II.⁷ One of the U.N.'s primary functions was the protection of global health. The World Health Organization (WHO) was established by the U.N. in order to fulfill this mandate.⁸ Its preamble expresses universal aspirations⁹

2. INT'L HEALTH REGULATIONS REVISION PROJECT, WORLD HEALTH ORGANIZATION: GLOBAL CRISIS—GLOBAL SOLUTIONS: MANAGING PUBLIC HEALTH EMERGENCIES OF INTERNATIONAL CONCERN THROUGH THE REVISED INTERNATIONAL HEALTH REGULATIONS 1, WHO Doc. WHO/CDS/CSR/GAR/2002.4 (2002), <http://www.who.int/csr/resources/publications/ihr/whocdsgar20024.pdf>.

3. International Sanitary Convention, Dec. 3, 1903, 35 Stat. 1770, 1 Bevans 359.

4. PAN AM. HEALTH ORG., PRO SALUTE NOVI MUNDI: A HISTORY OF THE PAN AMERICAN HEALTH ORGANIZATION (1992), http://165.158.1.110/english/pro_salute/contents.htm; *Agreement Between the World Health Organization and the Pan American Health Organization*, WHA Res. 2.91 (June 30, 1949), in WORLD HEALTH ORGANIZATION BASIC DOCUMENTS 38 (44th ed. 2003).

5. See DAVID P. FIDLER, INTERNATIONAL LAW AND INFECTIOUS DISEASES 24 (1999).

6. See *id.*

7. Article 55 of the U.N. Charter states that a primary objective of the U.N. is to promote "higher standards of living" and "solutions of international . . . health . . . problems." U.N. CHARTER art. 55.

8. Yukata Arai-Takahashi, *The Role of International Health Law and WHO in the Regulation of Public Health*, in LAW AND THE PUBLIC DIMENSION OF HEALTH 113 (Robyn Martin & Linda

stating that its “principles are basic to the happiness, harmonious relations and security of all peoples.”¹⁰ The WHO Constitution grants the agency the power to seek Member State adoption of conventions (Article 19),¹¹ promulgate regulations (Article 21), and make recommendations (Article 23).¹²

Pursuant to the agency’s Article 21 power, WHO Member States adopted the International Sanitary Regulations (ISR) on July 25, 1951. The ISR were renamed the International Health Regulations in 1969.¹³ The IHR initially applied to six diseases: cholera, plague, relapsing fever, smallpox, typhus, and yellow fever. The IHR were slightly modified in 1973 (particularly for cholera) and again in 1981 (to exclude smallpox, in view of its global eradication). The IHR currently apply only to cholera, plague, and yellow fever—the same diseases originally discussed at the first International Sanitary Conference in 1851. Thus the IHR have not been significantly changed since the ISR’s initial adoption in 1951 and predate modern health threats such as HIV/AIDS, SARS, and bioterrorism.

The 1995 World Health Assembly (WHA), in response to outbreaks of

Johnson eds., 2001).

9. Frank P. Grad, *The Preamble of the Constitution of the World Health Organization*, 80 BULL. WORLD HEALTH ORG. 981 (2002).

10. World Health Organization Constitution, July 22, 1946, pmbl., 62 Stat. 2679, 14 U.N.T.S. 185, 186-187; see also Yutaka Arai-Takahashi, *supra* note 8.

11. The only WHO Convention adopted pursuant to this power is the Framework Convention on Tobacco Control adopted in 2003. *WHO Framework Convention on Tobacco Control*, WHA Res. 56.1, World Health Assembly, 56th Ass., 4th plen. mtg, Agenda Item 13, Annex, WHO Doc. A56.VR/4 (May 21, 2003), http://www.who.int/tobacco/fctc/text/en/fctc_en.pdf.

12. See Allyn L. Taylor et al., *International Health Instruments: An Overview*, in OXFORD TEXTBOOK OF PUBLIC HEALTH 359 (Roger Detels et al. eds., 4th ed. 2002).

13. The current IHR contain several broad requirements for Member States: (1) Notifications—Countries must report to the WHO any case of these diseases, occurring in humans in their territories, and give further notification when an area is free from infection. (2) Health Standards at Points of Arrival and Departure—Countries must adopt hygiene measures at ports, airports, frontier posts, and on international cargo, goods, baggage, containers, and other articles. Hygiene measures include providing potable water and wholesome food; conducting inspections of equipment, installations, and premises; and maintaining facilities for isolation and care of infected persons, and for disinfecting, disinsecting, and deratting. (3) Health Documents—Countries may require health and vaccination certificates for travelers from infected to non-infected areas. (4) Maximum Measures—The health measures permitted by the IHR are “the *maximum measures* applicable to international traffic, which a State may require for the protection of its territory.” Gostin, *IHR*, *supra* note 1, at 2624.

cholera in Peru, plague in India, and Ebola hemorrhagic fever in Zaire, resolved to revise the IHR.¹⁴ Since that time, the WHA¹⁵ and other WHO governance structures¹⁶ have affirmed the importance of the reform process. The WHO Secretariat published a proposed revision of the IHR on January 12, 2004.¹⁷ Member States reviewed the draft during regional consultations and then in inter-governmental negotiations, with a view to adoption by the WHA in 2005.¹⁸

THE PROBLEMS OF SOVEREIGNTY, HORIZONTAL GOVERNANCE, AND ENTRENCHED POWER

Global health governance, then, is antiquated and sharply limited in scope. Even within its narrow reach, the WHO has experienced marked difficulties in enforcing the IHR in each content area.¹⁹ Why have nation states thus far resisted global health governance when they have acceded to global trade governance? Although perhaps not as readily quantifiable as economic gains from free trade, the trans-boundary effects of health hazards are profound. Biological, chemical, and radionuclear agents all have far-reaching consequences. With our modern system of global trade

14. *Revision and Updating of the International Health Regulations*, WHA Res. 48.7, World Health Assembly, 48th Ass., 12th plen. mtg. (May 12, 1995).

15. *Global Health Security: Epidemic Alert and Response*, WHA Res. 54.14, World Health Assembly, 54th Ass., 9th plen. mtg (May 21, 2001); *Revision of the International Health Regulations*, WHA Res. 56.28, World Health Assembly, 56th Ass., 10th plen. mtg. (May 28, 2003).

16. See *Revision of the International Health Regulations: Severe Acute Respiratory Syndrome (SARS)—Report by the Secretariat*, WHO Doc. A56/48 (May 17, 2003); *Revision of the International Health Regulations—Report by the Secretariat*, WHO Doc. EB111/34 (Dec. 15, 2002) [hereinafter *Revision of the IHR*].

17. *International Health Regulations: Working Paper for Regional Consultations*, Intergovernmental Working Group on the Revision of the Int'l Health Regulations, WHO Doc. IGWG/IHR/Working paper/12.2003 (Jan. 12, 2004) [hereinafter *IHR Revision—Working Paper*].

18. *Revision of the International Health Regulations: Report of the Secretariat*, WHO Doc. EB113/3 Rev.1 (Jan. 15, 2004).

19. See Gostin, *IHR*, *supra* note 1, at 2624 (“[M]ember States have: (i) not promptly reported notifiable diseases; (ii) not met hygienic standards at borders; (iii) required health certificates for non-listed diseases such as HIV/AIDS; and/or (iv) exceeded the allowable maximum measures by imposing bans on entry of travelers or goods without sufficient scientific justification. Member States do not comply for diverse reasons such as popular sovereignty or self-governance, political or economic interests, and incapacity due to lack of expertise or resources.”).

and international travel, nation states can no longer seal their borders to escape such hazards, if indeed they ever could. The health and economic effects of SARS and avian influenza, along with ongoing concerns about emerging infectious diseases and bioterrorism, may spur WHO Member States to agree to stronger forms of international health law. Continuing resistance to effective health regulation is most plausibly explained by countries' outdated assumptions about sovereignty, horizontal governance, and entrenched power.²⁰

Sovereignty

Sovereignty, although often criticized, remains an influential idea in international relations, particularly in matters of health. Sovereignty has multiple dimensions, but includes political authority over internal affairs, power to control border crossings, and freedom from external interference.²¹ The police power to protect the public's health and safety is a traditional prerogative of national sovereignty.²² Assertions of sovereignty, of course, are not always detrimental. A nation's decision to impose scientifically-based health regulations that are more stringent than required under international law is not simply a valid assertion of autonomy. Health regulations based on good science can provide increased protection for the state and its neighbors.

When used to preserve a poorly regulated status quo, however, assertions of sovereignty can severely harm global interests in health. Consider the potential adverse health effects within each of the three main spheres of sovereignty. First, state power to control internal affairs enables political leaders to set low standards for public health surveillance and regulation. Given the cross-boundary effects of health threats, a state's failure to identify and respond promptly to domestic health threats poses substantial risks to both its own citizens and other nations.²³ Second, the state's control over borders allows governments to ignore international health standards in regulating the flow of goods and people across its

20. David P. Fidler, *SARS: Political Pathology of the First Post-Westphalian Pathogen*, 31 J.L. MED. & ETHICS 485 (2003).

21. See John H. Jackson, *Sovereignty-Modern: A New Approach to an Outdated Concept*, 97 AM. J. INT'L L. 782 (2003).

22. LAWRENCE O. GOSTIN, *PUBLIC HEALTH LAW: POWER, DUTY, RESTRAINT* 47-51 (2000).

23. See, e.g., Joseph M. Schwartz, *On Doubting Thomas: Judicial Compulsion and Other Controls of Transboundary Acid Rain*, 2 AM. U. J. INT'L L. & POL'Y 361, 374-77 (1987) (noting the health risks posed to Canadians by the United States's failure to control its contribution to trans-boundary acid rain in the 1970s).

borders. The state may either set weak standards (facilitating the spread of disease) or overly strict standards (needlessly affecting travel and trade). Indeed, many international disputes arise from travel or trade restrictions imposed by international agencies or the states themselves.²⁴ Finally, a state's assertion of non-interference provides an ostensible justification for failing to comply with international health norms. A country may delay notifying the WHO of an emerging health threat, prevent its scientists from sharing information, or refuse to cooperate with international agencies.²⁵

Respect for sovereignty is particularly problematic because countries have built-in incentives for secrecy and inaction in the face of emerging health threats. Public notification of health hazards can adversely affect a country's economy and prestige. It can trigger media coverage or travel advisories affecting trade and tourism and adversely affect the reputation and electoral prospects of political leaders. One need only look at the political and economic effects of SARS in Asia and North America to understand the potentially perverse incentives of transparency in matters of health.²⁶

Horizontal Governance

Connected to the problem of sovereignty is the preference for horizontal governance of health threats. Under horizontal governance, nations regulate health threats through bilateral or regional agreements, eschewing the imposition of rules by international health agencies.²⁷ Indeed, since the European sanitary conferences in the nineteenth century, governments have focused primarily on border controls to prevent health threats. Horizontal governance is not a particularly effective method of protecting global health. Border controls can rarely prevent the

24. FIDLER, *supra* note 5, at 67-68.

25. Consider China's months-long failure to report the SARS outbreak. See Jerome Groopman, *The SARS Epidemic: Global Warning*, WALL ST. J., Apr. 23, 2003, at A22.

26. Gostin, *IHR*, *supra* note 1, at 2626 ("In many ways, it is in a country's [interests] to overlook WHO recommendations and regulations. . . . This dynamic was illustrated during the SARS outbreaks when China delayed notification to the WHO, and Ontario, Canada, resisted WHO travel advisories."); see also Keith Bradsher, *The SARS Epidemic: The Economic Impact*, N.Y. TIMES, Apr. 21, 2003, at A1 (describing the SARS epidemic as causing "the worst economic crisis in Southeast Asia since the wave of bank failures and currency devaluations that swept the region five years ago").

27. Fidler, *supra* note 20, at 487.

spread of disease, particularly if the threat is not detected promptly.²⁸

Vertical governance is likely to be far more effective by setting uniform standards for national health surveillance and regulation based on science.²⁹ Vertical governance means that international health agencies can set minimum public health capacities at the regional and national levels. Yet countries exhibit deep reservations about yielding their sovereignty to multinational authorities.³⁰ Vertical governance does not require countries to forego all autonomy, but greater devolution of power would enable the WHO to establish and enforce a system of global health preparedness that would make every country safer.

Entrenched Power

The current stagnation in global health governance may also be attributable to entrenched power structures. Economically and politically powerful countries, principally in Europe and North America, have had a disproportionate influence on the global health agenda.³¹ This geopolitical imbalance results in multiple problems for world health.

First, geopolitical centers of power have acted as if it were possible to protect themselves from the endemic diseases of the developing world. The bilateral and multilateral agreements in nineteenth-century Europe could be understood as an attempt to seal the Western European frontier to prevent the movement of epidemics from Africa and Asia.³² It is possible to see a similar dynamic today with border and immigration policies designed to fend off diseases such as hemorrhagic fever, tuberculosis, and HIV/AIDS.³³

Second, the developed world has an abiding interest in continuing its economic vitality through free trade agreements. It is perhaps for this reason that the IHR focus as much on commerce as health. The avowed “purpose of [the IHR] is to ensure the maximum security against the international spread of diseases with a minimum interference with world

28. *Id.* at 486.

29. Gostin, *IHR*, *supra* note 1, at 2626-27.

30. *Id.*

31. See, e.g., David P. Fidler, *Microbialpolitik: Infectious Diseases and International Relations*, 14 AM. U. INT'L L. REV. 1, 21 (1998) (noting that “infectious disease control as a matter of concern for the international system depends to a large extent on [the interests of] powerful states”).

32. FIDLER, *supra* note 5, at 30-31.

33. See *id.* at 13-14.

traffic.”³⁴ Yet, the SARS outbreaks demonstrated the need for decisive public health action, sometimes at the expense of commerce and trade.³⁵ Developed countries have similarly insisted on furthering their economic interests through the creation and protection of intellectual property rights for pharmaceutical companies, making lifesaving vaccines and drugs largely unaffordable in developing countries.³⁶ For example, although ninety-five percent of the burden of HIV/AIDS is in the developing world, only eight percent of those in need of antiretroviral treatments in this area have access to them.³⁷

Finally, developed countries have resisted systematic action to provide technical and financial assistance for health protection in poorer countries.³⁸ This failure to allocate resources equitably has powerful ramifications for world health. Resource-poor countries do not have the means to protect their own populations from the disproportionate burdens of endemic disease. The marked health disparities between the rich and poor regions of the world pose fundamental questions of fairness. At the same time, poor countries do not have the capacity for surveillance and response to emerging infections to prevent major outbreaks.³⁹ This is not simply a problem in developing countries but poses a major concern in the developed world. In an age of global travel and commerce, health hazards can move rapidly across the world.⁴⁰ Health protection is only as

34. WORLD HEALTH ASSEMBLY, INTERNATIONAL HEALTH REGULATIONS, ADOPTED BY THE TWENTY-SECOND WORLD HEALTH ASSEMBLY, BOSTON 2 (3d ed. 1983) (1969).

35. Lawrence O. Gostin et al., *Ethical and Legal Challenges Posed by Severe Acute Respiratory Syndrome: Implications for the Control of Severe Infectious Disease Threats*, 290 JAMA 3229 (2003) [hereinafter Gostin, SARS].

36. Giovanni Andrea Cornia, *Globalization and Health: Results and Options*, 79 BULL. WORLD HEALTH ORG. 834, 837 (2001) (noting that “even in the cases in which [the Agreement on Trade-Related Aspects of Intellectual Property Rights] allows parallel imports of cheap generic drugs, trade pressures by [developed countries] limits access to affordable drug imports”).

37. World Health Org., Coverage and Need for Antiretroviral Treatment (June 2004), at <http://www.who.int/3by5/coverage/en/> (noting that only eight percent of those in the developing world and four percent of those in Africa who require antiretroviral treatment were receiving antiretroviral treatment in June, 2004).

38. Consider the difficulties encountered in gathering adequate contributions from developed countries for the Global Fund to Fight AIDS, Tuberculosis and Malaria. See Donald G. McNeil, Jr., *World's Anti-AIDS Donations Slow, Cutting U.S. Contribution, Too*, N.Y. TIMES, Aug. 19, 2004, at A18 (detailing the lack of contributions to the Global Fund).

39. FIDLER, *supra* note 5, at 12-13.

40. LAURIE GARRETT, BETRAYAL OF TRUST: THE COLLAPSE OF GLOBAL PUBLIC HEALTH (2000) (arguing that the weakness of the public health infrastructure in developing

good as the weakest link, so low capacities in poor countries threaten every nation.

TOWARD A NEW CONCEPTION OF GLOBAL HEALTH GOVERNANCE

To overcome the problems of sovereignty, horizontal governance, and entrenched power, the international community should consider a new conception for global health based on the rule of international law.⁴¹ The WHO's proposed revision of the IHRs, if expanded, could serve as a model for effective public health governance.⁴²

The Salience of Health over Trade

The IHR should stress the salience of global health and the WHO's role in achieving that purpose. The WHO should dedicate itself to the protection and promotion of global health. Wherever possible, health rules should respect travel and trade, while assuring that promoting global health remains the WHO's primary mission. That is the vision of the WHO Constitution, which does not mention the protection of trade or commerce.

Wide Jurisdiction

The narrow scope of the IHR impedes the WHO in effectively dealing with modern health threats. The revised IHR cover "all events potentially constituting a public health emergency of international concern."⁴³ This new approach is preferable because it is flexible, prospective, and covers all hazards (radiological, chemical, and biological), whether naturally-occurring, accidental, or intentional. It does not require amendment of the IHR each time a novel health threat emerges.

Comprehensive Data Collection

Rapid and comprehensive data collection is crucial to global health. Yet surveillance is hindered by the reluctance of countries to fully cooperate.⁴⁴ Global surveillance can be dramatically improved by effective

countries threatens the health and security of developed nations).

41. See Gostin, *IHR*, *supra* note 1.

42. See *Revision of the IHR*, *supra* note 16.

43. *IHR Revision—Working Paper*, *supra* note 17, art. 5(1).

44. FIDLER, *supra* note 5, at 65 ("The IHR surveillance system has broken down because Member States regularly fail to notify WHO of outbreaks of diseases subject to the IHR.").

vertical governance. First, the WHO could establish criteria for uniform data sets, core informational requirements, and timely monitoring and reporting. These norms would help set a standard for national and global surveillance. Second, the WHO should expand its data sources beyond official government channels. "Small-world networks" consisting of scientists, health professionals, membership associations, and non-governmental organizations could considerably broaden the sources of health information. Finally, the WHO should utilize modern technology for surveillance, including electronic health records and the internet, to gather and analyze surveillance data. The WHO is already beginning this process, and it could be enhanced through the revised IHR.

National Public Health Preparedness

Uniformly strong public health capacities at the national level offer the best prospect for global health. Prompt and efficient monitoring and response at the national level is critically important to prevent the proliferation of disease.⁴⁵ To improve national competencies, the WHO should set minimum standards for laboratories, data systems, and response. By setting performance standards and measuring outcomes, the WHO could continually help member states evaluate their public health preparedness. Compliance with international health norms has been a serious problem that must be addressed by the WHO. This could be accomplished through a combination of hard and soft law: mediation, adjudication, and incentives.

A related problem is that poor countries cannot meet minimum standards for public health preparedness. The international community, therefore, should substantially increase technical and financial assistance for health system improvement in developing countries. This commitment would not be open-ended; nor would it necessarily be sufficient to meaningfully reduce global health disparities. However, at a minimum, the developed world should help assure that all nations have core public health capacities for surveillance and containment of emerging health threats of global importance. This kind of commitment not only allows progressive development of higher standards of health in resource-poor countries, but also is in the interests of the industrialized world.

Human Rights Safeguards

The IHR were promulgated before the development of international

45. See Gostin, *SARS*, *supra* note 35, at 3231.

human rights law. As a result, the IHR do not protect individual rights under international law. Many aspects of global health regulation affect human rights, including surveillance (privacy), vaccination and treatment (bodily integrity), travel restrictions (movement), and isolation and quarantine (liberty). Health measures may also be applied inequitably, leading to discrimination against unpopular groups, such as migrants and ethnic minorities. The IHR could demonstrate respect for human rights by incorporating the internationally accepted norms contained in the Siracusa principles, which require health measures to be necessary, proportionate, and fair.⁴⁶ Health measures should be based on the rule of law and provide due process for persons whose liberty is placed in jeopardy.

Good Public Health Governance

WHO member states have not always followed basic principles of good public health governance. They have sometimes acted in ways that are insular and discriminatory, without adequate regard to science. The WHO could set an example of good public health governance by complying with the principles of transparency, objectivity, and fairness. The agency's policies and recommendations should be established in an open manner, based on scientific evidence, and exercised equitably. The agency gains credibility by its adherence to science, the truthfulness of its disclosures, and its fair-dealings with countries, rich and poor alike.

The Future of Global Health Governance

More effective monitoring and management of international health threats is undoubtedly a global public good. Yet, the question arises whether international law is the most effective institutional vehicle to achieve this objective. After all, the WHO has been relatively impotent in enforcing the existing IHR. During the SARS outbreaks, moreover, the agency was active and effective without the need for formal international law.

Certainly, revised IHR will not assure capable leadership and sound governance by the WHO. Yet, the revision offers an opportunity for a renewed commitment by the international community to a shared vision of

46. U.N. ECON. & SOCIAL COUNCIL SUB-COMM. ON PREVENTION OF DISCRIMINATION & PROTECTION OF MINORITIES, SIRACUSA PRINCIPLES ON THE LIMITATION AND DEROGATION OF PROVISIONS IN THE INTERNATIONAL COVENANT ON CIVIL AND POLITICAL RIGHTS, U.N. Doc. E/CN.4/1985/4 (1985), *reprinted in* 7 HUM. RTS. Q. 3 (1985).

global health. The revision would give the WHO a clear mission, significantly enhanced jurisdiction, and formal power to set standards and make recommendations. By assenting to a far-reaching revision of the IHR, Member States would cede some control over health threats of international importance and grant to the WHO a measure of centralized authority.

International law can help forge a new conception of global health governance that assures:

- the salience of health over trade;
- broad jurisdiction over conditions of international public health importance;
- global surveillance through core data requirements and “small-world networks”;
- national public health preparedness by enforcing standards, creating incentives, and cultivating developmental and technical assistance;
- human rights protection through incorporation of the Siracusa principles; and
- good public health governance through transparency, objectivity, and fairness.

By adhering to the rule of law, the international community can take a vital step toward better protection against the biological, chemical, and radiological hazards posed in the modern age.

The Challenge of Assuring Continued Post-Trial Access to Beneficial Treatment

Christine Grady, R.N., Ph.D.*

Sam Jones has agreed to participate in a clinical trial testing an experimental drug as a possible treatment for his chronic disease. The primary outcome of effectiveness is a decrease in a specific disease marker after two months of taking the drug. Soon after Sam begins to take the drug, he reports feeling better and the level of disease marker in his blood is significantly lower at each study interval. The study reaches its predetermined endpoint, is stopped as planned, and the sponsor submits an application to a regulatory agency to license the drug for this indication. It is clear that Sam would clinically benefit from continuing to take the drug.

Is it the responsibility of the investigator or the research sponsor to ensure that Sam and other participants in this study continue to receive the drug or even to provide it to them after a study ends? This question is at the heart of recent controversy regarding post-trial benefits and may be one of the biggest ethical challenges facing clinical investigators, especially those involved in international research, over the next several years.

POST-TRIAL BENEFITS

Until recently, regulations and codes of research ethics have been silent about what should happen at the conclusion of a clinical study. Regulations and codes have focused on protecting the rights and welfare of individuals in clinical research by requiring that the research design was appropriate, risks were minimized, research was reviewed by an independent body, and the participant's consent was adequately informed

* Department of Clinical Bioethics, National Institutes of Health (NIH). The author is currently head of the section on human subject research at the NIH Department of Clinical Bioethics. She would like to thank Reidar Lie and Zeke Emanuel for their helpful discussions and critical review. The views expressed here are those of the author and do not necessarily reflect those of the Clinical Center, the National Institutes of Health, the Public Health Service, or the Department of Health and Human Services.

and voluntary. The Common Rule in the Code of Federal Regulations requires that investigators inform participants in advance of any interventions or compensation that will be provided if a research participant is injured during trial participation,¹ but it offers no guidance regarding what should happen to a person like Sam. Even the forward-looking *International Ethical Guidelines for Biomedical Research Involving Human Subjects*,² published by the Council for International Organizations of Medical Sciences (CIOMS) in 1993, did not address the issue of what should happen at the end of a clinical study to participants who are receiving beneficial treatment. The CIOMS guidelines did address the issue of compensation for research injury and also introduced the idea that the sponsoring agency of externally sponsored research “should agree in advance of the research that any product developed through such research will be made reasonably available to the inhabitants of the host community or country at the completion of successful testing.”³

Controversy in the late 1990s about the ethics of international HIV trials brought increased attention to the issue of post-trial benefits.⁴ Many agreed that in order to minimize the possibility of exploiting research participants in developing countries, a plan for how the benefits of research would be made reasonably available to the developing country or community was required.⁵ Most of the discussion focused on the requirement that products proven effective through research be made available to the wider community from which research participants were drawn. The question of continued treatment of participants like Sam has

1. 45 C.F.R. § 46 (2003).

