

TABLE OF CONTENTS

ARTICLES

- 507 Wrongful Abortion: A Wrong in Search of a Remedy**
Ronen Perry, LL.D. and Yehuda Adar, LL.D.
- 587 The FDA and the Tort System: Postmarketing Surveillance, Compensation, and the Role of Litigation**
Catherine T. Struve, J.D.

NOTE

- 671 Juvenile Mental Health Courts and Therapeutic Jurisprudence: Facing the Challenges Posed by Youth with Mental Disabilities in the Juvenile Justice System**
Patrick Geary

SYMPOSIUM—PHARMACEUTICAL INNOVATION AND COST: AN AMERICAN DILEMMA

- 713 Introduction**
Mark Siegler, M.D., Alix Weisfeld, and Richard A. Epstein, LL.B.
- 717 The Problem of New Uses**
Rebecca S. Eisenberg, J.D.
- 741 Regulatory Paternalism in the Market for Drugs: Lessons from Vioxx and Celebrex**
Richard A. Epstein, LL.B.
- 771 The Patient's Role in Choice of Medications: Direct-to-Consumer Advertising and Patient Decision Aids**
Marshall H. Chin, M.D., M.P.H.
- 785 The Effects and Role of Direct-to-Physician Marketing in the Pharmaceutical Industry: An Integrative Review**
Puneet Manchanda, M.Phil., Ph.D. and Elisabeth Honka

ARTICLES

Wrongful Abortion: A Wrong in Search of a Remedy

Ronen Perry, LL.D.* and Yehuda Adar, LL.D.†

INTRODUCTION

“Wrongful abortion” is an abortion that a pregnant woman is induced to undergo by negligent conduct (usually a medical misrepresentation).¹ As an example, early in her pregnancy a woman is told by her physician that a medication that she had taken will cause her baby to be born with a severe birth defect. Based on the expert opinion, she decides to undergo an abortion. Only after the abortion does she learn that the advice regarding the baby’s health was a negligent misrepresentation and that the termination of the pregnancy was unnecessary.²

In this Article, we argue that the law does not currently provide adequate incentives to avoid wrongful abortions, the consequences of which are often devastating. We suggest that the best solution to this problem may be built on the distinctive characteristics of the wrongful abortion setting. Validating the intuition that the status quo does not adequately respond to wrongful abortions requires a systematic and comprehensive analysis of existing law, and justifying our novel solution entails a thorough theoretical inquiry.

Accordingly, this Article addresses two interrelated questions. First, how is

* Professor, Faculty of Law, University of Haifa.

† Adjunct Professor, Faculty of Law, University of Haifa. Visiting Scholar, Faculty of Law, University of Toronto.

The authors are grateful to the editors of the *Yale Journal of Health Policy, Law, & Ethics* for their valuable comments.

1. As we shall see below, the term “wrongful abortion” may be used in a much broader sense, but we shall focus on its narrowly defined meaning for reasons also set out below.

2. See *Martinez v. Long Island Jewish Hillside Med. Ctr.*, 512 N.E.2d 538, 538 (N.Y. 1987).

existing law likely to respond to wrongful abortions? We intentionally ask how the law is “likely to respond” and not how it actually responds. The problem of wrongful abortion has been the subject of judicial opinion only in a few sporadic cases, making it practically impossible to generate a comprehensive analysis of case law directly on this point. Our effort will therefore focus on identifying the legal issues involved and resolving them within existing (and relevant) legal frameworks.

Second, how *should* the law respond to wrongful abortions? Wrongful abortions raise a unique problem to which current law does not provide an appropriate solution. Our objective is to discuss the various alternatives that policymakers might consider in response to this peculiar disparity.