2. COUNCIL FOR INT’L ORG. OF MED. SCI., *INTERNATIONAL ETHICAL GUIDELINES FOR BIOMEDICAL RESEARCH INVOLVING HUMAN SUBJECTS* (1993) (on file with author).

3. *Id.* at 45.

4. See Marcia Angell, *The Ethics of Clinical Research in the Third World*, 337 NEW ENG. J. MED. 847 (1997); Peter Lurie & Sidney M. Wolfe, *Unethical Trials of Interventions to Reduce Perinatal Transmission of the Human Immunodeficiency Virus in Developing Countries*, 337 NEW ENG. J. MED. 853 (1997); Harold Varmus & David Satcher, *Ethical Complexities of Conducting Research in Developing Countries*, 337 NEW ENG. J. MED. 1003 (1997).

5. See, e.g., NAT’L BIOETHICS ADVISORY COMM’N, *ETHICAL AND POLICY ISSUES IN INTERNATIONAL RESEARCH: CLINICAL TRIALS IN DEVELOPING COUNTRIES* (2001) [hereinafter NBAC]; NUFFIELD COUNCIL OF BIOETHICS, *THE ETHICS OF RESEARCH RELATED TO HEALTHCARE IN DEVELOPING COUNTRIES* 11 (2002), http://www.nuffieldbioethics.org/go/ourwork/developingcountries/publication_309.html [hereinafter NUFFIELD]; Leonard H. Glantz et al., *Research in Developing Countries: Taking “Benefit” Seriously*, 28 HASTINGS CENTER REP. 38 (1998); Eric M. Meslin & Harold T. Shapiro, *Ethical Issues in the Design and Conduct of Clinical Trials in Developing Countries*, 345 NEW ENG. J. MED. 139 (2001).

been less frequently attended to, with a few notable exceptions.⁶ The challenges of assuring that products proven effective during a trial are subsequently made reasonably available to the population in which the product was tested are formidable.⁷ Nonetheless, it is both practically and ethically a different challenge than that of assuring that individual research participants continue to receive beneficial treatments once the trial is over. Yet, the two issues are commonly conflated.⁸

CURRENT GUIDANCE

In the past few years, certain research ethics guidance documents,⁹ reports,¹⁰ and national guidelines¹¹ have begun to address the issue of continued post-trial treatment of participants like Sam who are receiving beneficial treatment. The Declaration of Helsinki, first published in 1964 by the World Medical Association (WMA), is internationally recognized as a major source of ethical guidance for the conduct of clinical research.¹² The WMA substantially revised the Declaration of Helsinki at its fifty-second assembly in 2000 due to intense public disagreement regarding the

6. See NBAC, *supra* note 5, at 12-13; COUNCIL FOR INT'L ORG. OF MED. SCI., INTERNATIONAL ETHICAL GUIDELINES FOR BIOMEDICAL RESEARCH INVOLVING HUMAN SUBJECTS cmt. on guideline 10 (2002) [hereinafter CIOMS 2002] ("[I]f an investigational drug has been shown to be beneficial, the sponsor should continue to provide it to the subjects after the conclusion of the study, and pending its approval by a drug regulatory authority."); JOINT UNITED NATIONS PROGRAMME ON HIV/AIDS, ETHICAL CONSIDERATIONS IN HIV PREVENTIVE VACCINE RESEARCH 13 (2000) [hereinafter UNAIDS].

7. See Participants in the 2001 Conference on Ethical Aspects of Research in Developing Countries, *Moral Standards for Research in Developing Countries: From "Reasonable Availability" to "Fair Benefits,"* 34 HASTINGS CENTER REP. 17 (2004).

8. See Editorial, *One Standard, Not Two*, 362 THE LANCET 1005 (2003) [hereinafter *One Standard, Not Two*].

9. See, e.g., WORLD MED. ASS'N, DECLARATION OF HELSINKI (2000), <http://www.wma.net/e/policy/pdf/17c.pdf> [hereinafter WMA] (clarified in 2002 and 2004); CIOMS 2002, *supra* note 6.

10. See, e.g., NBAC, *supra* note 5; NUFFIELD, *supra* note 5.

11. See, e.g., CLINICAL TRIALS WORKING GROUP OF THE S. AFRICAN DEP'T OF HEALTH, GUIDELINES FOR GOOD PRACTICE IN THE CONDUCT OF CLINICAL TRIALS IN HUMAN PARTICIPANTS IN SOUTH AFRICA ¶ 9.3.5 (2000), at http://196.36.153.56/doh/docs/policy/trials/trials_contents.html; INDIAN COUNCIL OF MED. RES., ETHICAL GUIDELINES FOR BIOMEDICAL RESEARCH ON HUMAN SUBJECTS 27 (2000), <http://icmr.nic.in/ethical.pdf>; NAT'L CONSENSUS CONF., GUIDELINES FOR THE CONDUCT OF HEALTH RESEARCH INVOLVING HUMAN SUBJECTS IN UGANDA (1997) (on file with author).

12. WMA, *supra* note 9.

aforementioned international HIV trials. The extensive changes included the addition of several new ideas that had not appeared in previous versions of the Declaration.¹³ One addition to the 2000 version of Helsinki, Paragraph 30, directly speaks to what should happen to Sam: "At the conclusion of the study, every patient entered into the study should be assured of access to the best proven prophylactic, diagnostic, and therapeutic methods identified by the study."¹⁴

Certain paragraphs in the 2000 version of the Declaration of Helsinki, particularly Paragraph 29 (regarding placebo-controlled trials) and Paragraph 30 (regarding post-trial treatment of participants), have continued to fuel considerable debate. A clarification to Paragraph 29, issued by the WMA in 2002, attempted (some would say unsuccessfully) to affirm a definitive stance about placebo-controlled trials.¹⁵ A call for clarification or an amendment to Paragraph 30 was also considered. The WMA voted early in 2004 not to amend Paragraph 30, but to consider issuing a clarification.¹⁶ In October 2004, a clarification of Paragraph 30 was issued by the WMA:

The WMA hereby reaffirms its position that it is necessary during the study planning process to identify post-trial access by study participants to prophylactic, diagnostic and therapeutic procedures identified as beneficial in the study or access to other appropriate care. Post-trial access arrangements or other care must be described in the study protocol so the ethical review committee may consider such arrangements during its review.¹⁷

WHAT IS THE ISSUE?

At first glance, it may appear to be common sense that if someone is doing well on a medication or treatment, even if that treatment is

13. See Heidi P. Forster et al., *The 2000 Revision of the Declaration of Helsinki: A Step Forward or More Confusion?*, 358 THE LANCET 1449 (2001).

14. WMA, *supra* note 9, ¶ 30.

15. Douglas P. Lackey, *Clinical Research in Developing Countries: Recent Moral Arguments*, 18 CADERNOS DE SAÚDE PÚBLICA 1455, 1457-58 (2002); Reidar K. Lie et al., *The Standard of Care Debate: The Declaration of Helsinki Versus the International Consensus Opinion*, 30 J. MED. ETHICS 190, 190 (2004).

16. WORLD MED. ASS'N, WORKGROUP REPORT ON THE REVISION OF PARAGRAPH 30 OF THE DECLARATION OF HELSINKI (2004), http://www.wma.net/e/ethicsunit/pdf/wg_doh_jan2004.pdf [hereinafter WMA WORKGROUP REPORT].

17. WMA, *supra* note 9, at 5; see also Press Release, World Med. Assembly, Clarification on Declaration of Helsinki (Oct. 11, 2004), http://www.wma.net/e/press/2004_24.htm.

investigational, it should be continued. Discontinuation of such a treatment may have poor health consequences for many research participants, and therefore should require compelling justification. Yet most research protocols do not include provisions to assure continued access to products that are providing clinical benefit to individual participants after the study concludes. Even for investigational drugs that eventually are approved, licensed, and made available through the health care system, there is usually a lag time during which a research participant could benefit from continued treatment. And in some cases, individual research participants will not be able to obtain such treatments even after approval.

Anecdotal evidence suggests that most investigators, at the very least, refer research participants to continued treatment through their regular health care providers when a trial is over. However, the realities of limited access to health care, primarily in the developing world, but also in many communities in the United States, suggest that referral to health care providers may often be inadequate. Research participants can also be referred to pharmaceutical company-sponsored patient assistance programs, which are designed to help patients obtain subsidized access to drugs they need that would otherwise be unavailable or unaffordable. Pharmaceutical companies sometimes pledge to continue to provide a drug that participants are receiving in a clinical trial for a predetermined period of time—often three to five years—and more rarely for the life of an individual participant.¹⁸ Investigators have been known to creatively seek out public assistance programs, social services, additional research protocols, and other sources of funds to be able to continue to provide beneficial drugs to research participants in the short term.¹⁹

Despite these efforts, there is no system in place for assuring continued treatment. And the question of whose responsibility or obligation it is to make sure a research participant continues to receive beneficial treatment remains unsettled, as noted by the editors of *The Lancet*:

The idea behind [the Declaration of Helsinki] language is straightforward: a person who participates in a trial should have a chance to benefit from what is learned from the trial—a principle that is particularly important for participants in the developing world. Trial

18. See Nancy Kass & Adnan A. Hyder, *Attitudes and Experiences of U.S. and Developing Country Investigators Regarding U.S. Human Subjects Regulations*, in NBAC, *supra* note 5, at B-1, B38-45.

19. *Id.*

participants in wealthy nations will usually be able to get the best available treatment after the trial is over. But in the developing world when the researchers pack up and go home, participants can be left with nothing.²⁰

As important as it is to recognize that research participants should have the chance to benefit from what is learned from a trial, it is also true that benefits can and probably should come in many other forms besides continued receipt of treatment. In addition, the Declaration of Helsinki is silent on *who* should assure continued treatment and assigning responsibility for assuring continued treatment is not straightforward.

MORAL REASONS FOR ASSURING CONTINUED TREATMENT

If a participant is deriving clinical benefit from an investigational therapy, withdrawing that therapy can be harmful. For many diseases, especially those requiring chronic treatment, exacerbation of symptoms or disease can occur if treatment is stopped. In accord with principles of non-maleficence and beneficence, patients, including those who are being treated as participants in research, should continue to receive a treatment they need as long as they are benefiting from it.

Further, stopping a clinically beneficial treatment simply because a clinical trial ends seems unfair and might contravene ethical obligations created by enlisting people to participate in research.²¹ Because research participants accept some risk for the good of society and the advancement of science, certain things are owed to them in return. The U.S. National Bioethics Advisory Committee described this as “justice as reciprocity.”²² Also, because research participants are asked to entrust certain aspects of their health to researchers, there is a corresponding responsibility of researchers to care for these aspects.²³ Although the basis for these obligations may seem indisputable, the extent of the obligation—or what exactly is owed to research participants and by whom—is at the center of the debate. Belsky and Richardson argue that the extent of researchers’ obligation to care for participants is influenced by several factors, including the strength of the relationship and the vulnerability of the participants.²⁴ Participants who are ill and participate in research with the hope of deriving some therapeutic benefit put partial trust in investigators.

20. *One Standard, Not Two*, *supra* note 8, at 1005.

21. Forster et al., *supra* note 13, at 1451.

22. NBAC, *supra* note 5, at 59.

23. See Leah Belsky & Henry S. Richardson, *Medical Researchers’ Ancillary Clinical Care Responsibilities*, 328 BRIT. MED. J. 1494 (2004).

24. *Id.* at 1495-96.

Dropping them completely may be an abandonment of investigator responsibilities and trust. A physician's role, even as clinical investigator, includes that of advocating for the welfare of the research participant.

Exploitation is another worry in clinical research. In research, exploitation occurs when the participant is taken unfair advantage of for the investigator's or sponsor's benefit. Is the research participant who receives beneficial treatment for a limited period of time exploited? Receiving beneficial treatment through trial participation even for a finite period of time may be perceived by participants with limited access to health care as a good option, or at least as a better option than no treatment at all. I would argue that it is not necessarily exploitative to offer time-limited access to beneficial treatment through research for a willing and informed participant. Yet, there are other moral reasons why researchers should take steps to assure post-trial access to beneficial treatment.

The moral obligation to assure that beneficial treatment is continued might well be fulfilled by referring a research participant like Sam to a physician who can prescribe the drug and, when necessary, by making provisions through the sponsor or a pharmaceutical company to provide the drug in the interim. But what if the drug is unlikely to be licensed in the jurisdiction for the foreseeable future? And what if, even if the drug is licensed, Sam or someone like him will not be able to afford it? Growth in international collaborative research has called attention to this problem because of the reality of limited availability and access to medical treatment in many developing countries.²⁵ It is also an important issue in countries like the United States, where health care services are unevenly available.

OTHER CHALLENGES TO PROVIDING CONTINUED TREATMENT

Even if the rationale for assuring continued treatment is compelling, the question of who should be responsible for assuring this and how it should be accomplished remains. It has been argued that if pharmaceutical companies and sponsors are made solely responsible for assuring continued access to beneficial treatment, this requirement could serve as a major disincentive for companies to engage in certain kinds of research.²⁶ This could also jeopardize the future of research in places with limited health care access,²⁷ especially for diseases that might require

25. NBAC, *supra* note 5, at 59.

26. See P. G. De Roy, *Helsinki and the Declaration of Helsinki*, 50 WORLD MED. J. 9 (2004).

27. See Bernard Pécoul et al., *Access to Essential Drugs in Poor Countries: A Lost Battle?*, 281

chronic or expensive treatment.²⁸ Commentators worry about the possibility of dampening research in developing countries where new treatments are needed the most.²⁹

A recent and dramatic example of this tension unfolded in a planned study of tenofovir for possible prevention of HIV in Cambodia. The study called for the inclusion of almost one thousand sex workers. The study was halted when the Women's Network for Unity, a Cambodian sex workers union, demanded a guarantee that participants would receive health care for thirty years following conclusion of the trial. Ironically, the Asian Pacific Network of Sex Workers concurrently denounced the trial in a protest at the World AIDS Conference for offering participants access to better treatment than they would have otherwise received.³⁰

A separate question also arises about the limits of such responsibility if it is assumed by a sponsor. In other words, how should it be decided whether it is sufficient for a sponsor to agree to provide continued treatment to participants for three years, for thirty years, as the Cambodian sex workers wanted, or for participants' lifetimes? The recent WMA clarification suggests that the adequacy of arrangements for post-trial access to beneficial treatments should be decided by ethics review committees.³¹ Ethics review committees still need guidance to determine the adequacy of such arrangements for post-trial treatment in each case. This determination should also take into consideration the need for monitoring and administration of supplied treatments.

Previous arguments for modification or clarification of Paragraph 30 of the Declaration of Helsinki focused on the specifics of Paragraph 30.³² Some have debated the exact meaning of "best proven," and have pointed out that a single trial almost never proves the effectiveness of an intervention. Others have noted that an ethical requirement for assuring post-trial treatment could never be absolute, since many research studies

JAMA 361 (1999).

28. See Helen Frankish, *WMA Postpones Decision to Amend Declaration of Helsinki*, 362 THE LANCET 963 (2003).

29. See, e.g., Jintanat Ananworanich et al., *Creation of a Drug Fund for Post-Clinical Trial Access to Antiretrovirals*, 364 THE LANCET 101 (2004).

30. Jon Cohen, *Cambodian Leader Throws Novel Prevention Trial into Limbo*, 305 SCIENCE 1092 (2004); Marilyn Chase & Gautam Naik, *Key AIDS Study in Cambodia Now in Jeopardy*, WALL ST. J., Aug. 12, 2004, at B1.

31. WMA WORKGROUP REPORT, *supra* note 16.

32. See WORLD MED. ASS'N, WMA SECRETARIAT REPORT ON THE REVISION OF PARAGRAPH 30 OF THE DECLARATION OF HELSINKI (2003), http://www.wma.net/e/ethicsunit/pdf/secretariat_report_rev_paragraph30.pdf; WMA WORKGROUP REPORT, *supra* note 16.

are not treatment trials and it is far from clear what obligation would be due to participants who do not benefit from the trial. How would an obligation to assure “the best proven methods . . . identified by the study”³³ be affected if it was known that a treatment not tested in the trial is more effective for the condition in question? All of these issues should be clarified in available guidance documents so that researchers and research sponsors know the extent of their obligations, and ethics review committees can determine whether proposed arrangements are acceptable. However, deliberations about particular details should occur against a background commitment to finding ways to assure continued access to beneficial treatment.

Certain commentators and reviewers have worried about the possibility of undue inducement from continued provision of treatment.³⁴ That is, if continued access to treatment is guaranteed, the treatment access may be so attractive that an individual might be unable to refuse participation even if he or she wanted to. For a study that is otherwise ethical, and from which the participant is deriving benefit and not subject to significant risk of harm, concerns about undue inducement from continuing beneficial treatment seem misplaced.³⁵

POSSIBLE STRATEGIES

Despite the many problems and challenges in assuring continued treatment after a trial, there is basic agreement that this should be done. The real challenge is to specify how this obligation should be understood and discharged. With that in mind, I will suggest some possible strategies to help Sam and others like him.

First, the various partners to a research study—investigators, sponsors, communities, national health systems, international organizations—should assume responsibility for this problem together. Possible ways of addressing continued access for research participants who benefit from investigational treatments should be discussed and negotiated prior to beginning a study. Researchers and sponsors cannot ignore this issue, believing it is someone else’s problem. At the same time, researchers and sponsors cannot be saddled with the sole responsibility of treating people

33. WMA, *supra* note 9.

34. See, e.g., Leah E. Hutt, *Freebies for Subject 641: A Discussion of the Ethical Prospect of Providing Drug Trial Subjects with Post-Trial Access to the Drug Tested—A Canadian Perspective*, 6 HEALTH L.J. 169, 185 (1998).

35. See Ezekiel J. Emanuel, *Ending Concerns About Undue Inducement*, 32 J.L. MED. & ETHICS 100 (2004).

who ought to be receiving treatment through the regular health care infrastructure. Expecting researchers and sponsors to fill that gap is not only an unrealistic expectation but would also act as a powerful negative disincentive. The aforementioned trial in Cambodia may be a good example. Involved parties should engage in good faith negotiations before a study begins to agree on how treatment will be assured after a study has concluded.

Those involved in research and health care delivery should be working together to come up with creative strategies to offer continued treatment to research participants who need it. One example of a creative strategy with multiple partners is the HIV Netherlands, Australia, Thailand Research Collaboration (HIV-NAT) co-payment and sliding scale drug fund program. HIV-NAT is a non-governmental, non-profit organization of three collaborators: the Thai Red Cross AIDS Research Center in Thailand, the National Centre in HIV Epidemiology and Clinical Research in Sydney, and the International Antiviral Therapy Evaluation Centre in Amsterdam.³⁶ Part of the motivation for creating the drug fund was eloquently described in a recent publication: "Although we stated clearly in consent forms that we could not promise post-trial drug supply, we were compelled to take action when faced with the tragic prospect of watching patients reversing their excellent quality of life gained while on antiretrovirals."³⁷ In the HIV-NAT drug fund program, patients who apply are initially assessed and then re-assessed annually by experienced social workers to determine their ability to pay, and the case is then reviewed by the drug fund committee who decides on an amount to be subsidized. The committee works with prescribing physicians to consider possible ways to reduce costs without jeopardizing the patient and also oversees the bulk purchase of drugs to obtain low prices. Patient support may come in the form of cash or drugs or a combination of the two.³⁸ Models such as this one can be adapted for other types of trials or other areas. In any case, it is clear that other approaches are sorely needed.

Second, the problem of post-trial access to beneficial treatments for participants should always be considered in the context of other considerations for ethical research. Continued treatment of research participants with medications they receive in a trial cannot make otherwise bad research ethical. Providing treatment to a small number of individuals during or after a trial does not eliminate or address concerns about

36. See Ananworanich et al., *supra* note 29.

37. *Id.* at 101.

38. *Id.*

exploitation. Negotiating fair benefits from research in order to minimize exploitation of participants and communities is a necessary part of collaborative research.³⁹

Third, the world health community must remain committed to finding ways to promote better access to needed health care and treatment globally. This will require the energy and creativity of policymakers, scientists, clinical providers, politicians, and communities. If patients everywhere had better access to needed treatments, continued access to treatment at the end of a trial would be primarily a temporary issue. Research is only one way of contributing to improved access to health care. It does so through the application of rigorous methods to search for health care interventions that are appropriate, affordable, safe, effective, and easy to administer. Even those with creative strategies, like the HIV-NAT group, articulate this larger need.⁴⁰

Fourth, sponsors and researchers should take responsibility for certain short-term solutions when appropriate. For example, sometimes it will be necessary to provide beneficial medications to participants while awaiting licensure, or to establish or support patient assistance programs for expensive treatments. Continued attention to reducing the costs of treatments for those who need them is also called for.

The goal of clinical research is to find new knowledge to improve health and health care. Participants in clinical research contribute invaluablely to this goal. Every effort should be made to find ways to assure that participants like Sam who are receiving beneficial treatments as part of their clinical trial continue to receive it after the trial ends.

39. See Participants in the 2001 Conference on Ethical Aspects of Research in Developing Countries, *Fair Benefits for Research in Developing Countries*, 298 SCIENCE 2133 (2002).

40. "For any developing country, long-term drug supply for patients at the end of a trial can only realistically be sustained if the government provides it. . . . A drug fund should be a temporary solution until the ultimate goal of access for all is achieved." Ananworanich et al., *supra* note 29, at 102.

The Most Important Health Care Legislation of the Millennium (So Far): The Medicare Modernization Act

Timothy Stoltzfus Jost, J.D.*

Whether or not one believes that the Medicare Prescription Drug, Improvement, and Modernization Act (MMA)¹ in fact improves or modernizes Medicare, the legislation obviously changes the program radically. The extent and nature of these changes make the MMA the most important piece of health care legislation to be adopted by Congress to date in this young millennium.² The MMA also contains what are arguably the most important amendments to the Medicare program since its creation.³ This Essay first describes the identifying characteristics of the current Medicare program, then examines the significant changes that the MMA makes in the program, and finally discusses the importance—and danger—of these changes.

I. TRADITIONAL MEDICARE

The Medicare program, as it was created in 1965 and has evolved over the past four decades, exhibits a number of distinguishing characteristics. First, Medicare is an entitlement program.⁴ According to the statutes

* Robert L. Willett Family Professor, Washington and Lee University School of Law.

1. Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Pub. L. No. 108-173, 117 Stat. 2066.

2. Indeed, the only legislation that could compete with the MMA for the title of most important health care legislation of the past decade would be the Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936.

3. The Medicare statute has been amended almost annually since the early 1970s, but most of the amendments have brought about only marginal changes in the program. See Timothy S. Jost, *Governing Medicare*, 51 ADMIN. L. REV. 39, 66 (1999). The MMA's most significant competitors would probably be the Medicare Catastrophic Care legislation of 1988, which also added a drug benefit as well as catastrophic care coverage, but was repealed a year later, and the Balanced Budget Act of 1997, which added the Medicare+Choice Program.

4. The word "entitled" appears over one hundred times in Title XVIII of the Social

creating both the Part A Hospital Insurance Program and the Part B Supplemental Insurance Program, any person who has qualified for benefits⁵ is legally entitled to go to any health care professional, institutional provider, or health care supplier in the United States that participates in the Medicare program (that is, most health care professionals and virtually all health care institutions in the United States) and receive any medically necessary product or service covered by Medicare.⁶ Medicare will then pay for the service after any cost-sharing obligations of the beneficiary have been met.⁷ A person denied eligibility or denied coverage for a particular service may appeal through a multi-layered appeals process and, ultimately, may seek judicial review.⁸

A second characteristic of the traditional Medicare program is that it does not discriminate among its beneficiaries with respect to premiums, cost-sharing, or coverage. All beneficiaries of the traditional Medicare program—from the poorest to the wealthiest—face the same cost-sharing obligations for Parts A and B and are expected to pay the same premiums.⁹ In fact, the premiums and cost-sharing obligations of very low income beneficiaries are covered by the federal/state Medicaid program, and wealthier beneficiaries often purchase individual Medicare supplement policies or have retiree benefits which cover their cost-sharing obligations,

Security Act (the Medicare title) in, for example, phrases referring to individuals' entitlement to benefits. *See, e.g.*, 42 U.S.C. §§ 426, 1395c, 1395d, 1395f(e), 1395k (2000); *see also* TIMOTHY S. JOST, *DISSENTIMENT* 30-32 (2003).

5. Any legal U.S. resident who is sixty-five or older or permanently disabled, who has paid sufficient quarters of Medicare payroll taxes or who, alternatively, enrolls in Medicare Part A and B and pays a Part A premium is eligible for Part A. 42 U.S.C. §§ 426, 1395c, 1395i-2. Persons who are over 65 or disabled are also eligible for Part B, 42 U.S.C. § 1395o; however, Part B is a voluntary program, and persons who opt to enroll in it must also pay Part B premiums. These premiums are currently equal to about one-quarter of the cost of the program, while the rest is subsidized by general revenue funds. *See* Press Release, U.S. Dep't of Health & Human Servs., HHS Announces Medicare Premium, Deductibles for 2005 (Sept. 3, 2004), <http://www.hhs.gov/news/press/2004pres/20040903a.html>.

6. *See* JOST, *supra* note 4, at 38-45.

7. *See* BARRY R. FURROW ET AL., *HEALTH LAW HORNBOOK* §§ 11-10 to 11-22 (2d ed. 2000) (describing Medicare payment provisions).

8. 42 U.S.C. § 1395ff(b) (2000).

9. *Id.* §§ 1395e, 1395l, 1395r. Those few beneficiaries who enroll voluntarily in Part A because they lack coverage based on their payroll contributions also pay the same premiums. *Id.* § 1395o. There are a few exceptions to this general principle, one of the most important of which is that enrollees who do not enroll in Part B at the date they become eligible for enrollment face a penalty for late enrollment. *Id.* § 1395r(b).

but Medicare itself treats all the same. Moreover, the same products and services are generally available to all beneficiaries regardless of where they live, although Medicare coverage varies to a limited extent across the country because Medicare contractors are authorized to make their own local coverage determinations.¹⁰ In this respect, Medicare resembles the national health services and social insurance programs of most other developed countries, under which all, or virtually all, citizens are covered, and all receive the same benefits.¹¹

Third, Medicare pays for virtually all covered products and services on an administered price basis. Initially, Medicare paid for services based on reasonable charges or reported costs, following the model relied on by the Blue Cross and Blue Shield plans of the time, but this proved wildly inflationary.¹² Beginning with diagnosis-related hospital prospective payment in 1983, Medicare has moved steadily toward administered price systems, with the movement virtually completed by the passage of the Balanced Budget Act of 1997.¹³ In practice, administered prices for health care products and services are set through a rather messy mixture of technical analysis and interest group politics. The Medicare payment under the resource-based relative value scale for a particular physician's service, for example, is based on a formula that includes a component for the physician's work,¹⁴ as well as components for the practice and malpractice costs associated with a given procedure (adjusted for geographic variation). These components are summed and multiplied by a conversion factor based on a "sustainable growth rate" to reach a final payment amount.¹⁵ The process is not purely technical, however. In response to physician arguments that they are underpaid and will not participate in the program unless they are paid more, Congress has consistently stepped in to upwardly adjust the payments that would have

10. Local medical review policies often address new technologies and utilization management issues. See Susan Bartlett Foote et al., *Resolving the Tug-of-War Between Medicare's National and Local Coverage*, 23 HEALTH AFF. 108 (2004).

11. See Timothy Stoltzfus Jost, *Why Can't We Do What They Do?*, 32 J.L. MED. & ETHICS 433 (2004).

12. FURROW ET AL., *supra* note 7, § 11-10.

13. *Id.* § 11-16.

14. The component of the formula that accounts for the physician's work is based on a technical evaluation of the time, mental effort, psychological stress, technical skill, and physical effort expended in producing a particular service.

15. See 42 U.S.C. § 1395W-4; FURROW ET AL., *supra* note 7, § 11-20 (describing 42 U.S.C. § 1395W-4).

resulted from this formula.¹⁶ In the end, however, this technical and political process has proved very successful as a cost-control strategy, holding increases in the cost of the Medicare program below increases in the private sector throughout the late 1990s and early 2000s, and indeed leading to an absolute decrease in the cost of the program during one of those years.¹⁷

Fourth and finally, traditional Medicare covered a limited bundle of products and services. Medicare was patterned after the Blue Cross and Blue Shield programs that dominated the health insurance industry at the middle of the twentieth century, and, like them, was focused on hospital and physician services.¹⁸ From the outset it covered a few other services as well (such as home health care and a limited amount of nursing home care), and with each decade its coverage expanded at the margins (most recently to cover more preventive care). In general, however, Medicare's coverage has in recent years been more limited than that found in commercial employee benefit plans.

As Medicare has evolved over the decades, it has strayed to some extent from these basic patterns. In particular, Medicare beneficiaries have long been able to receive both Part A and Part B benefits through Medicare managed care plans. Under the 1997 Balanced Budget Act the Medicare managed care program was renamed "Medicare+Choice" and designated as a new Part C. Medicare+Choice members were not entitled to obtain services from any Medicare-participating provider, but were instead limited in most instances to providers participating in their managed care plans.¹⁹ Service coverage varied among Medicare+Choice plans, with many covering drug benefits or preventive services similar to those covered by commercial managed care plans and not otherwise covered by Medicare.²⁰ Moreover, providers who participated in Medicare+Choice plans were paid on the basis of prices they negotiated with the plans, rather than based on prices set by the Medicare program.²¹

16. It did so, in particular, in the MMA, Pub. L. No. 108-173, sec. 601.

17. See Cristina Boccuti & Marilyn Moon, *Comparing Medicare and Private Insurers: Growth Rates in Spending over Three Decades*, 22 HEALTH AFF. 230 (2003).

18. See Nancy-Ann DeParle, *Medicare at 40: A Mid-Life Crisis?*, 7 J. HEALTH CARE L. & POL'Y 70, 76-77 (2004).

19. See GERALDINE DALLEK ET AL., LESSONS FROM MEDICARE+CHOICE FOR MEDICARE REFORM 4 (Commonwealth Fund Policy Brief No. 658, June 2003), http://www.cmwf.org/usr_doc/Dallek_lessonsM+C_658.pdf.

20. *Id.* at 3.

21. *Id.* at 4.

Yet, in most respects, the Medicare+Choice program still fits the basic Medicare model. For example, Medicare+Choice plans were required to cover all of the services covered by traditional Medicare.²² The Medicare+Choice plans themselves were paid on the basis of administered prices set by Medicare through the use of a statutory formula, which allowed the Medicare program to control its costs.²³ Most importantly, the vast majority of beneficiaries remained enrolled in traditional Medicare, and in the last few years the Medicare+Choice program shrank dramatically in size.²⁴

II. THE MEDICARE MODERNIZATION ACT

A. The Medicare Drug Benefit

The MMA promises to produce far greater changes in the traditional Medicare program and to change Medicare managed care as well. First, the MMA modernizes the Medicare benefits package—particularly by adding coverage for outpatient prescription drugs and additional preventive services, such as an initial screening physical and cardiovascular and diabetes screening tests.²⁵ The Medicare benefits package has been expanding for a number of years, and the package of benefits offered under the MMA bears a heightened resemblance to the packages of current commercial plans.

As the MMA expands Medicare benefits, however, it also makes them less uniform. Indeed, the changes that the MMA works in the Medicare benefit package go to the fundamental nature of the Medicare entitlement. The new prescription drug program will be provided through private prescription drug plans (PDPs) or Medicare Advantage (MA) managed care plans, which have replaced Medicare+Choice plans under an expanded new program. The statute states that a Medicare beneficiary is entitled to enroll in a qualified prescription drug plan or to obtain prescription drugs through a Medicare Advantage program. The Medicare program is supposed to ensure that each beneficiary has a choice between

22. See COLLEEN L. BARRY & JANET KLINE, MEDICARE MANAGED CARE: MEDICARE+CHOICE AT FIVE YEARS 3 (2002), http://www.cmwf.org/programs/medfutur/barry_fiveyears_ib_537.pdf.

23. *Id.* at 3-4.

24. DALLEK ET AL., *supra* note 19, at 2-4.

25. Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Pub. L. No. 108-173, secs. 101, 611-613, 117 Stat. 2066-2152, 2303-06 (2003).

at least two plans (either two PDPs or MAs).²⁶ But beneficiaries will only have this choice if private plans choose to participate in Medicare. Beneficiaries are only entitled to a drug benefit if PDPs or MA plans offer the benefit, and only on the terms that those plans choose to offer.²⁷ Each beneficiary will have to decide individually whether the prescription drug benefits offered by the available PDP or MA plans are worth what he or she will have to pay for those benefits.

The PDPs are not, in fact, wholly free to define the benefit packages that they will offer. The MMA sets those terms generally by describing a "standard benefit package." The standard benefit package under the program will include a \$250 deductible and a twenty-five percent co-payment for the first \$2000 in benefits beyond that. Beneficiaries whose drug costs exceed \$2250, but do not reach \$5100, fall in a "doughnut hole" and will have to pay all of their costs in excess of \$2250 out-of-pocket.²⁸ Once a beneficiary's out-of-pocket costs exceed \$3600 (i.e., when total costs exceed \$5100), catastrophic coverage kicks in, and thereafter the beneficiary must pay only five percent in co-insurance.²⁹ Both thresholds for co-payment will grow over the years if only because of inflation: In seven years, the program's deductible is projected to grow to \$445, and the catastrophic protection threshold to \$9066 in total drug spending.³⁰ Each PDP must offer coverage for at least two drugs from each therapeutic drug category or class.³¹ The average beneficiary will have to pay about \$35 per month for this package of benefits, initially,³² with the federal government picking up roughly three-quarters of the cost for the basic package plus

26. *Id.* sec. 101(a)(2), § 1860D-3, 117 Stat. at 2081-82 (providing that participants have a choice of plans).

27. If not enough plans offer to provide coverage on a risk-bearing basis, the Centers for Medicare and Medicaid Services can enter into contracts with "fallback" plans to provide coverage with Medicare bearing the full risk. *Id.* § 1860D-3(3)(b)(2), -11(g), 117 Stat. at 2081-82, 2092-99. Even this approach, however, will only work if a plan agrees to contract with Medicare on a fallback basis. Medicare is not authorized anywhere in the statute to provide prescription drug coverage itself.

28. *See id.* § 1860D-2(b)(4), 117 Stat. at 2076-77.

29. *Id.*

30. GERALDINE DALLEK, THE HENRY J. KAISER FOUND., CONSUMER PROTECTION ISSUES RAISED BY THE MEDICARE PRESCRIPTION DRUG, IMPROVEMENT, AND MODERNIZATION ACT OF 2003, at 3 (July 2004), <http://www.kff.org/medicare/loader.cfm?url=/commonspot/security/getfile.cfm&PageID=43996>.

31. *See* Pub. L. No. 108-173, sec. 101(a)(2), § 1860D-4(b)(3)(C), 117 Stat. 2066, 2085; 69 Fed. Reg. 46632 (proposed Aug. 3, 2004) (to be codified at 42 C.F.R. § 423.120(b)(2)).

32. DALLEK, *supra* note 30, at 3.

reinsurance for high-cost insureds.³³

The much-discussed parameters of the standard plan, however, almost certainly do not describe the prescription drug plan that most Medicare beneficiaries will be offered. PDPs may in fact offer any package they choose, as long as it is “actuarially equivalent” to the standard benefit package described in the MMA and meets other legal requirements.³⁴ Most plans will probably offer tiered benefit coverage, with different “copays” for generics, formulary brand name drugs, and nonformulary brand name drugs.³⁵ Each plan will also come up with its own formulary, and each may change its formulary or the preferred or tier status of drugs at will as long it gives appropriate notice (usually thirty days).³⁶ Thus, although beneficiaries will only be able to change plans once a year, the plans can change their benefits at any time more than thirty days after the beginning of the contract year, making for an odd and one-sided market.³⁷ A beneficiary who is denied coverage or preferred-tier status for an off-formulary drug can request coverage and appeal a denial of coverage if the beneficiary’s physician determines that the preferred or formulary drugs would not be as effective or would have adverse effects on the beneficiary.³⁸ But beneficiaries will otherwise have to live with the formulary and cost-sharing structure of the plan they have chosen, no matter how much this structure changes. Finally, premiums will also vary from plan to plan, based on the bids submitted by the plans.³⁹ Some plans will likely cost less than \$35, others more.

As noted above, enrollment in the new Medicare prescription drug

33. Pub. L. No. 108-173, sec. 101(a)(2), § 1860D-4(b)(3)(c), -11(g)(6), -13(a)(3), -15, 117 Stat. at 2085, 2098, 2103, 2113-20.

34. *Id.* § 1860D-2(c), 117 Stat. at 2079-80.

35. See DALLEK, *supra* note 30, at 9, 21-22; HAIDEN A. HUSKAMP & NANCY L. KEATING, THE NEW MEDICARE DRUG BENEFIT: POTENTIAL EFFECTS OF PHARMACY MANAGEMENT TOOLS ON ACCESS TO MEDICATIONS (July 2004), <http://www.kff.org/medicare/loader.cfm?url=/commonspot/security/getfile.cfm&PageID=40660>. A formulary is a list of drugs covered by the plan and usually includes drugs with respect to which the PDP has negotiated a favorable price with a drug manufacturer.

36. Pub. L. No. 108-173, sec. 101(a)(2), § 1860D-4(b)(3)(E), 117 Stat. at 2085; *see* 69 Fed. Reg. 46632 (proposed Aug. 3, 2004) (to be codified at 42 C.F.R. § 423.120(b)(5)).

37. *See* 69 Fed. Reg. 46632 (proposed Aug. 3, 2004) (to be codified at 42 C.F.R. §§ 423.120(b)(6), 423.36).

38. Pub. L. No. 108-173, sec. 101(a)(2), § 1860D-10(h), 117 Stat. at 2091.

39. *Id.* § 1860D-13(a), 117 Stat. at 2102-07; Ctrs. for Medicare & Medicaid Servs., Choices for Drug Coverage (Aug. 4, 2004), at <http://www.cms.hhs.gov/medicarereform/issueoftheday/08042004iotd.pdf>.

benefit will be voluntary. Enrollment in Medicare Part B has always been voluntary as well, but because the terms of Medicare Part B are so favorable, because private insurance is effectively unavailable to people over sixty-five, and, perhaps most importantly, because a person who is eligible must affirmatively opt out of the program to be excluded from it,⁴⁰ enrollment has been almost universal.⁴¹ Enrollment in the drug benefit is likely to be more selective, with many young and healthy beneficiaries concluding that the program is not worth its cost. If premiums in fact cost about \$35 per month, only beneficiaries whose drug costs exceed that figure (\$810 a year) will benefit from the program; currently, seventeen percent of beneficiaries face annual drug costs below \$250 per year.⁴² Beneficiaries who choose not to enroll when first eligible face significant penalties if they choose to enroll later, but they may not realize this fact until it is too late.⁴³ Many Medicare beneficiaries, therefore, likely will not participate in the program.

Many Medicare beneficiaries will also opt out of the drug program because they are otherwise covered by employment-related retiree drug benefit programs. It has always been the case that some Medicare beneficiaries have enjoyed richer health care coverage than others because they receive supplemental benefits under retiree programs. The MMA goes further, however, explicitly subsidizing employee benefit programs that offer benefits at least as generous as the Part D benefit.⁴⁴ The regulations proposed to implement the program suggest several approaches that retiree drug benefit programs may take to fulfill this role, but the bottom line is that many retiree plan members will be receiving federally subsidized drug coverage through private programs that will replace, rather than supplement, Medicare.⁴⁵ Members of Medicare Advantage plans are also likely to face different premiums and cost-sharing structures for their drug benefits than those who receive drugs through the

40. See 42 U.S.C. § 1395p (automatic enrollment provisions); *id.* § 1395r (premium setting provisions).

41. The participation rate for Part B is 95.5 %. Dahlia K. Remler & Sherry A. Glied, *What Other Programs Can Teach Us: Increasing Participation in Health Insurance Programs*, 93 AM. J. PUB. HEALTH 67, 68 tbl.1 (2003).

42. DALLEK, *supra* note 30, at 16.

43. *Id.* at 15-16; Pub. L. No. 108-173, sec. 101(a)(2), §1860D-13(b)], 117 Stat. at 2104-06.

44. Pub. L. No. 108-173, sec. 101(a)(2), § 1860D-22(a), 117 Stat. at 2125-27.

45. See Ctrs. for Medicare & Medicaid Servs., *The Retiree Drug Subsidy: More Secure Coverage for Retirees* (Aug. 11, 2004), at <http://www.cms.hhs.gov/medicarereform/issueoftheday/08112004iotd.pdf>.

traditional program, leading to further variation in the program.⁴⁶

B. Means-Testing Medicare

Some Medicare beneficiaries, moreover, will not have to pay any premium and will face much lower cost-sharing obligations for the drug benefit. The MMA for the first time means-tests part of the Medicare program. Dual-eligibles—i.e., persons who are eligible for both Medicare and Medicaid—will not need to pay any premiums for the Medicare drug program, while low-income beneficiaries with incomes slightly above the Medicaid eligibility level will receive assistance with premiums and cost-sharing.⁴⁷ High-income beneficiaries, on the other hand, will have to pay higher premiums than other beneficiaries under the MMA for the Part B program, albeit not for the drug benefit. Beginning in 2007, higher premiums will be charged to those with incomes above \$80,000 for an individual, \$160,000 for a couple.⁴⁸

Beginning in 2010, moreover, MA plans in some regions will be placed in direct competition with the traditional Medicare program, with the possibility that premiums for traditional Medicare beneficiaries will increase so that Medicare beneficiaries in some parts of the country will pay higher Part B premiums than beneficiaries in other parts of the country.⁴⁹

In sum, the simplicity of the traditional Medicare program—in which every one paid the same premiums and got the same services—is radically broken down under the MMA. This change will undoubtedly benefit some—Medicaid recipients in some states, for example, will receive more generous drug benefits—but the virtues of equality and solidarity that attended the old program will be lost, probably forever. These changes, however, may ultimately also sacrifice the widespread political support the program has always enjoyed, as some healthier and wealthier beneficiaries may realize that they receive fewer benefits or pay higher premiums than their neighbors and may begin to question whether they would be better

46. Medicare Advantage plans may, for example, offer enhanced drug coverage for no additional premium if they are required under Medicare Advantage payment formulas to offer additional benefits. *See* Ctrs. for Medicare & Medicaid Servs., New Medigap Options and Supplemental Options (Aug. 30, 2004), at <http://www.cms.hhs.gov/medicarereform/issueoftheday/08302004ioid.pdf>.

47. Pub. L. No. 108-173, sec. 101(a)(2), § 1860D-14, 117 Stat. at 2107-13.

48. *Id.* sec. 811, 117 Stat. at 2364-65.

49. *Id.* sec. 241, 117 Stat. at 2214-21.

off without a public program.⁵⁰ Experience with the United States's Medicaid program and the experiences of other countries have shown that where health care benefits and costs are shared universally or nearly universally, programs enjoy strong political support, but where wealthier persons opt out, public programs are weakened.⁵¹

C. Payment for Services

The most radical change wrought by the MMA, however, may not be in the way it provides benefits, but rather in the way it pays for services. As noted above,⁵² Medicare has in recent years paid for most products and services using an administered price system. The Medicare program will not pay for outpatient prescription drugs through this approach, however—indeed, the MMA expressly forbids the program from doing so.⁵³ Rather, PDPs will negotiate drug prices directly with pharmaceutical companies. The PDPs will in turn submit bids to Medicare, as noted above, and will be paid based on these bids, with Medicare paying approximately three quarters of a weighted average bid and the beneficiary paying the amount by which the PDP's premium exceeds this amount.⁵⁴ MA plans, after a transition period, will be paid based on a bidding process, but they will also receive heavy subsidies from the federal government, which will be even more substantial in some markets.⁵⁵

There seem to be two reasons why Congress has abandoned administered prices in the MMA. First, many in Congress genuinely believe that managed competition—that is, requiring PDPs or MAs to compete with each other through a managed bidding process that results in beneficiaries having to choose among plans based on marginal premiums—is the best way to control the cost of health care. There are several ironies in this reasoning. First, the generous payment formulas initially being used initially to pay MA plans, the subsidies built into the MA program, and the budget projections from the U.S. Congressional

50. See Jacob S. Hacker & Theodore R. Marmor, *Medicare Reform: Fact, Fiction and Foolishness*, PUB. POL'Y & AGING REP., Fall 2003, at 1, 20-23 (discussing the perils of means-testing Medicare).

51. See JOST, *supra* note 4, at 50-51, 270-73.

52. See *supra* text accompanying notes 13-15.

53. Pub. L. No. 108-173, sec. 101(a)(2), § 1860D-11(i)], 117 Stat. at 2098.

54. See *id.* sec. 101(a)(2), § 1860D-4(b)3(c), -11(g)(6), -13(a)(3), -15, 117 Stat. at 2085, 2098, 2103, 2113-20. Medicare will also provide PDPs with reinsurance for high cost beneficiaries. *Id.* § 1860D-15(b), -15(e), 117 Stat. at 2114-15, 2116-20.

55. *Id.* sec. 222, 117 Stat. at 2193.

Budget Office clearly indicate that the MA program will not save money in the foreseeable future; on the contrary, it is expected to cost a great deal more than traditional Medicare.⁵⁶ Second, it is unlikely that the drug and insurance lobbies that supported the MMA would have fought so hard for it if they had believed that it would result in them earning *less* than they would have under an administered price system. Third, years of experience with Medicare managed care have proven that managed care plans cannot, and will not, provide care at a lower cost than traditional Medicare.⁵⁷ Fourth, every other developed country has controlled health care costs better than the United States, and most of them use some form of administered prices—none use managed competition.⁵⁸ Nevertheless, such is the power of ideology that many members of Congress view managed care and managed competition as potent tools for controlling Medicare costs, despite all evidence to the contrary.

With respect to paying for drugs, however, more is at stake. Congress is very reluctant to allow Medicare to set prices for drugs in the same way that Medicare sets prices for hospital or physician care. This reticence undoubtedly has something to do with the clout of the pharmaceutical industry, which spends far more on lobbying Congress than does any other health care interest group.⁵⁹ But even independent of any lobbying efforts, members of Congress might genuinely fear that if administered prices for drugs are set too low, drug companies may cut back on their investment in research and development, in turn retarding efforts to find new miracle cures.

The question of how to pay for drugs is a complex topic and cannot be addressed fully here.⁶⁰ The arguments for Medicare cutting payments for

56. See BRIAN BILES ET AL., THE COST OF PRIVATIZATION: EXTRA PAYMENTS TO MEDICARE ADVANTAGE PLANS (May 2004), http://www.cmwf.org/usr_doc/biles_extrapayments_ib_750.pdf. Medicare Advantage cost 8.4% more than traditional Medicare—\$552 for each beneficiary. *Id.* at 1.

57. DALLEK ET AL., *supra* note 19, at 6-7.

58. See Timothy S. Jost, *Why Can't We Do What They Do? National Health Reform Abroad*, 32 J.L. MED. & ETHICS (forthcoming 2004) (describing how other countries manage health care costs).

59. See Steven H. Landers & Ashwini R. Sehgal, *Health Care Lobbying in the United States*, 116 AM. J. MED. 474, 474 (2004) (noting that pharmaceutical companies spent \$96 million on lobbying in 2000, compared to \$46 million for physicians and other health professionals).

60. See Ernst R. Berndt, *Unique Issues Raised by Drug Benefit Design*, 23 HEALTH AFF. 103 (2004) (discussing Newhouse's article); Joseph P. Newhouse, *How Much Should Medicare Pay for Drugs?*, 23 HEALTH AFF. 89 (2004) (discussing the role Medicare should play in "setting

drugs are well known: Drug companies make very high profits and spend a great deal on marketing practices that have questionable value to society, such as direct-to-consumer advertising and wining and dining doctors.⁶¹ Drug manufacturers are protected to a considerable degree from the normal pressures of competitive markets by the generous patent and market exclusivity protection that their products are afforded by federal law.⁶² Much of the cost of drug research is already borne by the federal government, which pays tens of millions of dollars for research each year, and which should get some return on its investment.⁶³ And much of the research being carried on by drug companies presently is not directed at “miracle cures,” but rather at lifestyle and “me-too” drugs.⁶⁴ Finally, citizens of other countries pay far less for drugs than we do and seem to have quite adequate access to drugs.⁶⁵

On the other hand, the United States likely subsidizes pharmaceutical research for the rest of the world, and if a major payer like Medicare were to cut drug prices sharply, the cuts would not all come out of marketing or profits.⁶⁶ The task of setting administered prices for drugs, moreover, is terribly difficult, and it is not clear that other countries are doing it right.⁶⁷ In the end, Congress found it easier to punt the problem of establishing payments for drugs over to the PDPs, which the pharmaceutical companies apparently believe will not interfere significantly with their profits.