As there are only a handful of cases on the subject, and since our topic has not been discussed in any detail in the academic literature,³ we find it almost unavoidable to open this Article with an analysis of the factual settings in which wrongful abortions occur and a systematic itemization of their consequences. In Section I.A, we define more accurately the term “wrongful abortion” and explain through contrast and analogy the settings in which wrongful abortion cases arise. In Section I.B, we survey the social costs of wrongful abortions within two distinct categories: (a) parental losses, and (b) “loss of potential life,” i.e., any loss that may be attributed to the destruction of potential human life. We point out where these two categories overlap and explain why they nonetheless merit separate discussion.

Parts II and III discuss the anticipated legal response to wrongful abortions. In Part II we demonstrate that the law may respond quite effectively (although somewhat imperfectly) to parental losses within the traditional framework of tort law. In contrast, we show in Part III that nationwide, all branches of the law currently leave the loss of potential life (except for any overlap with parental losses) unaccounted for in most cases of wrongful abortion. We believe this is a significant and disturbing anomaly in American law, given that states’ important and legitimate interest in preserving potential life has been well established in American legal thought.

Finally, in Part IV we endeavor to find an appropriate legal means to rectify this inconsistency. We introduce and critically evaluate three possible legal paths of resolution: extending criminal liability to cover negligent inducement of abortion; expanding civil liability to cover the elements of the loss of potential life not currently compensable under tort law; and a discretionary civil fine,

3. The topic was mentioned only once, and rather briefly, in an American legal periodical. See Kathy Seward Northern, *Procreative Torts: Enhancing the Common-Law Protection for Reproductive Autonomy*, 1998 U. ILL. L. REV. 489, 527-29. Northern’s article is mainly concerned with the violation of pregnant women’s procreative autonomy.

which we find the most attractive legal solution to the problem.

At the outset, we emphasize that we take no stand concerning the fierce ongoing battle between the pro-life and pro-choice movements. We have no desire to provide either side with academic ammunition. We hope that both sides will support our effort: The proposed solution enhances the legal protection of states' interest in preserving potential life and at the same time improves the legal protection of the pregnant woman's right to privacy. It should be remembered that prevention of a wrongful abortion fulfills the prospective mother's true will. Therefore, our endeavor to strengthen the legal protection of the public interest in preserving potential life does not weaken the pregnant woman's procreative autonomy. On the contrary, since in cases of wrongful abortion the public interest in preserving potential life and the prospective mother's will coincide, our proposal simultaneously reinforces both without taking a stance on the appropriate scope of the abortion right.

Supporters of abortion rights may argue that enhancement of the legal protection of potential life in various contexts may eventually undermine the mother's procreative autonomy. However, our proposal does not consider the unborn child to be a legal person.⁴ It is possible to protect a public interest in preserving certain life forms without recognizing their legal personhood. Legal protection of animals and plants serve as good examples. As adopting our proposal does not necessitate recognition of legal personhood in a fetus, it poses no risk to the mother's right to privacy (as long as and to the extent that it is recognized under the Constitution).

Moreover, even if we based our proposal on the assumption that the fetus were a legal person, this would not necessarily turn the non-viable fetus into a *constitutional* person. In American jurisprudence, the idea that a legal person may not have all the rights of a constitutional person is well-established. Since only the interests of a constitutional person might be strong enough to overcome the pregnant woman's constitutional rights, recognizing the legal personhood of a fetus (at any stage of development) would not directly affect the pregnant woman's right to choose abortion.⁵ There is, however, the real possibility that a sweeping recognition of the legal personhood of fetuses in cases of wrongful abortion and their like may lead the Supreme Court to reevaluate its prior decisions on abortion rights. For those who fear this result, our proposal may be

4. However, several criminal statutes do consider the unborn child to be a legal person. *See, e.g.,* Unborn Victims of Violence Act, 18 U.S.C.A. § 1841(a)(2)(C), (d) (West Supp. 2004) (stating that a fetus is a human being at any stage of gestation).