In the long run, however, the prescription drug program and the Medicare Advantage program are going to prove very costly, a fact that became increasingly clear after the MMA was adopted and information on

manufacturers’ prices”).

61. See Timothy Stoltzfus Jost, *Pharmaceutical Research and Manufacturers of America v. Walsh: The Supreme Court Allows the States To Proceed with Expanding Access to Drugs*, 4 YALE J. HEALTH POL’Y L. & ETHICS 69, 74-75 (2004).

62. DAVID G. ADAMS ET AL., 2 FUNDAMENTALS OF LAW AND REGULATION 180-84 (1997).

63. MICHAEL E. GLUCK, THE KAISER FAMILY FOUND., *FEDERAL POLICIES AFFECTING THE COST AND AVAILABILITY OF NEW PHARMACEUTICALS* 17 (July 2002), <http://www.kff.org/rxdrugs/loader.cfm?url=/commonspot/security/getfile.cfm&PageID=14078>.

64. See Thomas W. Croghan & Patricia M. Pittman, *The Medicine Cabinet: What’s in It, Why, and Can We Change the Contents?*, 23 HEALTH AFF. 23 (2004).

65. Gerard F. Anderson et al., *Doughnut Holes and Price Controls*, HEALTH AFF. W4-396, W4-396 (Web Exclusive July 21, 2004), at <http://content.healthaffairs.org/cgi/reprint/hlthaff.w4.396v1>.

66. Cf. Patricia M. Danzon, *Closing the Doughnut Hole: No Easy Answers*, HEALTH AFF. W4-405, W4-406 (Web Exclusive July 21, 2004), at <http://content.healthaffairs.org/cgi/reprint/hlthaff.w4.405v1>.

67. *Id.*

the legislation's cost embargoed during the congressional debate finally leaked out.⁶⁸ This brings us to the final provision of the MMA—its doomsday clause.⁶⁹ This provision requires the trustees of the Medicare trust funds to project each year whether during that year or any of the six succeeding years the proportion of Medicare expenditures funded by general revenue funds (i.e., that component not covered by beneficiary premiums, Part A payroll taxes, or other dedicated sources) is likely to exceed forty-five percent.⁷⁰ If, for two years in a row, the trustees project that this will happen during the current fiscal year or the succeeding six years, the President must take action.⁷¹ Specifically, the President must within fifteen days present to Congress proposed legislation to eliminate the “excess general revenue funding” problem. Legislation addressing the problem will then be handled under special rules in the Senate and House and be subject to very limited debate.⁷² Depending on who is the President at the time, this provision could very well open the door for eliminating the Medicare program as we have traditionally known it with very little opportunity for Congress to debate the change.

CONCLUSION

In sum, the MMA makes radical changes in the Medicare program, undermining in particular the commitment to solidarity and equality that has kept the program politically strong and loosening the cost controls that have kept it fiscally sustainable. The statute expands benefits for many beneficiaries in the short run, but adopts an approach to financing and distributing those benefits that may not be viable in the long run. In decades to come we may well look back at the MMA, the most important Medicare legislation of its decade, as a statute whose primary lasting effect was not to modernize the Medicare program, but rather to doom it.

68. See BILES ET AL., *supra* note 56 (discussing the high costs of privatization); U.S. Gov't Accountability Office, Dep't of Health & Human Servs., Chief Actuary's Communication with Congress, B-302911 (Sept. 7, 2004), <http://www.gao.gov/decisions/appro/302911.htm>.

69. Pub. L. No. 108-173, secs. 801-04, 117 Stat. at 2357.

70. *Id.* sec. 801, 117 Stat. at 2357-60.

71. *Id.* sec. 802, 117 Stat. at 2360.

72. *Id.* secs. 803-04, 117 Stat. at 2360-64.

The Swinging Pendulum: The Supreme Court Reverses Course on ERISA and Managed Care

Aaron S. Kesselheim, M.D., J.D.* and Troyen A. Brennan, M.D., J.D., M.P.H.[†]

INTRODUCTION

The critical issue in health policy is the cost of health care, and its importance will only rise further with the changing demographics of the U.S. population. The last twenty years have seen numerous efforts to control costs, beginning with regulatory mechanisms¹ and later dominated by the market-based approach of managed care.² At its peak, managed care led to a historic decrease in the rate of inflation in health care costs.³ Over the last five years, however, managed care has retreated significantly in favor of consumer-driven health care, in which individual patients are more exposed to the costs of care and thus choose more carefully which services to purchase.⁴

Federal and state regulation, as well as common law litigation, helped hasten the abandonment of managed care and the subsequent embrace of consumer-driven health care.⁵ In the 1980s and early 1990s, as a result of

* Division of Internal Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School.

† Professor of Medicine, Harvard Medical School and Professor of Law and Public Health, Harvard School of Public Health. The authors thank David Studdert and Michelle Mello for their insights on earlier drafts.

1. See TROYEN A. BRENNAN & DONALD M. BERWICK, *NEW RULES: REGULATION, MARKETS, AND THE QUALITY OF AMERICAN HEALTH CARE* (1996).

2. See JAMES C. ROBINSON, *THE CORPORATE PRACTICE OF MEDICINE* (1999).

3. See Katharine Levit et al., *Health Spending in 1998: Signals of Change*, 19 HEALTH AFF. 124 (2000).

4. See James C. Robinson, *Reinvention of Health Insurance in the Consumer Era*, 291 JAMA 1880 (2004).

5. See M. Gregg Bloche & David M. Studdert, *A Quiet Revolution: Law as an Agent of Health System Change*, 23 HEALTH AFF. 29 (2004).

the Employee Retirement Income Security Act (ERISA),⁶ insurance companies designed managed care strategies without significant state supervision.⁷ ERISA established national standards for employer-sponsored benefit plans, rendering certain state laws inapplicable for enrollees in employer-sponsored health plans and limiting their ability to sue their managed care companies. As consumer dissatisfaction with elements of managed care grew, however, federal courts found themselves under increasing pressure to restrict this pre-emptive effect and allow more effective oversight of managed care tactics.

Starting in 1995, and fueled by three decisions since 2000, the Supreme Court has helped spur a judicial movement to limit the boundaries of ERISA preemption. The Court seemingly acceded to popular concern about the role that courts' support for ERISA preemption had originally played in the growth of managed care. This retreat weakened managed care, as executives and shareholders of managed care companies grew concerned about the costs of litigation and complying with state regulation, and the publicity gave further voice to the opposition to managed care practices.⁸ Experts agreed, "[T]he free ride enjoyed by health maintenance organizations is now over."⁹

Thus it is an extraordinary surprise that in its recent decision in *Aetna Health v. Davila*,¹⁰ which involved a state statute intended to protect managed care enrollees, the Supreme Court reversed course and reiterated its pre-1995 broad ERISA preemption doctrine. Few, if any, health law experts anticipated this event.¹¹ In the wake of this reversal for consumer and physician interests, we reexamine the development of ERISA law prior to *Davila* and present a vision of the immediate implications of the decision. We conclude by suggesting that *Davila* may represent the first swing of the pendulum back toward managed care.

A REVIEW OF MANAGED CARE LITIGATION AND REGULATION

Much has been written about the development of managed care in this country.¹² Briefly put, due to the benefits of the Health Maintenance

6. 29 U.S.C. §§ 1001-1461 (2000).

7. See Wendy K. Mariner, *What Recourse?—Liability for Managed-Care Decisions and the Employee Retirement Income Security Act*, 343 NEW ENG. J. MED. 592 (2000).

8. See James C. Robinson, *The End of Managed Care*, 285 JAMA 2622 (2001).

9. Marc I. Machiz, *Hidden Blow to HMOs*, NAT'L L.J., July 3, 2000, at A19.

10. *Aetna Health Inc. v. Davila*, 124 S. Ct. 2488 (2004).

11. See Bloche & Studdert, *supra* note 5, at 35.

12. See, e.g., WALTER A. ZELMAN & ROBERT A. BERENSON, *THE MANAGED CARE BLUES AND*

Organization (HMO) Act of 1973,¹³ nascent managed care organizations (MCOs) gained a competitive advantage over traditional insurers in terms of the premiums that they could offer to employers. HMOs, particularly those that employed medical staff and emphasized the use of guidelines to dictate care delivery, controlled physician decision-making by inducing compliance with algorithms designed to provide the most cost-effective care.

Seeing reduced expenditures, traditional insurers organized their own managed care plans, relying on tighter networks of physicians who agreed to managed care techniques—such as prospective utilization review, primary care gate-keeping, and relatively careful prior approval.¹⁴ Though the Clinton Health Plan was rejected by Congress, some of its principles still diffused out into the marketplace. As a result, even the hospital industry began to reorganize along the lines of primary care gate-keeping and prospective capitated payment. In many metropolitan areas, hospital utilization and ancillary testing were reduced, and health care inflation slowed to historically low levels.¹⁵

Patients and their advocates, however, began to recognize that many managed care organizational structures reversed the financial incentives in the doctor-patient relationship. In indemnity care, the physicians increased income by providing more care; in managed care, physicians' profit motive was no longer aligned with elaboration of services, and patients began to worry whether that could lead to restrictions on needed care.¹⁶ Patient advocates appealed to state legislatures to help regulate MCOs,¹⁷ and patients brought increasingly potent suits alleging harm by MCOs for denying them appropriate benefits.¹⁸ But these efforts, which challenged care reduction techniques at the heart of managed care, often bumped up against the ERISA preemption doctrine.¹⁹

HOW TO CURE THEM (1998).

13. Federal HMO Act of 1973, Pub. L. No. 93-222, 87 Stat. 914.

14. Robinson, *supra* note 8, at 2624.

15. Levit et al., *supra* note 3, at 124.

16. See Jon Gabel, *Ten Ways HMOs Have Changed During the 1990s*, 16 HEALTH AFF. 134 (1997).

17. See Stephen R. Latham, *Regulation of Managed Care Incentive Payments to Physicians*, 22 AM. J.L. & MED. 399 (1996).

18. *Jones v. Chicago HMO Ltd. of Ill.*, 730 N.E.2d 1119 (Ill. 2000); *Petrovich v. Share Health Plan*, 719 N.E.2d 756 (Ill. 1999).

19. See Alice A. Noble & Troyen A. Brennan, *Managing Care in the New Era of "Systems-Think": The Implications for Managed Care Organizational Liability and Safety*, 29 J.L. MED. & ETHICS 290 (2001).

ERISA AND MANAGED CARE ORGANIZATIONS

ERISA sets a national administrative standard for employer-sponsored pension and benefit plans, making them more palatable for large employers who would otherwise be subject to fifty different state regulatory schemes. The law also establishes a national remedy for failure to provide ERISA-sponsored benefits, limiting damages to the costs of the denied benefit and attorneys' fees. ERISA removes employer-sponsored plans from the control of any state law that "relates to" the management of plan benefits.²⁰ The ERISA "savings clause" allowed legislation that regulates the general business of insurance to stand,²¹ but it was interpreted narrowly in early decisions and does not apply to regulation specifically directed at MCOs.²² Although ERISA does not establish a uniform standard for all health plan administration, it impacts the large number of consumers who obtain health insurance through their employers.²³

This design has had two significant results. First, state laws did not apply to some important business practices of MCOs.²⁴ A state, for example, that required plans to cover annual mammograms for women at a certain age could only apply to government-sponsored or individually-purchased health plans. Second, even if plans inappropriately delayed claims or denied coverage for a treatment, enrollees could not sue under their state's common law of negligence, thus denying them consequential damages, punitive damages, or compensation for emotional distress.²⁵ Instead, such enrollees could receive only the ERISA remedy. When a patient sued her health plan for negligently denying approval for an autologous bone marrow transplant and high-dose chemotherapy in treating her breast cancer, she won an \$89 million damage award in state court.²⁶ Her plan was state-sponsored; if she had been in an employer-sponsored benefit plan, her award would have been limited in federal court to the cost of the procedure and attorneys' fees. This so-called "regulatory vacuum"²⁷ likely contributed to the excesses of managed care

20. See 29 U.S.C. § 1144(a) (2000).

21. See *id.* § 1144(b)(2)(A).

22. See, e.g., *Pilot Life Ins. Co. v. Dedeaux*, 481 U.S. 41 (1987).

23. Peter D. Jacobson & Scott D. Pomfret, *ERISA Litigation and Physician Autonomy*, 283 JAMA 921 (2000).

24. *FMC Corp. v. Holliday*, 498 U.S. 52 (1990).

25. See *Corcoran v. United Healthcare*, 965 F.2d 1321 (5th Cir. 1992).

26. David M. Studdert & Troyen A. Brennan, *The Problems with Punitive Damages in Lawsuits Against Managed-Care Organizations*, 342 NEW ENG. J. MED. 280 (2000).

27. Jacobson & Pomfret, *supra* note 23, at 921.

during that era, as MCOs made large profits exploiting physicians and hospitals in negotiation tactics and limiting services to enrollees.²⁸

As a result, in the 1990s, advocacy groups pushed the federal government to amend ERISA to bring it in line with the modern reality of the health care marketplace.²⁹ While Congress did not reform ERISA, these efforts found a receptive audience in state legislatures, where popular opinion helped inspire legislators to try to rein in the well-publicized excesses of managed care. States protected enrollees by mandating certain inclusions in their health plans—for example, requiring certain benefits like mammograms,³⁰ providing independent review quickly when coverage was denied,³¹ and mandating prompt payment to physicians.³² Meanwhile, aggrieved enrollees looked to hold MCOs accountable for their business practices by making claims, akin to common law negligence arguments, that MCOs were liable for the injury-causing decisions of providers under their control. But as originally interpreted, ERISA preemption made these legislative and judicial efforts moot for many MCO enrollees.

A NEW PREEMPTION ANALYSIS

The Supreme Court's early comments on ERISA preemption solidified a long-standing broad judicial interpretation of the "relates to" clause.³³ In two 1987 cases, the Supreme Court called ERISA "a comprehensive civil enforcement scheme"³⁴ and seemed to favor arming ERISA with "extraordinary pre-emptive power."³⁵ In the 1990s, however, as MCOs expanded their influence over the health care system—and their profits—

28. See Wendy K. Mariner, *State Regulation of Managed Care and the Employee Retirement Income Security Act*, 335 NEW ENG. J. MED. 1986 (1996).

29. Cf. Bipartisan Patient Protection Act, S.1052, 107th Cong. (2001).

30. See Vicki L. MacDougall, *Medical Gender Bias and Managed Care*, 71 OKLA. CITY U. L. REV. 781 (2002).

31. See Aaron S. Kesselheim, *What's the Appeal? Trying To Control Managed Care Medical Necessity Decision Making Through a System of External Appeals*, 149 U. PA. L. REV. 873 (2001).

32. See Healthcare Fin. Mgmt. Ass'n, *State Prompt Payment Laws At-a-Glance*, at http://www.hfma.org/resource/focus_areas/commercial_payment/promptpay.htm (last visited July 1, 2004); Conomikes, *Summary of State Prompt Payment Laws*, <http://www.conomikes.com/PromptPay.htm> (last visited Dec. 5, 2004).

33. See Jana K. Strain & Eleanor D. Kinney, *The Road Paved with Good Intentions: Problems and Potential for Employer-Sponsored Health Insurance Under ERISA*, 31 LOY. U. CHI. L.J. 29 (1999).

34. *Pilot Life Ins. Co. v. Dedeaux*, 481 U.S. 41, 54 (1987).

35. *Metropolitan Life Ins. Co. v. Taylor*, 481 U.S. 58, 65 (1987).

the Supreme Court moved to limit the scope of preemption using three different legal avenues.

First, in *New York State Conference of Blue Cross & Blue Shield Plans v. Travelers*, the Supreme Court re-examined the “relates to” clause in the context of a New York state statute that imposed extra surcharges on patients covered by commercial insurers or HMOs.³⁶ A group of commercial health insurers sued to have the statute invalidated, and the lower federal courts agreed that ERISA preempted this state law that affected employee benefit plans by increasing their costs of doing business. On appeal, however, the Supreme Court unanimously upheld the statute because the impact of the rate-setting law was remote. Justice Souter concluded, “If ‘relate to’ were taken to extend to the furthest stretch of its indeterminacy, then for all practical purposes preemption would never run its course.”³⁷ He placed the first restriction on ERISA’s reach by limiting the scope of the “relates to” section, and in doing so prevented ERISA preemption from undermining a state law designed to reduce health care costs and improve access.

More generally, the Court seemed to change its attitude toward ERISA. Justice Souter confronted the broad path that preemption was carving in health care, remarking that nothing in ERISA indicated “that Congress chose to displace general health care regulation, which is traditionally a matter of local concern.”³⁸ The opinion implied that ERISA would not completely block reform efforts.

The Court demonstrated an awareness of its role, mediated by ERISA, in changing health policy. As a result, some statutes and cases once considered preempted under the broad reading of ERISA now found receptive lower courts. In *Dukes v. U.S. Healthcare*, for example, the plaintiff sued his MCO for not taking reasonable care in selecting and monitoring its physicians. The Second Circuit Court of Appeals held that ERISA did not preempt his claim, since it did not involve withheld or delayed benefits.³⁹ The Fifth⁴⁰ and Eleventh⁴¹ Circuit Courts also found instances where ERISA did not preempt suits against MCOs for negligence.

The *Dukes* distinction predicted the second technique the Supreme

36. N.Y. PUB. HEALTH LAW § 2807-c (McKinney 2002).

37. *N.Y. State Conference of Blue Cross & Blue Shield Plans v. Travelers*, 514 U.S. 645, 655 (1995).

38. *Id.* at 661.

39. *Dukes v. U.S. Healthcare*, 57 F.3d 350 (3d Cir. 1995).

40. *Roark v. Humana*, 307 F.3d 298 (5th Cir. 2002).

41. *Land v. Cigna Healthcare of Fla.*, 339 F.3d 1286 (11th Cir. 2003).

Court employed to limit ERISA—redefining MCO decision-making. In *Pegram v. Herdrich*,⁴² Dr. Lori Pegram, an employee of the physician-owned Carle HMO, found a mass in Cynthia Herdrich's abdomen. Instead of sending her to a local hospital at increased cost to Carle, Dr. Pegram recommended that Herdrich wait eight days for an ultrasound by colleagues within Carle's system. Herdrich's appendix ruptured. In her lawsuit, she charged that Carle failed its fiduciary function by directly influencing its physicians' decisions about medical treatment in such a way that caused her harm. It was a novel effort to find employer-sponsored health plans liable for their coverage decisions within the confines of ERISA, which requires its plans to act as fiduciaries in the best interest of their participants.

However, the Supreme Court refused to extend such a fiduciary duty to eligibility determinations. In a unanimous decision, Justice Souter distinguished between two major forms of health care decision-making—eligibility decisions made by health plan administrators to determine what services the plan might cover and treatment decisions made by providers regarding how to diagnose and manage patients' conditions. Justice Souter called Pegram's decision a "mixed eligibility and treatment decision"⁴³ where the question was not whether the ultrasound was covered, but whether the service was appropriate to use at that particular time. MCOs could not be held liable as a fiduciary, because that would strike at the very basis of managed care itself and "no HMO organization could survive without some incentive connecting physician reward with treatment rationing."⁴⁴

Despite dismissing the fiduciary claim, Justice Souter restricted MCOs' ERISA shield by separating the eligibility decisions from the mixed eligibility and treatment decisions. Whereas ERISA preemption clearly covers eligibility determinations, mixed decisions are not part of a health plan's administrative function. As a result, if MCOs or their agents make such mixed decisions, then it might be possible to hold them liable for negligence in doing so. Since Justice Souter did not set a clear distinction between eligibility and mixed decisions, it might be possible for MCOs' efforts to influence member physicians' practices to fall outside ERISA protection.

The Supreme Court again espoused a critical tone about MCOs in general, as Justice Souter critiqued their role in health care delivery and

42. *Pegram v. Herdrich*, 530 U.S. 211 (2000).

43. *Id.* at 229.

44. *Id.* at 220.

encouraged further debate about the need to oversee decision-making in the managed care system.⁴⁵ His dicta confirmed that the Court was monitoring ERISA's effect on the evolution of the American health care system and might now consider reexamining impediments previous decisions had placed on reform efforts.⁴⁶ The Supreme Court seemed to join the anti-managed care fray.

In two decisions in 2002 and 2003, the Supreme Court revealed a third approach to restricting the ERISA shield by expanding the "savings clause." In the first case, *Rush Prudential v. Moran*, an Illinois statute provided MCO enrollees the right to independent medical review if their MCO denied benefits they felt were contractually owed.⁴⁷ When Debra Moran had persistent pain and numbness in her hand, she sought to have a special procedure performed by a surgeon not associated with her HMO, Rush Prudential. Rush Prudential told her that it would only cover a more standard procedure performed by an affiliated physician. Moran sought independent review, but Rush Prudential denied her request and countered that the Illinois law that "relates to" administration of their benefits was preempted and invalid.

In a narrow 5-4 decision, the Supreme Court sided with Moran.⁴⁸ Rush Prudential lawyers argued that the independent review process was an illegal alternative to ERISA's system. MCO enrollees could avoid suing for the relatively meager ERISA statutory remedies by applying to a state-organized external appeals process. But Justice Souter, again for the majority, wrote that the "[e]ffect of eliminating insurer's autonomy to guarantee terms congenial to its own interests is stuff of regular insurance regulation" and therefore fell under ERISA's savings clause. He supported states' ability to enforce standards of reasonable medical care in the process of regulating insurance companies—as long as states do not come in direct conflict with ERISA.

The scope of the savings clause arose again in the context of another state law in *Kentucky Ass'n of Health Plans v. Miller*.⁴⁹ Most MCOs contract with specific providers to establish selective networks that only members can access. These contracts provide bargaining leverage for MCOs in negotiations with other providers looking to join the network. Kentucky passed a statute forbidding health insurers from discriminating against any

45. *Id.* at 221.

46. *Id.*

47. 215 ILL. COMP. STAT. 125/4-10 (1987).

48. *Rush Prudential HMO v. Moran*, 536 U.S. 355 (2002).

49. *Kentucky Ass'n of Health Plans v. Miller*, 538 U.S. 329 (2003).

provider in a given coverage area who wanted to join the network and could meet the standard MCO conditions for participation.⁵⁰ The Kentucky Association of Health Plans claimed that ERISA preempted this so-called “Any Willing Provider” law.⁵¹

In a unanimous decision authored by Justice Scalia, the Supreme Court ruled that the state law fell under the ERISA savings clause. For a general law that regulates insurance to be saved from ERISA preemption, it only had to substantially affect the risk pooling arrangement, rather than control the actual terms of insurance policies.⁵² Though by 2003 many health plans had voluntarily stopped using selective provider networks to drive cost savings,⁵³ *Kentucky Ass’n of Health Plans* relaxed the savings clause requirements and left the door open for more state regulations to impact other MCO management tactics.

By limiting the “relates to” clause, separating eligibility from mixed decisions, and expanding the savings clause, these Supreme Court decisions seemed consistent in narrowing ERISA’s reach and expanding state influence over employer-sponsored health plans. Finding no legislative relief at the federal level, consumer activists and provider groups continued to work for local regulatory reform. Had ERISA been transformed from an “extraordinarily preemptive power”⁵⁴ and given way to allow MCOs to be subject to tort liability or state control? Many experts believed so. Bloche and Studdert arguably spoke for the majority of health policy analysts when they stated, “the Supreme Court has sounded an ERISA ‘all-clear’ for state regulation of plans’ management practices.”⁵⁵ But the Supreme Court had a surprise in store this past June.

A REVERSAL OF FORTUNE

Davila consolidated several cases that arose in Texas when enrollees in employer-sponsored MCOs attributed their injuries to the decision-making of their health plan administrators. After Ruby Calad underwent a hysterectomy with a rectal, bladder, and vaginal repair, the utilization review nurse for Cigna, Calad’s health plan, arranged for her to be sent

50. KY. REV. STAT. ANN. § 304.17A-270 (Michie Supp. 2003).

51. *Kentucky Ass’n of Health Plans*, 538 U.S. at 332-33.

52. *Id.* at 338.

53. Tanya Albert, *High Court Punches Another Hole in the Federal Law Shielding HMOs*, AMNEWS, Apr. 21, 2003, at <http://www.ama-assn.org/amednews/2003/04/21/gvl10421.htm>.

54. *Rush Prudential HMO v. Moran*, 536 U.S. 355, 376 (2002).