5. *Cf. Jeffrey Rosen, A Viable Solution: Why It Makes Sense To Permit Abortions and Punish Those Who Kill Fetuses*, LEGAL AFF., Sept.-Oct. 2003, at 20, 21.

preferable to at least one of the alternatives.⁶

I. WRONGFUL ABORTION AND ITS SOCIAL COSTS

A. The Factual Setting

“Wrongful abortion” will be used in this Article to describe an abortion *instigated* by wrongful conduct, as opposed to an abortion which is *performed* in a wrongful manner,⁷ although this term may be construed to embrace both types of cases.⁸ We further limit the scope of the term “wrongful abortion” to negligent inducement of abortion,⁹ as opposed to intentional inducement. Negligently induced abortions seem to us more interesting from a theoretical standpoint.

To epitomize the various settings in which we believe an abortion is negligently induced, we use the seminal decision of the Supreme Court in *Roe v. Wade*¹⁰ as our starting point. This ruling articulated the scope of states’ authority to regulate abortions. The guidelines were set in Justice Blackmun’s opinion as follows:

(a) For the stage prior to approximately the end of the first trimester, the abortion decision and its effectuation must be left to the medical judgment of the pregnant woman’s attending physician.

(b) For the stage subsequent to approximately the end of the first trimester, the State, in promoting its interest in the health of the mother, may, if it chooses, regulate the abortion procedure in ways that are reasonably related to maternal

6. See *infra* Subsection IV.D.1.a.

7. See, e.g., *Williams v. Robinson*, 512 So. 2d 58 (Ala. 1987) (failure to diagnose ectopic pregnancy, leading to the utilization of an improper method of abortion); *Perguson v. Tamis*, 937 P.2d 347 (Ariz. Ct. App. 1996) (death of pregnant woman due to negligent performance of abortion); *Vuitch v. Furr*, 482 A.2d 811 (D.C. 1984) (improper treatment of laceration of uterine wall after abortion); *Shirk v. Kelsey*, 617 N.E.2d 152 (Ill. App. Ct. 1993) (incomplete abortion); *Ganapolskaya v. V.I.P. Med. Assocs.*, 644 N.Y.S.2d 735 (N.Y. App. Div. 1996) (perforated uterus caused by abortion procedure); see also Margaret E. Vroman, Annotation, *Medical Malpractice in Performance of Legal Abortion*, 69 A.L.R.4th 875 (1989) (analyzing various cases in which tort liability was imposed on physicians who negligently performed legal abortions).

8. But see *Collins v. Thakkar*, 552 N.E.2d 507, 509 (Ind. Ct. App. 1990), where the term “wrongful abortion” was used to describe a malicious and intentional termination of pregnancy without the expectant mother’s consent.

9. In this Article, the term “negligence” refers to ordinary negligence only and not to gross negligence, unless otherwise stated.

10. 410 U.S. 113 (1973).

health.

(c) For the stage subsequent to viability,¹¹ the State in promoting its interest in the potentiality of human life may, if it chooses, regulate, and even proscribe abortion except where it is necessary, in appropriate medical judgment, for the preservation of the life or health of the mother.¹²

In *Planned Parenthood v. Casey*, a plurality of the Court rejected the “rigid trimester framework,”¹³ but the Court affirmed the “essential holding” of *Roe*: a woman’s right to have an abortion prior to viability, states’ rights to regulate or proscribe abortions subsequent to viability in order to protect their interest in the potentiality of human life, and the states’ interests from the beginning of the pregnancy in protecting the health of the woman and the life of the fetus.¹⁴

Although *Roe* limited states’ power to regulate abortions, the statutory regulation of the subject in various states before this case was decided might shed some light on the circumstances in which wrongful abortion cases may arise. Until the late 1950s, most states proscribed any abortion (unless required to preserve the life of the mother).¹⁵ Yet prior to *Roe* and following the release of the Model Penal Code (drafted in 1962),¹⁶ several states had revised their abortion statutes to permit abortion in three types of cases: (1) where there was a substantial risk that continuance of the pregnancy would threaten the life or gravely impair the physical or mental health of the mother; (2) where there was a substantial risk that the child would be born with grave physical or mental defect; and (3) where the pregnancy resulted from rape, incest, or other felonious intercourse.¹⁷

Following *Roe*, all states must permit abortion even after viability in cases of the first type.¹⁸ Some sanction abortion after viability in one or two of the other

11. I.e., the stage when the fetus is capable of existing independently outside the mother’s womb. The Supreme Court held that viability occurs at twenty-four to twenty-eight weeks of gestation. *Id.* at 160.

12. *Id.* at 164-65.

13. *Planned Parenthood of Southeastern Pa. v. Casey*, 505 U.S. 833, 878 (1992).

14. *Id.* at 846.

15. Christopher P. Keleher, *Double Standards: The Suppression of Abortion Protesters’ Free Speech Rights*, 51 DEPAUL L. REV. 825, 834-35 (2002).

16. MODEL PENAL CODE § 230.3(2) (Proposed Official Draft 1962).

17. See Jeffrey L. Lenow, *The Fetus as a Patient: Emerging Rights as a Person?*, 9 AM. J.L. & MED. 1, 5 (1983); Tom Ginsburg, Book Note, 85 CAL. L. REV. 749, 755 (1997) (reviewing NEAL DEVINS, *SHAPING CONSTITUTIONAL VALUES* (1996)); see also, e.g., *Doe v. Bolton*, 410 U.S. 179, 202-05 (quoting the Georgia statute).

18. See, e.g., ALA. CODE § 26-22-3 (2004); ARIZ. REV. STAT. § 36-2301.01 (2004); CONN. GEN.

types of cases as well.¹⁹ Prior to viability, abortion is no longer limited to predetermined categories.²⁰ But even though cases of the three types mentioned above currently constitute a relatively small fraction of the various cases in which abortions are carried out in the United States,²¹ these three settings have a unique feature that makes them especially susceptible to the negligent instigation of abortion.

In all of them, the pregnant woman's decision regarding the continuance of her pregnancy is substantially dependent on *external* information provided by a qualified *professional* and is not a purely independent decision (as in other settings). If the woman receives inaccurate information, she may be induced to abort a fetus that she would otherwise wish to give birth to. If the inaccuracy stems from negligence, the abortion is "wrongful." Accordingly, we have detected at least two archetypal cases of wrongful abortion. In a case of the first type, a pregnant woman (with or without a companion) seeks medical counseling regarding the possible perils related to the continuance of her pregnancy. The adviser mistakenly maintains that the pregnancy is fraught with substantial risks for the woman, and she consequently decides to undergo an abortion. Later it is found that the information given by the adviser was wrong.²² In a case of the second type, the woman seeks advice concerning the health and bodily integrity of her fetus, and decides to undergo an abortion after being told that the fetus is deformed or disabled. Here, too, it is eventually realized that the information was wrong.²³ One may also consider a third type of case in which a woman is raped

STAT. § 19a-602 (2004); DEL. CODE ANN. tit. 11, § 651 (2004); FLA. STAT. ANN. § 390.0111 (West 2003); GA. CODE ANN. § 16-12-141 (2004); IND. CODE ANN. § 16-34-2-1 (2004); KAN. STAT. ANN. § 65-6703 (2003); LA. REV. STAT. ANN. § 14:87(B)(2) (West 2004).

19. See, e.g., ARK. CODE ANN. § 20-16-705 (2003) (allowing abortion of viable fetus in cases (1) and (3)); MISS. CODE ANN. § 97-3-3 (2004) (same); see also COLO. REV. STAT. § 18-6-101 (2003) (allowing abortion of viable fetus in cases (1) and (2)); UTAH CODE ANN. § 76-7-302(2), (3) (2004) (same). Lastly, see N.M. STAT. ANN. § 30-5-1 (Michie 2004) (allowing abortion of viable fetus in all three cases).