55. Bloche & Studdert, *supra* note 5, at 35.

home after a one-day hospital stay (contrary to the judgment of her doctor). Calad claimed that her early discharge contributed to the unspecified complications that arose a few days later and caused her to have to return to the emergency room.⁵⁶ In another case, Juan Davila suffered from arthritis and was prescribed rofecoxib by his physician. His Aetna HMO coverage required him to first try a less expensive non-steroidal anti-inflammatory drug such as naproxen. After three weeks of treatment with naproxen, he suffered severe gastrointestinal bleeding requiring blood transfusions and a five-day stay in a hospital intensive care unit.⁵⁷

Both Davila and Calad sued under the 1997 Texas Health Care Liability Act (THCLA), which requires MCOs to “exercise ordinary care when making health care treatment decisions” and makes plans liable for damages if they are negligent in meeting the ordinary care standard.⁵⁸ If the plan did not cover the desired health care service, then no liability could arise. The law thus sought to mirror the dichotomy between eligibility and mixed treatment-eligibility decisions set in *Pegram* and impose accountability accordingly. The plaintiffs claimed that their injuries resulted from such mixed decisionmaking. In response, the health plan lawyers invoked ERISA’s preemption over their claims and recast the claims in federal court to make the plaintiffs entitled, at most, to collecting the benefits denied—in Davila’s case, arguably the cost of a rofecoxib prescription.⁵⁹

The Fifth Circuit Court of Appeals gleaned a modern interpretation of ERISA preemption as one preventing states from exactly duplicating the terms of ERISA. Since ERISA provides a “means of collecting benefits,”⁶⁰ and THCLA provides a duty of reasonable care, the Fifth Circuit Court reasoned that the Texas statute fell outside of ERISA. It was a reading of ERISA so narrow as to make the statute completely toothless; no injured MCO enrollees would use the ERISA scheme merely to collect benefits. Rather, all would choose to sue for negligence and the resulting damages under state law.⁶¹

The Supreme Court rejected this opportunity to remove ERISA from the health care regulation equation. With the support of a unanimous

56. *Roark v. Humana, Inc.*, 307 F.3d 298, 302 (2002).

57. *Id.* at 303.

58. TEX. CIV. PRAC. & REM. CODE ANN. § 88.001-.003 (Vernon Supp. 2003).

59. *Aetna Health Inc. v. Davila*, 124 S. Ct. 2488, 2498 (2004).

60. *Id.* at 2499.

61. *Roark*, 307 F.3d at 310.

court, Justice Thomas found that the issue of exercising ordinary care under THCLA was inextricably bound up in the administration of medical services under health plan contracts, which was ERISA's regulatory domain.⁶² Even though THCLA enforced a somewhat different duty than ERISA, the same set of facts could invoke both state and federal law under the Fifth Circuit's interpretation, so state law could be used to completely supplant ERISA. This would go against "Congress' intent to make the ERISA civil enforcement mechanism exclusive."⁶³

More significantly, Justice Thomas readdressed some prior Supreme Court ERISA health law holdings to cast them in a new light. He quoted liberally from the Supreme Court's 1987 cases, and once again classified ERISA as a "comprehensive remedial scheme."⁶⁴ He also closed the door that *Pegram* had opened with regard to mixed decisions. He considered the plan administrators' actions with respect to Davila and Calad as pure eligibility decisions, remarking that only a treating physician also acting as the administrator of health plan coverage decisions can make mixed eligibility-treatment decisions.⁶⁵ His decision effectively placed plan administrators' utilization review decisions back under the ERISA shield for liability purposes.

Finally, he invoked an overpowering federal policy implicit in ERISA to tighten the scope of the savings clause. He limited *Rush Prudential* to its facts, implying that the decision did not support the principle that states could freely formulate novel alternative forms of regulation outside of ERISA, such as independent appeals processes, without fear of preemption.⁶⁶ The dissent in *Rush Prudential*, also written by Justice Thomas, rejected Illinois' independent appeals law, in part due to ERISA's rejection of overlapping remedies.⁶⁷ A unanimous majority now directly invoked a position formerly held by a four-person minority—a sign of how critically the court may view other alternative remedies in the future. Justice Thomas did not even cite *Kentucky Ass'n of Health Plans*, perhaps reflecting his view that, with the Court's new perspective on ERISA, that case's reformulation of the savings clause was of minimal importance to future ERISA jurisprudence. Texas, as well as the other states with similar

62. *Davila*, 124 S. Ct. at 2497.

63. *Id.* at 2499.

64. *Id.* at 2500.

65. *Id.* at 2499.

66. *Id.*

67. *Rush Prudential HMO v. Moran*, 536 U.S. 355, 388 (2002) (Thomas, J., dissenting).

statutes,⁶⁸ found this effort to hold MCOs accountable for negligent coverage decisions to be invalid.

WHAT'S NEXT FOR ERISA HEALTH CARE LAW?

The language reinvigorating ERISA in *Davila* is hard to reconcile with the previous trend of cases limiting the reach of preemption, unless we impute to the Court an awareness of health policy. In *Davila*, the Supreme Court may have finally reached the end of how far it could stretch ERISA. More interesting, however, was the change in tone from previous discussions in *Pegram* or *Rush Prudential*. The Court retreated from language in those decisions that seems to favor local health care regulation over the business tactics MCOs use to administer care. Even Justice Ginsburg's concurrence agreed that the decision is "consistent with our governing case law," that "virtually all state law remedies are preempted," and that the Court's hands were tied by the federal ERISA law as currently framed.⁶⁹ These words should strongly discourage future efforts to reinterpret ERISA at the state legislative or judicial contexts, and we are not likely to see another *Davila*-like case soon. Instead, patient advocates will likely turn their efforts towards federal legislative reform of the ERISA statute itself.

Our explanation for this swing is that the Supreme Court was uncomfortable at the vanguard of the anti-managed care movement. As discussed in the introduction, the Supreme Court, in its own decisions since 1995 and the direction it therefore gave to lower courts, had helped restrict managed care by limiting ERISA preemption. As increased litigation gave voice to the public backlash, stockholders lost interest in those insurers who persisted in capitated managed care.⁷⁰ Managed care collapsed, to be replaced, at least in the rhetoric of health policy, by consumer-driven health care.

But consumer-driven health care has its own problems. First, the theory of consumer choice relies on competitive markets, and there are few signs that such market conditions are developing in the health care

68. At least ten other similar state statutes have been adopted. *See, e.g.*, Managed Health Care Insurance Accountability Act of 1999, 1999 Cal. Legis. Serv. 536 (West); *see also* Anne Gearan, *High Court Hears Test of Patient Protection Laws*, LEGAL INTELLIGENCER, Mar. 24, 2004, at 4.

69. *Davila*, 124 S. Ct. at 2503 (quoting *DeFelice v. Aetna U.S. Healthcare*, 346 F.3d 442, 456 (3d Cir. 2003)).

70. *See* CHARLES BOORADY ET AL., HEALTHCARE SERVICES: MANAGED CARE (2000).

sector.⁷¹ Second, the central features of consumer choice are higher co-payments and deductibles, which represent the thinning out of employer-provided insurance. Reduction of employee benefits is a major theme in the effort to make American companies more competitive,⁷² but this creates problems with access to care and is only a short-term solution to limiting costs. As a result, Robinson has warned that the consumer-driven approach will not be sufficient to control costs and improve quality; some aspects of managed care will have to be revived.⁷³

Perhaps then, in *Davila*, the Supreme Court was being appropriately cautious in not allowing the Texas law to take a further step to cripple the business model of managed care. More to the point, as Justice Ginsburg's concurring opinion relates, the key guidance on health policy must come from Congress, which has done little recently to address ERISA.⁷⁴ Judicial capacity to address such complex issues is limited,⁷⁵ and the Supreme Court's insistence in other cases that it must defer to administrative expertise in health policy indicates that it recognizes this.⁷⁶

Our view of the Supreme Court, then, is that the Justices are more aware of the role their decisions play in health policy than has previously been appreciated, and their intent for now is not to be activist. The effort to reverse course and halt the momentum of the attack on ERISA is important evidence that the Supreme Court will not allow itself to be the instrument of health care reform. That is a role it wants to defer to the legislative branch, as the role of consumer choice, and the re-emergence of managed care, define the policy battleground in the effort to reduce health care costs.

71. Bloche & Studdert, *supra* note 5, at 39-40.

72. Micheline Maynard, *United Appears Boxed in as Trouble Percolates*, N.Y. TIMES, Aug. 1, 2004, at C1.

73. Robinson, *supra* note 4, at 1886.

74. See *Davila*, 124 S. Ct. at 2503; Gail B. Agrawal & Mark A. Hall, *Managed Care Liability Beyond the ERISA Shield*, 47 ST. LOUIS U. L.J. 235 (2003).

75. See Michelle M. Mello & Troyen A. Brennan, *The Controversy over High-Dose Chemotherapy with Autologous Bone Marrow Transplant for Breast Cancer*, 20 HEALTH AFF. 101 (2001).

76. See Michelle M. Mello et al., *The Pharmaceutical Industry Versus Medicaid—Limits on State Initiatives To Control Prescription-Drug Costs*, 350 NEW ENG. J. MED. 608 (2004).

Inequality, Infections, and Community-Based Health Care

Evan Lyon, M.D.* and Paul Farmer, M.D., Ph.D.†

Advocates for better health care for the world's poor are fond of the mantra that "infections know no boundaries." Part of this logic evokes the reality of our global community, connected by the easy and frequent movement of people across national borders. But this mantra is also meant as a warning, reminding those of us in wealthier nations that we just might not be safe from exposure to the poor, huddled, coughing masses. HIV, tuberculosis, and other infectious diseases on the African continent have been declared a U.S. national security priority.¹ When the global extent of the multi-drug resistant tuberculosis epidemic was being uncovered—in part by the community-based efforts of our small non-profit health care organization working in the slums of Lima, Peru²—news of exposure to drug-resistant tuberculosis on international flights made headlines.³ The emergence of SARS and the worldwide fear it evoked mobilized

* Resident in Internal Medicine, Brigham and Women's Hospital. Dr. Lyon is a recent graduate of Harvard Medical School and divides his clinical time equally between Boston and central Haiti.

† Presley Professor of Medical Anthropology, Harvard Medical School. Dr. Farmer, a physician and anthropologist, is the founding director of Partners In Health.

1. NAT'L INTELLIGENCE COUNCIL, NIE 99-17D, THE GLOBAL INFECTIOUS DISEASE THREAT AND ITS IMPLICATIONS FOR THE UNITED STATES (2000), <http://www.cia.gov/cia/reports/nie/report/nie99-17d.html>; Barton Gellman, *AIDS Is Declared Threat to Security*, WASH. POST, Apr. 30, 2000, at A1.

2. Carole Mitnick et al., *Community-Based Therapy for Multidrug-Resistant Tuberculosis in Lima, Peru*, 348 NEW ENG. J. MED. 119, 120 (2003).

3. Thomas A. Kenyon et al., *Transmission of Multidrug-Resistant Mycobacterium Tuberculosis During a Long Airplane Flight*, 334 NEW ENG. J. MED. 933, 933 (1996) ("The transmission of *Mycobacterium tuberculosis* that we describe aboard a commercial aircraft involved a highly infectious passenger, a long flight [Honolulu-Chicago-Baltimore-Chicago-Honolulu], and close proximity of contacts to the index patient."); see also M.A. Miller et al., *Tuberculosis Risk After Exposure on Airplanes*, 77 TUBERCULE & LUNG DISEASE 414, 415 (1996) ("The index case [a Russian refugee] in this study flew from Moscow, Russia to Frankfurt, Germany in March 1993. There he boarded a flight originating in Bombay, India destined for New York City. In New York, he changed aircraft and flew to Cleveland, Ohio.").

unprecedented resources in a very short period of time.⁴

While it may seem that our increasingly connected world is getting smaller, the boundary of inequality that separates the world's rich and poor remains very much intact. In fact, the gap between the haves and the have-nots is widening.⁵ And if we admit that there are material differences between the living conditions of the vast majority of the world's poor in the global South and those living in the developed world, then inequalities—not just economic, but also gender, racial, ethnic, and religious—emerge as important determinants of health.

The relationship between poverty, other forms of inequality, and poor health remains whether making comparisons between countries or within a nation. For example, our poorest patients in Boston, Massachusetts do not suffer from malaria or typhoid (as our patients in rural Haiti do), but they are at increased risk for diabetes, cardiovascular disease, obesity, and disability from mental illness or addiction. Certainly, infectious diseases are not equitably distributed. Sub-Saharan Africa holds ten percent of the world's population, but is home to two-thirds of people living with HIV.⁶ In the United States, more than half of new HIV infections are in the black community, which represents only thirteen percent of the population.

4. Writing about SARS, Jerome Singh claims that there are fast and well-funded responses to epidemics threatening affluent countries. Jerome Singh, *SARS, A Challenge from the South*, 423 NATURE 585, 585 (2003); see also Paul Farmer, *SARS and Inequality*, 276 THE NATION 6, 24 (2003).

5. The World Bank estimates that over one billion people live on less than one U.S. dollar per day; 2.7 billion—representing over fifty percent of the world's population at the last calculation in 2001—live on less than two U.S. dollars per day. For more information, see The World Bank Group, *Global Poverty Monitoring*, <http://www.worldbank.org/research/povmonitor/> (last visited Nov. 5, 2004). A World Bank summary paper found that “there was a net decrease in overall incidence of consumption poverty over 1987-98. But it was not enough to reduce the total number of poor by various definitions.” SHAOHUA CHEN & MARTIN RAVALLION, DEV. RESEARCH GROUP, WORLD BANK, *HOW DID THE WORLD'S POOREST FARE IN THE 1990s?* 1 (2000), <http://www.worldbank.org/research/povmonitor/pdfs/methodology.pdf>. The authors “point to two main proximate causes of the disappointing rate of poverty reduction: too little economic growth in the poorest countries and persistent inequalities that inhibited the poor from participating in the growth that did occur.” *Id.* For in-depth analysis of the relationship between poverty, international financial structures, inequality, and health, see JIM Y. KIM ET AL., *DYING FOR GROWTH: GLOBAL INEQUALITY AND THE HEALTH OF THE POOR* (2000); and MEREDITH FORT ET AL., *SICKNESS AND WEALTH: THE CORPORATE ASSAULT ON GLOBAL HEALTH* (2004).

6. Joint United Nations Programme on HIV/AIDS [UNAIDS], *Sub-Saharan Africa*, at <http://www.unaids.org/EN/Geographical+Area/By+Region/sub-saharan+africa.asp> (last visited Nov. 5, 2004).

Gender inequality is also embodied in differential rates of disease: Among young people ages fifteen to forty-nine in sub-Saharan Africa, women are 1.2 times more likely than men to be infected with HIV.⁷ Moreover, the prevalence of HIV in adults in seven southern African countries is now over twenty percent.⁸ African-American and Hispanic women represent less than one-fourth of all women in the United States, but they account for seventy-eight percent of AIDS cases among women.⁹ While tuberculosis has become less prevalent in the United States, it still disproportionately affects the marginalized. One recent study showed that non-Hispanic blacks suffer from tuberculosis at rates eight times greater than non-Hispanic whites.¹⁰

Indeed, the relationship between poverty and disease is perhaps clearest when we consider an airborne infectious disease such as tuberculosis. Conditions of urban poverty—overcrowding, poor housing, and poor nutrition—continue to encourage the spread of tuberculosis worldwide. While treatment of tuberculosis is highly effective, it was the improvement of living conditions—not treatment—that first changed the trajectory of tuberculosis in the developed world. The state-of-the-art system of sanatoria that existed in the United States and early, expensive antibiotics like streptomycin were primarily available to the wealthy.¹¹ In the 1940s, prior to the advent of effective tuberculosis antibiotics, the rate of active tuberculosis in the United States plummeted—due mainly to the post-World War II economic boom and the migration of people from tenements in cities to single-family homes in the suburbs.¹² Tuberculosis then made a striking reappearance in the urban United States in the 1990s, fueled by HIV and structurally associated with overcrowding in prisons, increased rates of homelessness, and the deterioration of public

7. *Id.*

8. *Id.*

9. Ctrs. for Disease Control & Prevention, HIV/AIDS Among U.S. Women: Minority and Young Women at Continuing Risk (May 2002), <http://www.cdc.gov/hiv/pubs/facts/women.pdf>.

10. For information on tuberculosis and race in the United States, see *Racial Disparities in Tuberculosis—Selected Southeastern States, 1991-2002*, 53 MORBIDITY & MORTALITY WEEKLY REP. 556, 556 (2004).

11. For a comprehensive history of tuberculosis in the United States, see FRANK HERBERT, *THE WHITE PLAGUE* (1983). For other works of historical interest, see THOMAS M. DANIEL, *CAPTAIN OF DEATH: THE STORY OF TUBERCULOSIS* (1997); and RANDALL M. PACKARD, *WHITE PLAGUE, BLACK LABOR: TUBERCULOSIS AND THE POLITICAL ECONOMY OF HEALTH AND DISEASE IN SOUTH AFRICA* (1989).

12. See HERBERT, *supra* note 11.

health infrastructure.¹³ This outbreak shows us that public health safeguards—disease surveillance, active case finding, and universally accessible and organized medical treatment—cannot be safely removed without addressing the inequalities that continue to put many at risk.

As physicians who frequently travel between an academic tertiary care hospital in Boston and clinics in the deeply impoverished Central Plateau of rural Haiti, among other places, we can attest that these boundaries of inequality persist. Inequality puts the poor at greater risk for sickness and disease while also leaving poor communities with fewer resources to respond to new and ongoing health challenges. Long after tuberculosis in the United States had been relegated to prisons, homeless shelters, and immigrant populations infected in their home countries, it remains the leading infectious killer of adults in Haiti, less than two hours by plane from Miami. Risk of becoming sick with active tuberculosis is sixty times greater in Haiti than in the United States or Canada.¹⁴ While AIDS has recently overtaken tuberculosis as the leading infectious killer of adults worldwide, the less-visible tuberculosis epidemic is still raging in poor communities around the world, now compounded by the deadly synergy of co-infection with HIV.

INEQUALITY WITHIN POVERTY

Consider the story of one of our patients in Haiti. Joseph Pierre¹⁵ comes from the mountains above Maïssade, four hours on foot from our clinic in the regional capital of Hinche. When we first met Joseph, he had been sick for more than a year with a cough and night-time fever. He had wasted away to just seventy-five pounds, which hung loosely on his 5'8" frame. Our first suspicion was that he suffered from tuberculosis and the advanced stages of HIV disease, but a rapid blood test revealed that he was

13. Karen Brudney & Jay Dobkin, *Resurgent Tuberculosis in New York City: Human Immunodeficiency Virus, Homelessness and the Decline of Tuberculosis Control Programs*, 144 AM. REV. RESPIRATORY DISEASE 745, 747-49 (1991).

14. Tuberculosis incidence in Haiti in 2004 is 296 per 100,000. This is compared to an incidence in the United States of 4.2 and in Canada of 5.8. Haiti's neighbors also harbor less tuberculosis. The Dominican Republic, which shares the island of Hispaniola with Haiti, has a tuberculosis incidence of 87.8 per 100,000, while Cuba, Haiti's neighbor to the west, has an incidence of 9.8—nearer to the rates seen in the developed world. For additional statistics and more information about the global tuberculosis epidemic, see WORLD HEALTH ORG., GLOBAL TUBERCULOSIS CONTROL—SURVEILLANCE, PLANNING, FINANCING (2004), at http://www.who.int/tb/publications/global_report/2004/en/.

15. A fictitious name is used here to protect patient confidentiality.

HIV-negative. A chest x-ray confirmed the presence of tuberculosis, and we started Joseph on anti-tuberculosis antibiotics the same day. He stopped coughing within a week and gained almost ten pounds in his first month of treatment. By the time this Essay is published, he will almost certainly be cured and back to work as a subsistence farmer.

Haiti is, notoriously, the poorest nation in the Western hemisphere. And its population, not surprisingly, suffers some of the worst public health problems in the world.¹⁶ Why, then, is the story of one young man with treatable tuberculosis important? Joseph Pierre's story reminds us that even in settings of universal poverty, gradients of inequality exert their effect on health. Joseph Pierre is poor even by Haiti's standards: He has no family; he lives in a small village far from medical care; he had to rely on the kindness of his neighbors—many of whom are themselves hungry—when he could not tend his own crops. After he became ill, stigma and fear forced him to live in a small shelter on the edge of his community, and he became further isolated.

Joseph came to care through the intervention of a community health worker trained by Zanmi Lasante (ZL), the non-governmental branch of a public-private partnership that is expanding access to HIV, tuberculosis, and basic health care services across Haiti's Central Plateau. Joseph's community health worker was trained to recognize the signs and symptoms of tuberculosis, and he was able to refer this patient to our clinic. Joseph will continue tuberculosis treatment for six months with the help of an *accompagnateur*, a neighbor who will deliver directly observed therapy

16. Life expectancy at birth in Haiti is the lowest in the Western hemisphere at fifty years and falling. WORLD HEALTH ORG., THE WORLD HEALTH REPORT 2004, at 114 (2004), http://www.who.int/whr/2004/en/report04_en.pdf. Less than half the population has access to an improved water source. U.N. DEV. PROGRAMME, UNDP HUMAN DEVELOPMENT REPORT 2004 149 (2004), <http://hdr.undp.org/reports/global/2004>. One-quarter of childbirths are attended by a health professional. *Id.* at 159. Half the population is undernourished. *Id.* at 163. Many areas—including parts of the Central Department, where we work—are suffering from famine-like conditions. Maternal mortality is high, at 680 per 100,000 live births. *Id.* at 171. Per capita GDP has “grown” at a rate of -3.0 percent between 1990 and 2002. *Id.* at 187. HIV prevalence is estimated to be as high as 6.1 percent, the highest in the Western Hemisphere. In Haiti, AIDS kills an estimated 24,000 people annually, has orphaned 200,000 children cumulatively through 2002, and has also aggravated an already severe tuberculosis epidemic. *See, e.g.*, UNAIDS, EPIDEMIOLOGICAL FACT SHEETS ON HIV/AIDS AND SEXUALLY TRANSMITTED INFECTIONS: HAITI 2 (2004), <http://www.unaids.org/en/geographical+area/by+country/haiti.asp>; UNAIDS, EPIDEMIOLOGICAL FACT SHEETS ON HIV/AIDS AND SEXUALLY TRANSMITTED INFECTIONS: HAITI 2 (2002), http://www.who.int/emc-hiv/fact_sheets/pdfs/haiti_en.pdf.

(DOT) and accompany Joseph through his illness. In 1988, when ZL began systematic treatment of tuberculosis in central Haiti, cure rates were unacceptably low. In response, ZL developed a community-based DOT system, and deaths from tuberculosis nearly disappeared.¹⁷ This experience, and the human infrastructure constructed in response to the tuberculosis epidemic in rural Haiti, has become the backbone of our expansion of HIV treatment and prevention services.¹⁸

Many in public health circles point to a lack of infrastructure as one reason to limit investment in and expansion of treatment, in favor of disease prevention efforts. We have written in other settings that construction of a system of community health workers and *accompagnateurs* may be the answer to this missing infrastructure.¹⁹ Our experience with tuberculosis treatment in rural Haiti shows that this method is clinically successful, and preliminary results from our AIDS treatment efforts—DOT with antiretroviral medicines—suggest the same. Furthermore, in rural Haiti—as in most communities with high burdens of disease, poverty, and unemployment—a rich surplus of human resources remain untapped.²⁰ In Haiti, we have many more applicants to become community health workers than we have jobs available. In our experience, only basic literacy is needed to provide effective DOT, and literacy training is part of our comprehensive approach to community-based health care.²¹ Full

17. An early study by ZL suggested treatment relying only on monthly clinic visits—even when consultation and treatment were provided free of charge—led to an unacceptable level of treatment failures. When a comprehensive approach to tuberculosis control was initiated, including DOT, modest financial and nutritional support, and active case-finding through contact screening, tuberculosis rates and deaths from tuberculosis fell. Patients now at greatest risk for developing tuberculosis are those that live outside our catchment area, which is covered by community health workers and *accompagnateurs*. Paul Farmer et al., *Tuberculosis, Poverty and "Compliance": Lessons from Rural Haiti*, 6 SEMINARS IN RESPIRATORY INFECTIONS 254, 255-56 (1991).

18. Paul Farmer et al., *Community-Based Approaches to HIV Treatment in Resource-Poor Settings*, 358 THE LANCET 404, 404 (2001); see also Joia S. Mukherjee et al., *Tackling HIV in Resource Poor Countries*, 237 BRIT. MED. J. 1104 (2003).

19. Heidi L. Behforouz et al., *From Directly Observed Therapy to Accompagnateurs: Enhancing AIDS Treatment Outcomes in Haiti and in Boston*, 38 CLINICAL INFECTIOUS DISEASES S429, S430-31 (2004).

20. Precise employment statistics are not available for rural Haiti, but wage-earning employment is rare. The majority of the population practices subsistence farming. While the stipends we provide to community health workers and *accompagnateurs* do not replace the need for continued subsistence agriculture, they are one of the few available sources of income. With a modest budget, we have become the largest employer in the region.

21. ZL operates both primary and secondary schools that serve over a thousand students

community participation also assures that priorities are set by the local population rather than from the boardrooms of Geneva or Washington, D.C.

ZL's community-based structure was recently tested by the coup d'état which removed Haiti's popularly elected president, Jean-Bertrand Aristide, on February 29, 2004.²² During the months leading up to the coup, much of the nation's already-fragile public health system was further disrupted. The medical staff at the University of Haiti's General Hospital in Port-au-Prince—the only national referral hospital—was on strike, protesting the lack of security. Their doors were closed for weeks during the coup and its aftermath. Many private and public clinics also closed during the months of February, March, and April 2004. But despite working in the Central Plateau—an area of concentrated rebel activity near Haiti's border with the Dominican Republic—and despite ZL's collaboration with the Aristide administration (which was unprecedentedly supportive of nationwide efforts to expand HIV treatment and prevention), all six of ZL's clinics remained open during the unrest. From our staff of over one thousand employees, only six—the entire non-Haitian staff, including the authors—were displaced temporarily by the coup. ZL provides directly observed therapy to over 1100 AIDS patients and 1500 patients with active tuberculosis. Almost none of these patients missed a single dose of their medications, which is important not only for the health of individual patients but also in the prevention of drug resistance.²³ This unwelcome test of ZL's community-based health care system in central Haiti shows that a decentralized, community-based infrastructure can survive even the most disruptive social and political upheaval.

During the most dangerous moments of the coup, we saw a decrease in the number of ambulatory visits at our clinics as villagers stayed home out of fear, all too familiar from the intimidation and violence inflicted by the military and various militias throughout Haiti's dangerous history. The average number of patient visits at four of our expansion clinics dropped

and twenty-five adult literacy centers in partnership with the Haitian government's nationwide literacy efforts.

22. For more on the February 2004 coup and subsequent politics, see NOAM CHOMSKY ET AL., *GETTING HAITI RIGHT THIS TIME* (2004); Paul Farmer, *Who Removed Aristide?*, 26 LONDON REV. BOOKS 28, 28 (2004); Paul Farmer, *Political Violence and Public Health in Haiti*, 350 NEW ENG. J. MED. 1483 (2004); and Peter Hallward, *Option Zero in Haiti*, 27 NEW LEFT REV. 23 (2004).

23. See Sally Blower et al., *Predicting the Impact of Antiretrovirals in Resource-Poor Settings: Preventing HIV Infections Whilst Controlling Drug Resistance*, 3 CURRENT DRUG TARGETS—INFECTIOUS DISORDERS 345, 351 (2003).

from 189 per day in January 2004 to 113 per day in March 2004.²⁴ The number of patients seen by ZL has since exceeded pre-coup levels. The real repercussions of this widespread fear and decreased use of the public health system will never be known, but they almost certainly account for the most damaging health consequences of the coup. Nevertheless, the fact that a grassroots health care movement functioned under such extreme circumstances gives weight to our belief that community participation is essential in the struggle to provide equitable health care to the world's poor. This unique stability adds to the benefits already mentioned—namely, local priority setting, strengthening of underutilized community resources, expansion of literacy and educational opportunities, and broad-based active case-finding of the kind that brought Joseph Pierre to our attention and care.

CONCLUSION

Inequality will persist as long as there is poverty. And inequality will continue to drive the vulnerable toward increased sickness and early death—whether in the prisons and inner-cities of the developed countries or among the billions living in poverty worldwide. Public health decisions that ignore this reality are doomed to fail. In our opinion, building a community-based, stable public health structure is a first step toward better health, alleviating poverty, and closing the gap of inequality that keeps poor communities uniquely vulnerable to suffering and disease. One recent meta-analysis showed that direct costs related to a death from HIV disease, in various settings, often amount to more than fifty percent of a household's annual income—and are sometimes greater than one hundred percent—even before accounting for the extensive indirect costs of lost labor, decreased agricultural production, and general social and economic instability.²⁵ For most of our patients in central Haiti, a return to health means they are better able to provide for themselves and their families and to participate in social and political change in their troubled nation. This is a small step in a setting of grinding poverty, but it is a step in the right direction.

24. Averages were calculated from daily reports from ZL clinics in Boucan Carré, Thomonde, Lascachobas, and Belladère (on file with authors).

25. Steven Russell, *The Economic Burden of Illness for Households in Developing Countries: A Review of Studies Focusing on Malaria, Tuberculosis, and Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome*, 71 AM. J. TROPICAL MED. & HYGIENE 147, 148 (2004). For more on the economic and social burden of HIV, see *Myth Eight: Limited Resources*, in GLOBAL AIDS: MYTHS AND FACTS 135-51 (Alexander Irwin et al. eds., 2003).

A “basic minimum package” to support a complex intervention—in our case, the provision of antiretrovirals under DOT—was identified with input from ZL’s *accompagnateurs* and the larger community. This basic package is organized around four pillars: HIV prevention and care; tuberculosis diagnosis, treatment, and active contact screening; STD case-finding and treatment; and comprehensive women’s health services. Relating our experience in the town of Lascahobas, one of ZL’s five expansion sites in Haiti’s Central Department, we showed numerous health improvements, some even beyond the four pillars framework. These included a rise in tuberculosis diagnosis and treatment from nine to over two hundred patients within fourteen months; a ten-fold rise in the number of patient visits per day; increased access to prenatal care, including HIV testing to help prevent mother-to-child transmission; increased access to vaccines; improved staff morale; and greater community participation in health education activities.²⁶ Our experience has shown that the expansion of complex health interventions, such as providing comprehensive HIV prevention and care, not only enhances the life of each patient, but also increases public health and primary care capacity. Integrated, community-based care is essential if equity in health and health care is to become a reality for the world’s poor.

26. David A. Walton et al., *Integrated HIV Prevention and Care Strengthens Primary Health Care: Lessons from Rural Haiti*, 25 J. PUB. HEALTH POL’Y 137, 152 (2004).

Medicare Reform and Social Insurance: The Clashes of 2003 and Their Potential Fallout

Theodore R. Marmor, Ph.D.* and Jacob S. Hacker, Ph.D.†

Medicare pays for at least half of the hospital and medical expenses incurred by America's elderly and disabled.¹ It is also periodically the object of intense political debate, marked by exaggerated claims about how the sky will fall unless some fundamental change is made in the financing, benefits, or administration of the program. Over the past decade and a half, this political attention has had less and less to do with legitimate concerns about budget deficits and Medicare's real (if usually overstated) faults. Instead, it has become principally fueled by the alarmist rhetoric of those who ideologically oppose Medicare's social insurance structure. Most of these critics, mindful of Medicare's broad popularity, mask their underlying hostility to the program with a veneer of public-minded concern. Unfortunately, their rhetoric of crisis clouds more than it illuminates what is fundamentally at issue in all these disputes.

In this Essay, we concentrate on one such confused aspect of the recent debate that is emblematic of the contemporary politics of Medicare: the debate over "means-testing" Part B of Medicare, the medical insurance program. In a significant break with Medicare's history, the reform legislation of 2003—the Medicare Modernization Act (MMA)—imposes sharply higher premiums on wealthy beneficiaries.² The story of how this

* Professor of Public Policy and Management and Professor of Political Science, Yale University.

† Peter Strauss Family Assistant Professor of Political Science, Yale University. This Essay draws upon an article jointly published in December 2004 for the Gerontological Society of America's special public policy report on Medicare reform legislation. Jacob Hacker & Theodore R. Marmor, *Medicare Reform: Fact, Fiction and Foolishness*, 13 PUB. POL'Y & AGING REP. 1 (2003).

1. CRAIG CAPLAN, AARP PUB. POL'Y INST., WHAT SHARE OF BENEFICIARIES' TOTAL HEALTH CARE COSTS DOES MEDICARE PAY? 1 (2002), http://research.aarp.org/health/dd78_costs.pdf.

2. Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Pub. L. No. 108-173, § 811, 117 Stat. at 2364-65. Income scaling more than tripled the monthly premium for participants with incomes over \$80,000 (single) or \$160,000 (couples). Julie Appleby, *The Ins and Outs of Medicare Changes*, USA TODAY, Nov. 25, 2003, at 3B ("Instead of

came to pass has been largely lost in the crowded pages of American journalism. But at least one journalist noted that “House and Senate negotiators, struggling for accord on a plan to redesign Medicare, have agreed in principle that wealthy older Americans should pay more for doctor visits and other outpatient care, reprising an idea that has proved politically explosive.”³ The income or means-testing dispute was just one small part of the larger struggle over the shape of the prescription drug benefit that President George W. Bush signed into law in early December 2003.⁴ Yet the dispute was symbolic of the confusion that surrounds the questions of whether and how Medicare ought to be restructured.

THE ORIGINS OF MEDICARE AND THE 2003 REFORM LEGISLATION

The historical context of Medicare’s overarching structure is worth considering—it enables one to evaluate subsequent reforms, and the

paying the current 25% of the Medicare Part B premium, . . . wealthier seniors would pay up to 80% of that premium”). Together, these beneficiaries comprise only a small fraction of the current Medicare population. See Ctr. for Medicare Advocacy, *Quick Reference Medicare Facts & Statistics* (Nov. 8, 2004), at http://www.medicareadvocacy.org/FAQ_QuickStats.htm (reporting that, in 1999, ten percent of Medicare beneficiaries had incomes greater than \$80,000).

3. Amy Goldstein, *‘Means Test’ Deal Near on Medicare; Wealthy Seniors Would Be Charged More Under Plan*, WASH. POST, Oct. 16, 2003, at A1. The fact that this provision emerged from the conference committee of the House and Senate requires additional comment. The leader of the House conferees, Republican Congressman Bill Thomas from California, managed to exclude all but two of the designated Senate conferees, and the House conferees from the Democratic caucus refused to participate in such an imbalanced and, in their view, illegitimate practice. The counter-factual—what would have emerged from an ordinary conference bargaining process—is necessarily speculative. But the process employed was irrefutably controversial and contrary to long-established congressional norms. Moreover, as became known later, the passage of this legislation was marked by a number of apparently scandalous actions, ranging from alleged bribes to House Republicans initially opposed to the legislation to widely believed claims that the former head of Medicare, Tom Scully, threatened to fire the independent actuary if he sent Congress his cost estimates, which were much higher than those of the Congressional Budget Office. See Ted Marmor, *The US Medicare Programme in Political Flux*, 10 BRIT. J. HEALTH CARE MGMT. 140, 142 (2004); Timothy Noah, *A Drug-Company Bribe?: The Medicare Vote Scandal, Continued*, SLATE, Dec. 8, 2003, at <http://slate.msn.com/id/2092242>; Timothy Noah, *Information Is Treason: Why Bush Is Worse Than Reagan*, SLATE, Mar. 16, 2004, at <http://slate.msn.com/id/2097268>.

4. Ctrs. for Medicare & Medicaid Servs., *Medicare Modernization Act*, at <http://www.cms.hhs.gov/medicarereform/> (last modified Nov. 5, 2004).

agendas they reflect, in light of the aims of the program's framers. The development of hospital insurance, Medicare's Part A, was the focus of attention from the beginning of the Kennedy Administration in 1961 and became the legislative aim of President Johnson in the mid-1960s.⁵ As such, it incorporated the traditional elements of American social insurance programs—compulsory taxes known as "FICA" (for "Federal Insurance Contributions Act") contributions, a ceiling on the wage and salary income on which those taxes were paid, and broad eligibility without restrictions based on means or assets.⁶ Part B, or supplemental medical insurance, pays for physicians' fees and a variety of other outpatient expenses.⁷ This feature, unexpectedly included with Republican backing, introduced premiums—rather than payroll contributions—as a source of financing.⁸ Part B was enacted as a voluntary insurance program, though with subsidies so substantial that the overwhelming proportion—some ninety-six percent of those eligible—have enrolled.⁹

In 1965, an overwhelmingly Democratic Congress secured enactment of Medicare.¹⁰ In 2003, the concerted push to legislate a prescription drug benefit for Medicare arose because of the *absence* of clear partisan control of either the Senate or the House: For a decade or more, each political party had fought to make sure the other could not take credit for introducing such an expansion of insurance coverage, with stalemate regularly the result. In 2003, however, Republican and Democratic leaders in Congress and in the Administration came to believe that continued stalemate might well provide the other side with an effective electoral

5. See generally THEODORE R. MARMOR, *THE POLITICS OF MEDICARE* 45-85 (2d ed. 2000) (describing the origins and subsequent fate of Medicare).

6. See MARMOR, *supra* note 5, at 15-16 (describing the early appeal of financing Medicare hospital insurance through Social Security contributions); Ctr. for Medicare Advocacy, Medicare Summary (Dec. 3, 2004), at http://www.medicareadvocacy.org/FAQ_MedicareSummary.htm (providing an overview of Part A's financing, eligibility, and benefits). But see Cathy A. Cowan et al., DataView, *Business, Households, and Government: Health Spending, 1994*, HEALTH CARE FINANCING REV., Summer 1996, at 157, 161 (noting that "the cap on the maximum amount of taxable earnings" for the Medicare Hospital Insurance Trust Fund was eliminated in 1993).

7. See Ctr. For Medicare Advocacy, The Medicare Part B Benefit (Sept. 22, 2004), at http://www.medicareadvocacy.org/FAQ_PartB.htm#Payment.

8. See MARMOR, *supra* note 5, at 45-61.

9. See David J. Gross & Normandy Brangan, AARP Pub. Pol'y Inst., The Medicare Program (Apr. 1998), at http://research.aarp.org/health/fs45r_medicare.html.

10. See Theodore R. Marmor & Jerry L. Mashaw, Commentary, *Prescription Drug Bill Makes Only Politicians Feel Better*, L.A. TIMES, June 17, 2003, at B15.

battle weapon for 2004.¹¹ Opposing the other's reform, without offering a feasible alternative, appeared electorally dangerous. Both parties consequently were prepared to sacrifice crucial features of their traditional policy aspirations. Democrats, who otherwise might have insisted on a generous drug benefit for all beneficiaries, agreed to a plan that largely failed to satisfy this goal. For Republicans, passing any drug benefit represented a strategic compromise; they ultimately agreed to expand an entitlement program that they had long criticized.¹²

The decisions of the two parties to promote legislative change altered the calculations of pharmaceutical industry strategists. The industry for years had opposed adding drug coverage to Medicare. However, once the enactment of a drug benefit seemed likely, as it did in 2003, the industry threw its support behind the Republican version—recognizing that this would be preferable to a drug benefit that might be passed in the future by a potentially Democratic-controlled Congress.¹³

The resulting legislation purports to expand Medicare by offering a drug benefit, yet it includes an array of provisions that clearly constrain and even obstruct the Medicare program. The first portion of the MMA provides a much-needed, if modest and complex, drug benefit that will allow Medicare beneficiaries to buy government-guaranteed—although, in most cases, privately provided—drug plans.¹⁴ While this new benefit is generous for some low-income seniors, it appears likely to raise out-of-

11. See Robert Pear, *Tentative Medicare Pact Offers Drug Benefit to Elderly*, N.Y. TIMES, Nov. 13, 2003, at A28.

12. Proponents of charging the affluent elderly more seem to have forgotten the politics of catastrophic coverage reform in 1987-1988. Then, as now, reformers argued that it was commonsensical to charge the affluent elderly more. Then, unlike now, there was much to be said for the real improvement in Medicare that catastrophic coverage would have brought for all Medicare beneficiaries. But, within a year of passage, Congress "took the extraordinary step of repealing the law." Goldstein, *supra* note 3. Such "[e]fforts to charge comparatively wealthy Medicare [beneficiaries] more for their care have a long, divisive history." *Id.*

13. See Ctr. for Am. Progress, *Medicare Bill Greased by Corporate Dollars* (June 23, 2004), at <http://www.americanprogress.org/site/pp.asp?c=biJRJ8OVF&b=98500>.

14. Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Pub. L. No. 108-173, § 101, 117 Stat. 2066, 2071. Although the comprehensive drug benefit does not take effect until 2006, a purchasable drug discount card and limited assistance for low-income seniors were to be offered in the interim. Pub. L. No. 108-173, § 101(a)(2), 117 Stat. at 2072 (initiation date of Jan. 1, 2006); *id.* at 2131-48 (discount card and transitional assistance programs); see also Andrea Stone, *Benefits Start in '06, but Help Available Sooner*, USA TODAY, Nov. 26, 2003 at A3.

pocket drug costs for some other poor beneficiaries, namely several million low-income seniors who will lose the generous coverage they now enjoy under state Medicaid programs.¹⁵ Further, because the initiative is poorly designed for controlling drug costs—it does not allow Medicare to use its massive buying power to demand price reductions¹⁶—the plan is likely to ultimately leave many seniors little better off than they are today.

The remainder of the MMA consists of provisions that have little or nothing to do with drug coverage, but seem consistent with the demands of interest groups and aligned with a basic ideological hostility toward Medicare. In addition to sparing drug companies their greatest fear (i.e., Medicare's utilization of its monopsony power), the MMA contains other elements that risk further degeneration of Medicare's all-in-the-same-boat structure. To begin, the legislation provides for substantial new subsidies for private insurers—thereby favoring those who use private health insurance plans.¹⁷ The bill also introduces a new standard for program “insolvency” that could force substantial shifts of expense from Medicare to seniors.¹⁸ Finally, and central to our discussion, the MMA uses what is essentially a ‘means-test’ to determine premiums for Part B premiums.¹⁹ This reform threatens the basic principle of social insurance that holds that having large pools, with common benefits and regulations, is crucial. It seems to represent a stealth effort to transform the fundamental structure of Medicare in the long-term.

The rhetorical appeal of means-testing is obvious and may explain why so many political pundits came to accept charging higher premiums to upper-income elderly as common sense: Why, many asked, should we have a flat premium when some of the elderly are so rich?²⁰ Why shouldn't we

15. Ctr. for Medicare Advocacy, Will the Medicare Act of 2003 Really Do That? Myths and Realities About the New Law (Apr. 1, 2004), at http://www.medicareadvocacy.org/reform_Actof2003_WillItReallyDoThat.htm. In addition, despite the bill's subsidies for employers who retain coverage, some employers will likely drop retiree drug coverage in response to the MMA; some seniors who presently have good coverage under such plans may thus be made worse off.

16. Pub. L. No. 108-173, § 101(a)(2), 117 Stat. at 2098 (2003).

17. *Id.* § 222, 117 Stat. at 2913.

18. In a provision that has received relatively little attention and was not in either the original House or Senate legislation, the bill creates a new standard for Medicare “insolvency.” It defines the program as insolvent whenever, in two consecutive years, general revenues finance more than forty-five percent of Medicare's Part B costs. *Id.* §§ 801-04, 117 Stat. at 2357.

19. *Id.* § 811, 117 Stat. at 2364-65.

20. This idea appears to unite those New Democrats who rail against “corporate

link Medicare benefits to ability to pay? However, as we shall argue, the idea of means-testing Medicare is fiscally misleading, programmatically threatening, and—if extended as its advocates desire—philosophically at odds with the very principles that have made Medicare such a popular, relatively stable, and successful program.

THE IMMEDIATE FISCAL IMPACT

The fiscal fraudulence that lies behind the means-testing in the MMA is only apparent if one understands how Medicare is financed. Medicare Part B—or supplemental medical insurance—pays for physicians' fees and a variety of other outpatient expenses. As passed in 1965, Part B is a voluntary program that is, as noted, substantially subsidized by the government. Each Medicare beneficiary pays the same individual premium,²¹ with general taxes covering the remainder of the costs. The original idea was that premiums would finance half of Part B's outlays, and general taxes would pay for the other half. Over time, the ratio has shifted so that currently one-fourth is covered by premiums, and three-fourths are covered by general taxes.²² As a result, Part B is financed largely through the federal income tax, which is a progressive tax on all Americans, including the upper-income elderly. Similarly, Medicare's Part A—hospital insurance—is financed by a small proportional tax on taxable wage and salary income.²³

welfare," fiscal conservatives worried about future deficits, and a number of Republicans who are usually staunch defenders of the well-to-do. (The latter two groups do not spend equal time lamenting the Bush-era tax cuts, which are undoubtedly more consequential to the nation's fiscal future).

21. The Medicare Part B premium for 2004 was \$66.60 per month. *Medicare Buy-In Part II: Using Examples To Illustrate Key Eligibility Issues*, 3 BENEFITS PLANNER 82, 84 (2003), http://www.ilr.cornell.edu/edi/publications/BenefitsPlanners/BenefitsPlanner_Fall2003.pdf.

22. See Jill Berstein, *Should Higher Income Beneficiaries Pay More for Medicare?*, NAT'L ACAD. SOC. INS. MEDICARE BRIEF 3 (May 1999) ("When Medicare was first created, the Part B Premium was designed to cover about half the Part B program costs. As these costs increased faster than inflation, Congress chose to limit the increases charged to beneficiaries to the Social Security cost-of-living increases"), http://www.nasi.org/usr_doc/medicare_brief_2.pdf; see also Robert Pear, *Medicare Premium To Increase By 13.5 Percent Next Year*, N.Y. TIMES, Oct. 16, 2003, at A22 (noting that "The basic Medicare premium" is statutorily "set at the level needed to cover about 25 percent of the cost of Part B").

23. It is a 2.9% payroll tax split evenly between employer and employee. CRAIG CAPLAN & RYAN COOL, AARP PUB. POL'Y INST., THE STATUS OF MEDICARE PART A AND PART B TRUST

For this reason, by the time higher-income Americans reach the age of sixty-five, they have generally paid far more into the program than would have been required for private health insurance and far more than lower-income Americans. In other words, viewing the financing of Medicare over different time periods shifts the resulting portrait of its distributive features. By ignoring the realities of Medicare's financing over the course of the lifespan, advocates of "means-testing" present a misleading image of who contributes what to the program.²⁴ Evaluating social insurance programs properly requires not a financial snapshot at one point in time, but a view of who pays and who receives what over time. We will return to the question of why this fact is not adequately weighed in the public discourse on Medicare and other social insurance programs.

The means-testing idea is fiscally misleading in other respects as well. The revenues raised by such proposals would—from the standpoint of Medicare's overall fiscal viability—be trivial.²⁵ Because the premiums for Part B pay for only a quarter of program costs and because most Medicare beneficiaries have modest incomes,²⁶ targeting the richest of those who pay the premium makes little difference for Medicare's financial future. Many, like Henry Aaron of the Brookings Institution, who do not reject means-testing out of hand, still agree that "the number of well-to-do elderly is too small" to make a big difference in Medicare's fiscal future.²⁷

To be sure, the revenues raised by such income-scaled premiums would scarcely be trivial in absolute dollar terms. One to two percent of Medicare's outlays over ten years could easily amount to twenty billion dollars.²⁸ The relevant fiscal question, however, is not whether twenty

FUNDS: THE TRUSTEE'S 2004 ANNUAL REPORT 1 (2004), http://research.aarp.org/health/dd96_medicare.html.

24. See, e.g., Editorial, *Medicare Robbery*, WASH. POST, July 17, 2002, at A20.

25. Henry J. Aaron, Op-Ed, *The Grand Delusion*, CENTURY ISSUES, Oct. 15, 2003, <http://www.brook.edu/views/op-ed/aaron/20031015.htm>; see also Berstein, *supra* note 22, at 8 tbl.2, http://www.nasi.org/usr_doc/medicare_brief_2.pdf (describing two possible income-scaling plans and projecting that neither would raise revenues by more than two percent of expected Medicare outlays over a five-year period).

26. As of 1995, more than three out of four Medicare beneficiaries had annual incomes below \$25,000, and only three percent of Medicare spending went to recipients with incomes over \$50,000. The Commonwealth Fund, *Medicare Turns Thirty*, at http://www.cmwf.org/publications/publications_show.htm?doc_id=221620 (last visited Dec. 5, 2004) (excerpts from Senate Finance Committee testimony of Karen Davis, Commonwealth Fund President on February 28, 1995).

27. Aaron, *supra* note 25.

28. The late Senator Everett Dirksen is often reported to have said, "[A] billion here

billion dollars might improve Medicare's fiscal circumstances. It certainly could, if only modestly. Rather, the important evaluative question is whether the revenues raised are worth their price in terms of administrative hassle, bad social insurance precedent, and any consequent undermining of Medicare's political support. A glance at the expected effects of means-testing suggests that all except those ideologically opposed to social insurance would answer these questions in the negative.²⁹ Even Robert Reischauer, a defender of means-testing who argues that "making affluent beneficiaries pay more than those with fewer resources is eminently sensible," concedes that it is "not the long-term solution" to Medicare's solvency.³⁰

THE LONGER-TERM POLITICAL CONSEQUENCES

Certainly when one considers the long-term ideological and political ramifications of means-testing, the 2003 reform is unlikely to help sustain Medicare. When Medicare was created, it was deliberately designed to encompass both rich and poor, sick and well among its senior citizen beneficiaries. This universalistic impulse remains clear in the Part A hospital program, which is mandatory and financed by proportional contributions during one's working life.³¹

The use of proportional contributions³² or progressive contributions³³

and a billion there, and pretty soon you're talking real money." See The Dirksen Congressional Ctr., "A billion here, a billion there . . .", at http://www.dirksencenter.org/print_emd_billionhere.htm (last visited Nov. 23, 2004) (noting that Dirksen would have approved of the quotation's sentiment, although he never actually made the attributed statement).

29. One cannot rule out the possibility that some who would disagree here are simply uninformed about the principles of social insurance, rather than opposed to them. Teaching about social policy for more than thirty years suggests that this might well be the case for many college-educated persons under the age of fifty. Although for anyone educated in the social sciences between 1900 and 1960, there was a high probability that sociology, economics, and political science courses would comment on social insurance, its differences from private insurance, and the significance of social insurance in the American public household, coverage of these topics in the classroom has declined sharply since World War II. See THEODORE R. MARMOR ET AL., *AMERICA'S MISUNDERSTOOD WELFARE STATE: PERSISTENT MYTHS, CONTINUING REALITIES* (1992).

30. Robert D. Reischauer, Commentary, *This Isn't Such a Bitter Pill*, L.A. TIMES, Oct. 19, 2003, at M5.

31. JULIE LEE ET AL., *THE DISTRIBUTIONAL EFFECTS OF MEDICARE 3* (Nat'l Bureau of Econ. Research, Working Paper No. 6910, 1999), <http://www.nber.org/papers/w6910.pdf>.

32. Proportional contributions are also required for American social security pensions.

is workable for mandatory social insurance programs. By their very nature, such programs (unlike commercial insurers) do not take into account the specific characteristics of the individual or evidence about individual risks and circumstances.³⁴ For risks that all of us face—like disability, job accidents, unemployment, retirement, and medical expenses—social insurance provides income protection which reflects policy decisions, not the risk selection and underwriting that characterize private commercial insurance.

While particular distributive models vary, social insurance programs are generally premised on mandatory contributions. The important assumption underlying this is that the political stability and economic security of such programs depend upon the broad acceptance of the legitimacy of the programs themselves. In the history of the welfare state, social insurance emerged as an alternative to private and public charity—the hated poor house and the benevolent squire distributing alms at holiday time. A sense of entitlement to a benefit was widely presumed to flow from contributing to the common fund. Hence, what are otherwise compulsory taxes become, in the language of social insurance, “contributions.”³⁵

This set of considerations, however, does not apply to voluntary plans. The introduction of steep income-related premiums will likely prompt those with high incomes, good health, and catastrophic health insurance options to consider not paying the new, higher Part B premium. Faced with stiff new premium hikes, healthy and wealthy senior citizens would have good reason to opt out of Part B.³⁶ This, in turn, could very well

33. Progressive contributions are employed by Western European sickness funds.

34. In commercial insurance, premiums reflect the expected costs of individuals or groups. For example, residents of high-crime areas pay substantially higher theft insurance premiums than those in low-crime areas. In social insurance, the aim is to protect against the risk, but not to concentrate higher costs on those who happen to incur the risk more frequently.

35. See ROBERT BALL, *SOCIAL SECURITY: TODAY AND TOMORROW* (1978); THEODORE R. MARMOR ET AL., *supra* note 29, at 1-53.

36. In 1999, one source reported that ten percent of Medicare beneficiaries generate sixty-percent of the program's costs, while half of the program's beneficiaries “account for only 1.6% of the expenses.” Harold C. Sox, *Defined Contribution Programs and Their Effects on Medicare*, ACP-ASIM OBSERVER, Feb. 1999, <http://www.acponline.org/journals/news/feb99/defined.htm>. The exit motive would be particularly salient if congressional conservatives were to enact large new tax breaks for IRA-like medical savings accounts, which are favored by private insurers. The MMA already provides for a type of health savings account that allows “individuals or families to establish a tax free fund for the entire

undermine the diversified risk pool and widespread popular support that has sustained Medicare since its inception.³⁷ This incentive structure potentially initiates a vicious cycle; a reduction in the overall health of the program's population produces higher premiums over time, which, in turn, could trigger further departures. American insurers would no doubt deliver high-deductible plans for the healthy and wealthy, plans that protect against devastating illness costs, but at relatively low monthly premiums. Therein lies the greatest threat both to Medicare's programmatic design and to its long-term political stability—a breaking up of the Medicare risk pool.³⁸

So what, the skeptic might ask, given that the premiums paid by two percent of the elderly are a trivial part of the financing of Part B? The answer, of course, is political:³⁹ Over time, this dynamic could seriously compromise Medicare, especially if those who leave the program lose

amount of their health insurance deductible over \$1,000 (or \$2,000 for families)” up to the lesser of \$2250 individually or \$4500 per family. Melissa Ganz, iBrief, *The Medicare Prescription Drug, Improvement, and Modernization Act of 2003: Are We Playing the Lottery with Healthcare Reform?*, 2004 DUKE L. & TECH. REV. 11, ¶ 19 & nn.58-60, at <http://www.law.duke.edu/journals/dltr/articles/2004dltr0011.html>; see Pub. L. No. 108-173, § 1201, 117 Stat. at 2469-79 (2003) (establishing and describing health savings accounts).

37. Medicare's Part A, the hospital insurance program, clearly reflects social insurance principles. Payments are compulsory for wage earners during their working life, and there is no connection between the proportional taxation and what is covered or what is paid during retirement. All providers are paid according to the same rules, and there are no wedges between beneficiaries in connection with current income. This inclusiveness greatly increases the attention to Medicare in congressional tussles about its future. AARP, for example, closely monitors the program's politics, figuring rightly that its millions of members care a lot. This would be less true for a divided Part B program.

38. Even experienced Democratic social policy strategists who arguably should have recognized this risk, including former Social Security Commissioner Robert Ball and budget specialist Robert Greenstein, saw no fundamental problem in, for example, tripling the premiums wealthy beneficiaries would pay from about \$700 per year to over \$2100 per year. Robert Pear, *Medicare Plan Raises the Cost for the Affluent*, N.Y. TIMES, Oct. 6, 2003, at A1.

39. Anyone who has observed the fate of Medicaid in the decades since its enactment (with Medicare) in 1965 will know the differences in experience. Medicaid has had a boom and bust cycle and is poorly protected when state revenues are threatened by economic downturns. In part, this precarious situation is the consequence of state constitutional prohibitions against deficit financing. But another part of the explanation is that Medicare's supporters are more numerous, more powerful, and more obvious. Why, we ask, should supporters tinker with the program in a way that threatens the source of its political stability? While there has not been paralysis in Medicare policymaking, reformers have faced organized, committed backers when promoting change.

interest in supporting the program electorally or even choose to advocate for the increased support of private insurance alternatives.⁴⁰ Moreover, an income-related premium would require the creation of new administrative machinery for distinguishing among beneficiaries on the basis of current income in order to charge differential premiums. Doing so would use of some of the modest savings that the higher premiums themselves promise. More importantly, once this program feature is created, it would provide the ideological basis and administrative means for further distinctions in the future. Once the richest two percent were charged a premium surcharge, for example, the advocates of means-testing could, and almost certainly would, call for lowering the income level at which the surcharge applies—making more and more seniors the targets of private options.⁴¹

For proponents of social insurance, important principles are at stake in the means-testing provision of the recently passed legislation. Because of fundamental concerns about maintaining a broad risk pool, social insurance scholars have long rejected means-testing when it refers to limits on eligibility based on wealth or income. Although the current legislation does not go so far as to place wealth or income limits on eligibility, it shares important philosophical roots with the critics of social insurance. The new program of income-conditioned premiums, at least rhetorically, sets the stage for more substantial means-testing in the future. No matter how well cloaked they are in the language of egalitarianism, populist hostility to the rich, or the rhetoric of necessary reforms, these are serious threats to the future of social insurance.

CONCLUSION: THE PRESCRIPTION DRUG BILL

That the reforms of 2003 were the result of political bargaining is not surprising. Politics frequently requires, and results in, compromises. What is startling about the 2003 legislation is just how deeply the compromises—

40. On CNN's *Inside Politics* program, William Schneider suggested a likely reason for Republican interest in changing Medicare's financing: "Democrats suspect that by imposing a means test, Republicans are not trying to save Medicare. They're trying to destroy it. How? By turning Medicare into welfare . . ." *Judy Woodruff's Inside Politics: Is it Payback Time for Iraq? Interview with John Edwards* (CNN television broadcast, Oct. 16, 2003), <http://transcripts.cnn.com/TRANSCRIPTS/031016/ip.00.html>.

41. It is naïve to believe, however, that such plans, if enacted, would remain limited to only the very high income elderly. The typical policy pattern is not to index the threshold income levels to inflation. As a result, more and more elderly will likely be affected by this change in policy over time. And therein lies a central political issue for the future of Medicare.

or more accurately, the concessions to ideology and private interests—undercut the stated goals of the law, namely drug coverage for seniors.

The MMA, as written, will yield a drug benefit program rife with inefficiencies that will likely benefit private interests at the public's expense.⁴² This is a consequence, in part, of the MMA's subsidies for health savings accounts⁴³ and private health plans,⁴⁴ which have markedly higher overhead costs than the public Medicare program.⁴⁵ Ultimately, the MMA's drug benefit is convoluted and rather meager—covering only a limited share of seniors' expected overall drug spending.⁴⁶ Credible estimates

42. A more sensibly designed bill could yield far greater coverage—perhaps twice as much—with the expected increases in Congressional and personal Medicare spending over the next decade. In 2004, the *Washington Post* projected that the total ten-year cost of the drug benefit would be \$564 billion. Ceci Connolly, *Premiums To Rise by 17.5%; Percentage Increase Biggest in 15 Years*, WASH. POST, Sept. 4, 2004, at A1. An oft-cited Congressional Budget Office projection estimated that the prescription drug benefit would result in \$400 billion in new spending over ten years. See, e.g., Robert Pear, *Deal 'In Principle' for Medicare Plan To Cover Drug Costs*, N.Y. TIMES, Nov. 15, 2003, at 1. The 17.5% increase in monthly premiums for Medicare beneficiaries in 2005 is the “largest premium increase in 15 years.” Connolly, *supra*.

43. “A health savings account is a tax-sheltered savings account similar to the IRA, but earmarked for medical expenses.” MSA (&HAS) Info.net, Info on Health Savings Accounts, at <http://www.msainfo.net/> (last visited Nov. 22, 2004).

44. See *supra* note 17 and accompanying text.

45. Critics allege that the overhead costs of private plans are at least five times those of public insurance. David Himmelstein of the Harvard Medical School asserts, “Medicare is actually much more efficient than the HMOs—it has 2 percent overhead, whereas they have 15 percent overhead.” Press Release, Common Dreams, Assessing Bush's Pharmaceutical Cards (July 12, 2001), <http://www.commondreams.org/news2001/0712-04.htm>. Elise Gould of the Economic Policy Institute similarly reports that the overhead costs of traditional Medicare, at less than four percent, are “super-low.” Elise Gould, *Bush Strikes Out on Health Care, Making Sense*, Dollars & Sense (May 2004), reprinted in Viewpoints, Economic Policy Institute, at http://www.epinet.org/content.cfm/webfeatures_viewpoints_healthcare_reform. She asks, “[I]f private insurance companies are so efficient, why do they need higher reimbursement fees?” *Id.* According to the federal Medical Payment Advisory Commission, Medicare payments to private plans total “an average of 107 percent of what it would cost to cover their patients under the traditional fee-for-service program.” Robert Pear, *Private Plans Costing More for Medicare*, N.Y. TIMES, Sept. 17, 2004, at A16.

46. It promises to reimburse the 251st dollar of drug spending, but not the 2251st dollar: For 2006, standard coverage under Part D of the MMA provides for a \$250 deductible, seventy-five percent coverage of allowable costs between \$251 and \$2250, zero percent coverage of costs between \$2251 and \$5100 (referred to as the “doughnut hole”), and ninety-five percent coverage above \$5100 in allowable costs. For allowable costs above \$5100, members are actually expected to pay the greater of either five percent of costs or

suggest that, except for the very poor and very sick, drug spending will consume a larger share of seniors' incomes in the coming years than it does now, despite the new legislation.⁴⁷ This is not just because of the gaps in coverage, but also because the bill fails to authorize the very negotiation strategies that large corporations and public programs like the veterans' health plan use to moderate skyrocketing drug prices.⁴⁸ Under the MMA, Medicare is expressly forbidden from using its bargaining power to negotiate for lower pharmaceutical prices.⁴⁹

These limitations help to explain why, according to polls, seniors are so critical of the reform. A University of Pennsylvania survey in December 2003, for example, showed opposition to the bill outweighing support by two percentage points among the general public, while opposition outweighed support by sixteen points among Americans over sixty-five.⁵⁰

co-payments of \$2 for generics and \$5 for brand name drugs. Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Pub. L. No. 108-173, §101(a)(2), 117 Stat. 2066, 2076-77. For references to the "doughnut hole," see, for example, Patricia Barry, *The New Medicare—And You*, AARP BULL., Jan. 2004, <http://www.aarp.org/bulletin/medicare/Articles/a2003-12-24-newmedicare.html> (last visited Dec. 4, 2004); Patricia M. Danzon, *Closing the Doughnut Hole: No Easy Answers*, HEALTH AFF. W4-405 (Web Exclusive, July 21, 2004), at <http://content.healthaffairs.org/cgi/reprint/hlthaff.w4.405v1>.

47. See, e.g., GAIL SHEARER, SKIMPY BENEFITS AND UNCHECKED EXPENDITURES: MEDICARE PRESCRIPTION DRUG BILLS FAIL TO OFFER ADEQUATE PROTECTION FOR SENIORS AND PEOPLE WITH DISABILITIES, Consumer's Union (June 2003), <http://www.consumersunion.org/pdf/medicare-603.PDF>.

48. See Robert B. Reich, *Buying Drugs in Bulk: It's Time To Negotiate*, AM. PROSPECT ONLINE, June 16, 2004, reprinted at Global Action on Aging, at <http://www.globalaging.org/health/us/2004/bulk.htm>.

49. Ironically enough, the *New York Times* reported in July 2004 that Medicare officials were announcing a plan to reduce the payments the program would be making in the future for drugs—especially cancer-related ones—administered in physicians' offices. Gardiner Harris, *Proposal Would Cut What Medicare Pays for Cancer Drugs*, N.Y. TIMES, July 27, 2004, at C1. This program, itself a part of Medicare known primarily by experts, patients, and their families, was part of the original 1965 legislation and has been gradually expanded to cover more and more drugs. See Thomas R. Oliver et al., *A Political History of Medicare and Prescription Drug Coverage*, 82 MILBANK Q. 283 (2004).

50. See Press Release, Annenberg Public Pol'y Ctr. of the Univ. of Pa., National Annenberg Election Survey, Public Split on Medicare Bill but Elderly Are Opposed, Annenberg Survey Shows (Nov. 24, 2003), http://www.annenbergpublicpolicycenter.org/03_political_communication/archive/naes/2003_03_support-medicare_pr.pdf; cf. Press Release, Annenberg Public Policy Ctr. of the Univ. of Pa., National Annenberg Election Survey, Public Initially Supports Medicare Law but Quickly Shows Its Doubts (Dec. 27, 2003), http://www.annenbergpublicpolicycenter.org/naes/2003_03_support-for-new

Ironically, Republicans who hoped to take Medicare off the political agenda, as it was an issue with which they had been battered for years, are similarly likely to see their hopes for the legislation frustrated. By pushing through such an unwieldy piece of legislation, they virtually ensured that Medicare will remain a contentious issue in American politics in the coming decade.

Some Democrats are hopeful that the bill will, in the long term, prove to be a stepping stone to a good drug benefit and more sensible Medicare reforms.⁵¹ Making the benefit more rational and generous, especially for low-income seniors and those with high, but not catastrophic, drug costs, is essential. The MMA, however, is unlikely to be a strong foundation for refinement and improvement down the line. The near-term issue will not be the expansion of benefits, but figuring out how to make the enormously complex legislation work. Furthermore, efforts to upgrade the benefit will run headlong into the massive budget deficit, and the fact that the profligate legislation has no effective cost-control mechanisms. The legislation's one concession to cost control—its resetting of the standard for program insolvency—will, in any case, create conflict highly unfavorable to those seeking to expand and rationalize benefits.⁵² Finally, the MMA's means-testing for Part B premiums may itself constitute a substantial barrier to future improvements, refinements, or expansions of Medicare. Practically speaking, by creating Medicare Part B premiums that will vary with income, Congress has established a system that will surely be cumbersome to administer. More broadly, as we have argued, the introduction of means-testing may provide a convenient cover for parties trying to produce an objectionable ideological transformation in the Medicare program.

Those committed to the central role of social insurance in modern America should understand the challenge to social insurance principles implicit in this debate. Advocates of means-testing on the right found a political wedge issue that split Medicare supporters on the left. But individuals, regardless of political orientation, who are genuinely concerned about America's low-income citizens should recognize that making well-to-do Medicare beneficiaries pay much more for Part B

-medicare-law_pr.pdf.

51. Conversation with John Rother, Chief Legislative Official, AARP, at Case Western Univ. (Oct. 4, 2004).

52. See *supra* note 18 and accompanying text. The requirement of presidential response to such insolvency is more likely to cause benefit cuts and premium hikes rather than benefit expansions.

coverage is not a sensible expression of decent social priorities. In light of both the historical structure of the Medicare program and its current fiscal circumstances, this seemingly innocent step forward represents a fundamental step backward.

The Pedagogical Significance of the Bush Stem Cell Policy: A Window into Bioethical Regulation in the United States

O. Carter Snead, J.D.*

The enormous significance of the Bush stem cell funding policy has been evident since its inception. The announcement of the policy on August 9, 2001 marked the first time a U.S. president had ever taken up a matter of bioethical import as the sole subject of a major national policy address. Indeed, the August 9th speech was the President's first nationally televised policy address of any kind.¹ Since then, the policy has been a constant focus of attention and discussion by political commentators, the print and broadcast media, advocacy organizations, scientists, elected officials, and candidates for all levels of office (including especially the 2004 Democratic nominee for President, Senator John Kerry, who made his opposition to the Bush policy a centerpiece of his domestic campaign, mentioning it explicitly in his acceptance speech at the Democratic National Convention).² The biotechnology industry has taken a keen

* General Counsel, The President's Council on Bioethics. The author would like to thank the *Yale Journal of Health Policy, Law, & Ethics* for soliciting this contribution. Special thanks also to Michelle Powers, Yuval Levin, Eric Cohen, John A. Ritsick, and Leigh Fitzpatrick Snead for their comments and support. All views expressed in this Essay are the author's own and are not meant to reflect the official position of the Council or the United States government.

1. See Amy Goldstein & Mike Allen, *Bush Backs Partial Stem Cell Funding*, WASH. POST, Aug. 10, 2001, at A1; Katharine Q. Seelye, *Bush Gives His Backing for Limited Research on Existing Stem Cells*, N. Y. TIMES, Aug. 10, 2001, at A1.

2. There have been a number of Congressional hearings illustrating the prominence of the issue. See *Embryonic Stem Cell Research: Exploring the Controversy: Hearing Before the Senate Commerce Subcomm. on Science, Technology & Space*, 108th Cong. (2004); *Adult Stem Cell Research: Hearing Before the Senate Commerce Subcomm. on Science, Technology & Space*, 108th Cong. (2004); *Hearing on Advances in Adult and Non-Embryonic Stem Cell Research: Hearing Before the Senate Commerce Subcomm. on Science, Technology & Space*, 108th Cong. (2004); *Opportunities and Advancements in Stem Cell Research: Hearing Before the Subcomm. on Criminal Justice, Drug Policy & Human Research of the Comm. on Government Reform*, 107th Cong. (2002); see also Laurie McGinley, *Stem-Cell Research Stirs Passionate Debate and Changing Politics*, WALL ST. J., July 9, 2001, at A30 (describing efforts of various advocacy organizations and

interest in stem cell research as a possible avenue for medical therapies; one study suggests that as of 2002 private sector companies had spent an aggregate of \$208 million on research and development of stem cell technologies.³ In response to the policy, there has been a flurry of state legislation proposed and enacted, with some states affirming and others condemning the Administration's approach.⁴ Finally, the great prominence of the national and international debate on human cloning has drawn further attention to the issue of embryonic stem cell research (and by extension, the Bush policy), given that one application of somatic cell nuclear transfer is the production of cloned human embryos from which stem cells may be derived (so-called "Therapeutic Cloning").⁵

To date, the significance of the Bush stem cell policy has been framed and publicly debated in terms of its practical import: Does it impede the scientific and medical progress that the research seems to promise? Is it adequately protective and respectful of embryonic human life? Aside from its great practical significance, however, the Bush policy is arguably one of the most important recent legal developments for the field of bioethics for an additional reason: its deep pedagogical significance. The Bush policy provides an unparalleled window into the nature and substance of "bioethical regulation" within the unique framework of the American system of government. And it does so in dramatic fashion, against the backdrop of some of the most enduring and vexing questions in all of bioethics: What is owed to developing human life, and how does this

scientists); Alexa H. Bluth & Laura Mecoy, *Boxer, Jones Split on Stem-Cell Issue*, SACRAMENTO BEE, Aug. 11, 2004, at A3; Judith Graham, *Quest for Cures Spurs Fierce Debate*, CHI. TRIB., July 27, 2004, at C16; Senator John Kerry, Speech at the 2004 Democratic National Convention (July 29, 2004), at http://www.johnkerry.com/pressroom/speeches/spc_2004_0729.html.

3. Michael J. Lysaght & Anne L. Hazlehurst, *Private Sector Development of Stem Cell Technology and Therapeutic Cloning*, 9 TISSUE ENGINEERING 555, 557 (2003).

4. See Judith Graham, *States Are Wrestling with Stem-Cell Issues*, CHI. TRIB., Apr. 6, 2004, at C1 (noting state efforts in California, New Jersey, Illinois, Connecticut, Maryland, Massachusetts, Minnesota, Pennsylvania, New York, Rhode Island, Tennessee, and Washington); Carl Ingram, *Stem Cell Initiative Certified for Balloting*, L.A. TIMES, June 4, 2004, at A1; Jeffrey Krasner, *Massachusetts Legislature Debates Stem Cell Research Bill*, BOSTON GLOBE, May 2, 2003, at A1. For a list of pending and recently enacted state laws relating to embryonic stem cell research, see Lori Andrews, *Legislators as Lobbyists: Proposed State Regulation of Embryonic Stem Cell Research, Therapeutic Cloning and Reproductive Cloning*, in PRESIDENT'S COUNCIL ON BIOETHICS, MONITORING STEM CELL RESEARCH (2004) [hereinafter PRESIDENT'S COUNCIL].

5. See, e.g., Woo Suk Hwang et al., *Evidence of Pluripotent Human Embryonic Stem Cell Line Derived from a Cloned Human Blastocyst*, 303 SCI. EXPRESS 1669 (2004).

obligation stand in relation to the aim of science to advance knowledge with the ultimate aspiration of alleviating human suffering? Reflecting on the nature and scope of the policy yields insights into a number of crucial matters that are central to the problem of whether and how to govern science and medicine according to bioethical principles. This Essay will briefly explore five areas in which the Bush policy is thus instructive: (1) the conceptual understanding of “regulation” as a legal category; (2) the principles of federalism; (3) the significance of federal funding; (4) the nature of governance according to a particular type of moral principle (e.g. “bright line”); and (5) the influence of political prudence and respect for pluralism.

I. THE BUSH POLICY

Before proceeding to a discussion of the lessons of the Bush policy, it is useful to articulate briefly the contours of the policy itself. To understand the current policy in its full context, one needs a brief account of the federal government’s historical role in the regulation of human embryo research. In 1975, a federal rule was enacted providing that “[n]o application or proposal involving human *in vitro* fertilization may be funded by the Department [of Health and Human Services] [until it] has been reviewed by the Ethical [later “Ethics”] Advisory Board and the Board has rendered advice as to its acceptability from an ethical standpoint.”⁶ In 1979, the Ethics Advisory Board (EAB) issued a report concluding that it is ethically acceptable to provide federal funding for embryo research under certain circumstances.⁷ The Department did not act on this recommendation, however, and the charter of the EAB expired in 1980. Thereafter, the EAB was not reconstituted, though the federal rule requiring EAB approval for federal funding for any research involving *in vitro* embryos remained in effect. The result was a de facto moratorium on federal funding for research involving human embryos until 1993, when Congress (acting at the insistence of the newly elected Clinton Administration) rescinded the EAB approval requirement,⁸ effectively clearing the way for the federal funding of embryo research. Before any proposals were funded, however, the newly elected Congress intervened,

6. Ethical Advisory Boards, 45 C.F.R. § 46.204(d) (1982) (later repealed).

7. HEW Support of Research Involving Human *In Vitro* Fertilization and Embryo Transfer: Report of the Ethics Advisory Board, 44 Fed. Reg. 35,033, 35,055-58 (June 18, 1979).

8. See National Institutes of Health Revitalization Act of 1993, Pub. L. No. 103-43, § 121(c), 107 Stat. 122 (1993) (repealing 45 C.F.R. § 46.204(d)).

attaching language to the 1996 Departments of Labor, Health and Human Services, and Education, and Related Agencies Appropriations Act that formally precluded the use of federal funds for "the creation of a human embryo or embryos for research purposes; or [for] research in which a human embryo or embryos are destroyed, discarded, or knowingly subjected to risk of injury or death greater than that allowed for research on fetuses in utero" under the controlling human subjects protection regulations.⁹ This language (known as the "Dickey Amendment," after its original sponsor) has been re-enacted in every HHS appropriations bill since 1996.

Three years later, in the wake of widespread enthusiasm and excitement in the scientific community over the first reported isolation of human embryonic stem cells, the General Counsel of the Department of Health and Human Services urged an interpretation of the Dickey Amendment that would allow for federal funding of research involving embryonic stem cells.¹⁰ The General Counsel argued that because the Dickey Amendment only precluded the provision of federal funding to research in which embryos were destroyed, it would be legally permissible to authorize federal funding for researchers who worked with stem cells acquired from embryos that had been destroyed with only private funding.¹¹ Supporters of the Dickey Amendment, including Representative Dickey himself, strenuously objected to this interpretation, arguing that it contradicted the spirit of the federal law by allowing the use of public funds in a way that would create incentives for the destruction of embryonic human life.¹² Secretary Shalala and President Clinton rejected this critique, and made preparations for the federal funding of embryonic stem cell research.¹³ Before the Clinton funding policy was implemented, however, President Bush was elected.

9. Pub. L. No. 104-99, § 128, 110 Stat. 26 (1996).

10. Memorandum from Harriet S. Raab, General Counsel of the Department of Health and Human Services, to Harold Varmus, Director of the National Institutes of Health, Federal Funding for Research Involving Human Pluripotent Stem Cells (Jan. 15, 1999) (on file with the National Archives).

11. *See id.*

12. *See* Letter from Representative Jay Dickey, to Secretary of Health and Human Services Donna E. Shalala (Feb. 11, 1999) (on file with author) (signed by seventy members of Congress).

13. *See* Letter from Secretary of Health and Human Services Donna E. Shalala, to Representative Christopher H. Smith (Feb. 23, 1999) (on file with author); *see also* National Institutes of Health Guidelines for Research Using Human Pluripotent Stem Cells, 65 Fed. Reg. 51,976 (Aug. 25, 2000).

Against the backdrop of this twenty-five year history, President Bush was confronted with the question of whether and how to fund stem cell research. President Bush accepted the legal analysis of the former HHS General Counsel, but pursued a policy that sought to combine that analysis with the principle animating the Dickey Amendment, namely, that human life is worthy of profound respect at all of its developmental stages (from zygote to adult), and therefore, at the very least, the federal government should not provide financial incentives for its destruction, even for the sake of beneficial scientific research.¹⁴ President Bush thus formulated a stem cell funding policy that would, in his words, “aggressively promote stem cell research” without violating his aforementioned principle of respect for human embryonic life.¹⁵ In practice, the Bush policy authorizes federal funding for all forms of stem cell research that do not create incentives for the destruction of human embryos.¹⁶ Accordingly, research using stem cells derived from non-embryonic (commonly referred to as “adult”) sources (e.g., from bone marrow, umbilical cord blood, etc.) are subject to unlimited funding, as the derivation of these cells does not cause significant or lasting harm to the donor. Similarly, federal funding is authorized for research on those human embryonic stem cell lines derived prior to the date of the announcement of the policy (provided that they were obtained in a manner that observed the traditional standards of research ethics—*informed consent of the donor, etc.*).¹⁷ So as not to encourage future destruction of human embryos, no federal funding is permitted for research on embryonic stem cell lines derived after August 9, 2001. For fiscal year 2003, the Bush Administration, through NIH, allocated \$190.7 million for adult stem cell research, and \$24.8 million for

14. See George W. Bush, *Stem Cell Science and the Preservation of Life*, N.Y. TIMES, Aug. 12, 2001, at D13.

15. *Id.*

16. See *id.* (“Federal funding for research on existing stem cell lines will move forward; federal funding that sanctions or encourages the destruction of additional embryos will not.”).

17. There are seventy-eight such lines that are “eligible” for federal funding. However, before a stem cell line becomes “available” for use and distribution, it must be grown into a stable cultured population (a scientific process) and the relevant “Material Transfer Agreements” must be negotiated and executed (a legal process). In the summer of 2002, only one of the eligible lines was available. In the fall 2003, that number rose to twelve. As of August 2004, there are twenty-two lines available for use and distribution. See Nat’l Insts. of Health, Information on Eligibility Criteria for Federal Funding of Research on Human Embryonic Stem Cells, at <http://stemcells.nih.gov/research/registry/eligibilitycriteria.asp> (last modified Aug. 11, 2004).

embryonic stem cell research.¹⁸ Additionally, the NIH, acting at the behest of the Bush Administration, has created a "Stem Cell Task Force" which seeks to "accelerate the pace of stem cell research by identifying the rate limiting resources (both material and human) and [to] develop initiatives to enhance these resources."¹⁹ The Bush policy imposes no restrictions on privately funded embryonic stem cell research; indeed the Administration clarified a previously enacted administrative rule so as to make it simpler for otherwise federally-funded scientists and institutions to pursue embryonic stem cell research using private funds.²⁰

II. THE PEDAGOGICAL SIGNIFICANCE OF THE BUSH POLICY

What, then, is the pedagogical significance of the Bush policy? As noted above, a careful consideration of the policy's scope and substance yields at least five different (yet related) insights into the nature of bioethical regulation in the United States. Each will be discussed separately.

A. "Regulation" as a Legal Concept

The first lesson is that "regulation" is a complex and multifaceted concept in American law. Regulation is not simply a matter of proscription and permission. Rather, it is a spectrum of legal activity by which the government can voice (or not voice, as the case may be) the values and priorities of the polity. At one end of the spectrum lies prohibition, in which the government forbids a given activity. The most obvious and dramatic example of this form of regulation is criminal proscription. At the other end lies affirmative encouragement, whereby the government

18. See U.S. Dep't of Health & Human Servs., HHS Fact Sheet: Embryonic Stem Cell Research (July 14, 2004), at <http://www.hhs.gov/news/press/2004pres/20040714b.html> [hereinafter HHS Fact Sheet]. This marks the first time in history that federal funds have been allowed for research that requires the destruction of human embryos.

19. Nat'l Insts. of Health, NIH Stem Cell Task Force, at <http://stemcells.nih.gov/policy/taskForce/> (last modified Oct. 28, 2004). Such initiatives have included grants and awards for infrastructure and training to improve distribution and development of approved cell lines and programs to train researchers in areas such as culture techniques. More recently, the NIH announced plans to open a "National Embryo Stem Cell Bank" for approved lines, in an effort to increase their availability to researchers. See Letter from Secretary of Health and Humans Services Tommy G. Thompson, to Representative J. Dennis Hastert, Speaker, U.S. House of Representatives (July 14, 2004) (on file with author); see also HHS Fact Sheet, *supra* note 18.

20. See HHS Fact Sheet, *supra* note 18.

rewards behavior that the polity deems worthwhile and useful. The most obvious example of this form of regulation is the provision of government funding. In between these poles of prohibition and encouragement lie myriad mechanisms by which the government speaks, including, among other things, recordkeeping (showing the government's view that "attention must be paid"), silence (signifying governmental permission without explicit endorsement), and permission with conditions (signaling a qualified endorsement of the underlying activity, with acknowledgement that some measure of oversight is required).

The Bush policy vividly illustrates the complexity of "regulation" as a legal concept by incorporating elements from across this spectrum. The policy adopts a posture of silence toward privately sponsored stem cell research that involves the destruction of human embryos, signaling the federal government's permission without explicit endorsement of this practice. At the same time, it provides significant financial incentives for stem cell research that does not involve the present and future destruction of human embryos, showing the federal government's endorsement and approval of this species of research. The significance of federal funding is further expanded and amplified in Section II.C.

B. Principles of Federalism

The Bush policy further illustrates how matters of federalism—both horizontal²¹ and vertical²²—are implicated in the context of bioethical governance. Principles of horizontal federalism play an important role in the formulation and implementation of public policy that touches and concerns bioethics. In making such policy, each co-equal branch must act within the boundaries of its own enumerated powers, while respecting the prerogatives and domains of the others. This process is brought into sharp relief by a reflection on the Bush policy's origins and operation, described above. The Bush policy was written against the backdrop of the nearly thirty-year history of give and take between the executive and legislative branches over the question of federal funding for embryo research. As discussed previously, this inter-branch dialogue culminated in the enactment of the Dickey Amendment, whereby the legislative branch,

21. "Horizontal federalism" refers to the relationship between and among the legislative, executive, and judicial branches of the federal government. *See, e.g.*, L. Harold Levinson, *The Legitimate Expectation That Public Officials Will Act Consistently*, 46 AM. J. COMP. L. 549, 551-52 (1998).

22. "Vertical federalism" denotes the relationship between the federal government and state governments. *See id.* at 552.

acting pursuant to its constitutionally enumerated spending power, formally proscribed the use of federal funds for research in which human embryos are destroyed or discarded. In formulating a policy governing stem cell research and its funding, the Bush Administration (like the Clinton Administration before it) was required to work within the framework provided by Dickey out of respect for the federalist principle of separation of powers. The Bush policy accepted the Clinton Administration's refined interpretation of Dickey, but chose a policy that upheld a broad conception of the principle of respect for embryonic human life that provided the foundation for the original amendment. Thus, the Bush policy demonstrates both an acknowledgement of Congress's sole authority to appropriate federal funds and a robust exercise of the President's authority as head of the executive branch to allocate the appropriated funding according to the Administration's priorities.

In similar fashion, reflection on the Bush policy lends key insights into principles of vertical federalism in the context of bioethical governance. In enacting public policy, both state and federal governments are limited by their respective jurisdictional mechanisms. By virtue of the general police power to safeguard the health, welfare, and morals of citizens, states enjoy wide latitude to legislate according to bioethical principles.²³ By contrast, the federal government is somewhat more limited in its options, consigned to act only pursuant to powers enumerated by the Constitution.²⁴ This division of responsibility allows in some cases for action and reaction between and among the federal and state governments.

Such is the case with the Bush policy. The Bush policy illustrates the use of the jurisdictional nexus of federal spending: The Administration is able to set ethical conditions on those practices to which it provides financial assistance, while remaining silent (and thus uninvolved) with respect to privately funded stem cell research. This leaves the state governments free to affirm or reject the policy within their own borders.

23. *See, e.g., Gibbons v. Ogden*, 22 U.S. (9 Wheat.) 1, 203 (1824) ("Inspection laws, quarantine laws, health laws of every description, as well as laws for regulating the internal commerce of a State No direct general power over these objects is granted to Congress; and, consequently, they remain subject to State legislation."); *see also Washington v. Glucksberg*, 521 U.S. 702 (1997) (upholding assisted-suicide ban as rationally related to legitimate state interest).

24. *See United States v. Morrison*, 529 U.S. 598, 607 (2000) ("Every law enacted by Congress must be based on one or more of its powers enumerated in the Constitution."); *United States v. Lopez*, 514 U.S. 549, 552 (1995) (same); *McCulloch v. Maryland*, 17 U.S. (4 Wheat.) 316, 405 (1819); *Marbury v. Madison*, 5 U.S. (1 Cranch) 137, 176 (1803).

Many states have taken this opportunity. On one end of the spectrum, there are states such as Louisiana, which bans destructive embryo research altogether.²⁵ On the other end of the spectrum, there are states such as New Jersey²⁶ and California,²⁷ which have explicitly endorsed embryonic stem cell research and cloning for biomedical research.²⁸

C. The Significance of Federal Funding

The Bush policy also offers noteworthy lessons regarding the nature and significance of federal funding. The U.S. government is a major provider of funds and resources for scientific and medical research.²⁹ This is reflective of the esteem in which the American polity holds the scientific enterprise, as well as its great concern for the alleviation of human suffering. Federal funding has long played a significant role in the regulation of medicine and science according to bioethical principles. In the first instance, it is a jurisdictional nexus, allowing for the regulation of activities that might otherwise lie beyond the enumerated powers of the federal government by attaching certain conditions to the provision of funds.³⁰ But perhaps more importantly for the present discussion, federal funding is a powerful device whereby the government expresses the polity's approval, disdain, or studied neutrality toward specified conduct. The government is under no obligation to provide federal funding for most activities—including those activities in which individuals may engage as a matter of constitutional right.³¹ Thus, the provision of federal funding can

25. See LA. REV. STAT. ANN. § 9:129 (West 1986). It is important to note, however, that Louisiana defines "embryo" as the product of the union of egg and sperm, thus excluding cloned embryos from legal protection. See LA. REV. STAT. ANN. § 9:121 (West 1986).

26. See N.J. STAT. ANN. 26:2Z-2 (West 2004).

27. See S.B. 322, 2003-04 Sess. (Cal. 2004) (signed by Governor Davis on Sept. 24, 2003).

28. In November 2004, California voted in a statewide referendum on a measure that both amends the state constitution to establish a "Right to Conduct Stem Cell Research," and calls for the issuance of three billion dollars of general obligation bonds to provide funding for stem cell research. See California Stem Cell Research and Cures Initiative, Proposition 71 (Cal. 2004), www.voterguide.ss.ca.gov/propositions/prop71text.pdf.

29. For a detailed breakdown of the funding from the National Institutes of Health for various research projects, see Nat'l Insts. of Health, Estimates of Funding for Various Diseases, Conditions, Research Areas, at <http://www.nih.gov/news/fundingresearch/areas.htm> (last visited Oct. 15, 2004).

30. See, e.g., *United States v. Butler*, 297 U.S. 1, 66 (1936) ("[T]he power of Congress to authorize expenditure of public moneys for public purposes is not limited by the direct grants of legislative power found in the Constitution.").

31. See *Rust v. Sullivan*, 500 U.S. 173 (1991); *Harris v. McRae*, 448 U.S. 297 (1980);

confer legitimacy on a given enterprise, signaling its worthiness for the allocation of otherwise scarce funds. The withholding of federal funds can signify a variety of sentiments: a lack of faith in the worthiness (moral or otherwise) of the enterprise, moral caution or affirmative disdain for the activity in question, or simply the judgment that there are more important priorities worthy of the expenditure of limited resources.

The Bush policy is instructive in this regard. It does, as mentioned above, utilize funding as a jurisdictional nexus. But it also conveys a message regarding the priorities of the Administration. First, it requires the federal government to adopt a posture of neutrality in the debate over the moral propriety of destructive embryo research. The Bush policy affirmatively and deliberately withholds the federal government's official approval for such practices, though it does allow these practices to proceed in the private sector. As such, no taxpayer is compelled to pay for and encourage an activity (i.e., embryo destruction) that a significant portion of the American public finds morally troublesome.³² At the same time, the Bush policy was designed in an effort to reflect the government's commitment "to fully exploring the promise and potential of stem cell research"³³ without running afoul of the particular moral and ethical principles set forth and embraced by President Bush in announcing the policy.

D. Governance According to a "Bright Line" Moral Principle

The Bush policy provides a rich and complex example of one particular approach to "bioethical governance." It is not driven by a utilitarian weighing of commensurate values, but rather begins with a clear moral standard that may not be transgressed. In his August 9, 2001 speech, and in an editorial printed in the *New York Times* three days later, President Bush said: "There is at least one bright line: We do not end some lives for the medical benefit of others. For me, this is a matter of conviction: a

Maher v. Roe, 432 U.S. 464 (1977).

32. See Matthew Nisbit, *Public Opinion About Stem Cell Research and Human Cloning*, 68 PUB. OPINION Q. 131, 135 (2004) (noting that in two separate Gallup Polls asking respondents whether they found medical research using cells obtained from human embryos to be morally acceptable or morally wrong, more than half in both polls said it is "morally wrong"). It is important to note, however, that polls regarding public support for embryonic stem cell research have varied widely: Many show widespread support for the practice, while others show widespread opposition. See generally NAT'L INSTS. OF HEALTH, REPORT OF THE HUMAN EMBRYO RESEARCH PANEL 44-45 (Sept. 1994).

33. See HHS Fact Sheet, *supra* note 18.

belief that life, including early life, is biologically human, genetically distinct, and valuable.”³⁴ This is the moral and ethical foundation upon which the Bush policy is erected. The Administration’s stated desire to better the human condition by eradicating dreaded diseases and debilitating injuries, and its attendant enthusiasm and support for scientific research aimed at these goals, are thus expressed and acted upon within the boundaries of this moral framework. Accordingly, the Bush policy is designed to endorse and actively promote all stem cell research (including embryonic) that does not encourage the future instrumentalization and destruction of human embryos.

This bright line policy stands in contrast to the balancing approaches espoused by other commentators and governmental advisory bodies. Both the NIH Human Embryo Panel and President Clinton’s National Bioethics Advisory Commission (NBAC) promoted the view that while human embryos deserve special respect as a form of developing human life, it is possible to balance this respect against the benefits of scientific research that might be achieved through research that necessarily requires the destruction of such embryos.³⁵ This approach led both bodies to recommend that such research was ethically acceptable, and that the federal government should fund such research, subject to various conditions. The NIH Human Embryo Panel went somewhat further, issuing a qualified endorsement of the creation of embryos solely for the sake of research.³⁶

The policy also teaches that policies originating from this species of bioethical governance—based on a bright line moral principle rather than a balancing of values or “compromise” (in the conventional sense)—are not alterable on the basis of a showing that the benefits of transgressing the established boundary would be higher than originally thought—even by orders of magnitude. Put concretely, the Bush policy (given the species of bioethical regulation that it represents) would almost certainly not be revised or reversed, even if tomorrow there were incontrovertible evidence that greater benefits could be realized by federally funding future derivations of embryonic stem cell lines. To do so would undermine the very “bright line” that animates the entire policy, namely, that destruction of human embryos should not be encouraged or incentivized by the promise of future federal funding.

34. George W. Bush, *Stem Cell Science and the Preservation of Life*, N.Y. TIMES, Aug. 12, 2001, at D13.

35. See PRESIDENT’S COUNCIL, *supra* note 4, at 82-84.

36. See NAT’L INSTS. OF HEALTH, *supra* note 32, at 44-45.

E. Political Prudence and Respect for Pluralism

While the Bush policy provides insight into a particular species of moral governance, it also teaches one way in which the formulation of bioethical policies is influenced by considerations of political prudence and respect for pluralism. Although the moral foundation of the Bush policy is a view that human beings are worthy of maximal respect regardless of their developmental stage and that ending some human lives for the medical benefit of others is unethical, the Bush policy does not seek to ban destructive embryo research altogether. To the contrary, it steers a more moderate course, merely withholding the government's affirmative endorsement of the practice by way of federal funding. What is the significance of this tension between the Bush policy's moral principle and its practical effect? One can only speculate, but there are several possibilities (or combinations of possibilities).

First, this apparent disconnect might reflect the Administration's acknowledgement of the moral, ethical, and legal discourse as it has evolved over the past thirty years. As a matter of historical context, the debate over the federal government's role vis-à-vis embryo research has been consistently framed in terms of funding rather than permission and proscription more broadly. The battle lines, so to speak, were drawn before the Bush Administration came on the scene. Indeed, the issue of stem cell research was presented to the Administration in the form of a question about funding and in the legal context of interpreting and implementing the Dickey Amendment.

Second, the modest nature of the Bush policy might be interpreted as a certain type of incrementalism. That is, it might bespeak the Administration's desire to avoid overreaching in such a controversial bioethical context. There is a deep divide within the American polity on the question of what is owed to human embryos, and an Administration that sought to impose a novel and restrictive policy in an area where there has historically been little government involvement risks polarization and backlash. Thus, it is possible to construe the Bush policy as reflecting the view that a judicious incrementalism is the most appropriate course for winning public support over the long term for policies that originate from the bright line principle that embryonic human life is inviolable. The limit on federal funding might thus be interpreted as laying the groundwork for a larger effort to convince the American public of the Administration's views regarding the respect owed to human embryos more generally.

Finally, the restrained nature of the Bush policy might also (or alternatively) serve to demonstrate how considerations of pluralism can

affect the formulation of bioethical public policy. While the Bush approach begins with the moral judgment that human embryos should not be instrumentalized or destroyed for the sake of another's medical benefit, the ultimate legal expression of this policy implicitly acknowledges that there is great division among the American citizenry on this point by remaining neutral on the ultimate question of the legal permissibility of embryo research. The policy does not ban the destruction of human embryos to derive embryonic stem cells, but it does withhold the government's official approval and refuses to compel American taxpayers to subsidize an activity that is a source of great moral and ethical disquiet for a significant portion of the population.³⁷ The Bush policy could thus be seen as an example of how the government can express its ethical approval (or disapproval) of a particular type of scientific activity while respecting the deep disagreements that persist in society.

37. Polling in this area has reached varied results, not surprisingly turning largely on how the question is framed and what information is provided to respondents. In polls in which respondents are asked if they support "stem cell research" but are not explicitly told that the derivation of embryonic stem cells requires the destruction of human embryos, opposition ranges from twenty-eight percent to thirty-five percent. See Press Release, Nat'l Annenberg Election Survey, Public Favors Stem Cell Research, Annenberg Polling Data Show (Aug. 9, 2004), http://www.annenbergpublicpolicycenter.org/naes/2004_03_stem-cell_08-09_pr.pdf (finding that twenty-eight percent of respondents opposed "federal funding of research on diseases like Alzheimer's using stem cells *taken from human embryos*"); Press Release, Pew Forum on Religion & Public Life, Cloning Opposed, Stem Cell Research Narrowly Supported (Apr. 9, 2002), <http://pewforum.org/publications/surveys/bioethics.pdf> (finding that thirty-five percent of respondents oppose federal funding for "stem cell research"). By contrast, in polls where respondents are explicitly told that the research requires the destruction of human embryos (but without explicit reference to the possible therapeutic benefits of the research), opposition increases to between fifty-three percent and sixty-one percent. See *Poll: Americans Oppose Destroying Human Embryos, Cloning*, TIDINGS, Aug. 27, 2004, <http://www.the-tidings.com/2004/0827/stemcell.htm> (showing that 61.4% of respondents oppose federal funding of stem cell research in which "embryos are destroyed in their first week of development"); Press Release, Nat'l Right to Life Comm., Majority Opposes Tax Funding of Stem Cell Research That Kills Human Embryos (Aug. 23, 2004) (showing that fifty-three percent of respondents opposed "using tax dollars to pay for the kind of research that requires the killing of human embryos."), http://www.nrlc.org/Killing_Embryos/Release082304.html. It does not appear that any public survey has yet been conducted in which respondents are advised explicitly of *both* the possible therapeutic benefits *and* the fact that embryos are destroyed in the derivation process.

CONCLUSION

In sum, the Bush stem cell policy is one of the most significant recent legal developments with bioethical implications—not merely for its practical impact on scientific research or the use and disposition of human embryos, but also for what it teaches about the nature of bioethical regulation in the United States. As this Essay has attempted to show, a careful consideration of the Bush policy leads one to key insights relating to the manifold character of regulation, principles of federalism, the significance of federal funding, the nature of governance according to a “bright line” moral principle, political prudential judgments, and the impact of a respect for pluralism in the bioethical context. Such insights are crucial to a robust understanding of the still raging debate over the federal regulation of and support for embryonic stem cell research, as well as to a comprehensive appreciation of “bioethical regulation” more generally.

Erratum

In Volume 4, Issue 2 (Summer 2004) of the *Yale Journal of Health Policy, Law, and Ethics*, Nicolas P. Terry's degree was incorrectly identified as a J.D. Professor Terry, the author of *Prescriptions sans Frontières*, holds an LL.M.