20. Interestingly, at least one state statute enumerates the aforementioned cases as factors that should be considered as justifying abortion before viability. See IDAHO CODE § 18-608 (Michie 2004).

21. See Nikki Katz, Abortion Statistics, <http://womensissues.about.com/cs/abortionstats/a/aaabortionstats.htm> (last visited Mar. 31, 2005).

22. *Baker v. Gordon*, 759 S.W.2d 87 (Mo. Ct. App. 1988). In that case, the doctor recommended an abortion in order to treat the mother's dysplasia. Later it was found that she had no dysplasia. The court held that the doctor was not negligent in recommending an abortion.

23. *Martinez v. Long Island Jewish Hillside Med. Ctr.*, 512 N.E.2d 538, 538 (N.Y. 1987). The facts of this case were outlined in the Introduction. See also *Johnson v. United States*, 810 F. Supp. 7 (D.D.C. 1993) (resolving a dispute where a pregnant woman was informed that she had HIV, and that consequently her baby would be born with AIDS; after she had an abortion it was discovered

shortly after having consensual sex with another and then discovers that she is pregnant.²⁴ After being told by her physician that her partner is not the father of the fetus, she chooses to terminate her pregnancy, only to discover that the physician was wrong. It seems to us that cases of this type would be rather rare, albeit not completely impossible.

We now attempt to further the understanding of our topic through contrast and analogy with other types of birth-related claims (BRC). One category of BRC consists of cases in which negligence by the defendant resulted in the birth of a healthy yet unwanted child. The negligence may manifest itself in the manufacture, provision, or installation of contraceptives;²⁵ in the performance of vasectomy²⁶ or tubal ligation;²⁷ or in the carrying out of an abortion.²⁸ These cases are frequently dealt with under the label of *wrongful pregnancy* (or *wrongful conception* in the appropriate cases).²⁹ In a way, they represent a mirror image of wrongful abortion cases, although they are not exact reflections. In

that she did not have HIV); *Breyne v. Potter*, 574 S.E.2d 916 (Ga. Ct. App. 2002) (resolving a dispute where a pregnant woman was informed that her fetus had Down's syndrome; after she had an abortion she discovered that the lab results were misinterpreted by her doctor); *Kupat Holim v. Dayan*, 55(1) P.D. 765 (1999) (Isr.) (resolving a dispute where a pregnant woman was told by her doctors that the fetus she was carrying was a male with severe bodily abnormalities, and urged her to terminate the pregnancy; after the abortion the woman saw that the fetus was a normal female).

24. We added a "shortly after having consensual sex" qualification because if the victim of rape were not involved in some kind of consensual sexual activity immediately before the occurrence of the crime, there would be no doubt with regard to the identity of the father. Theoretically, the identity of the father may also be uncertain where the victim of rape is involved in a consensual intercourse shortly *after* the rape. But this scenario seems to us unlikely given the psychological implications of rape.

25. *See, e.g.*, *Yasar v. Cohen*, 483 So. 2d 1099 (La. Ct. App. 1986) (inserting negligently an intrauterine device); *Troppi v. Scarf*, 187 N.W.2d 511 (Mich. Ct. App. 1971) (misfilling of a prescription for birth control pills); *Jackson v. Bumgardner*, 347 S.E.2d 743 (N.C. 1986) (failing to replace an intrauterine device after surgery); *J.P.M. v. Schmid Labs., Inc.*, 428 A.2d 515 (N.J. Super. Ct. App. Div. 1981) (manufacturing a defective condom).

26. *See, e.g.*, *Univ. of Ariz. Health Sci. Ctr. v. County of Maricopa Super. Ct.*, 667 P.2d 1294 (Ariz. 1983); *McFarlane v. Tayside Health Board*, [1999] 4 All E.R. 961 (Eng. H.L.).

27. *See, e.g.*, *Coleman v. Garrison*, 327 A.2d 757 (Del. Super. Ct. 1974), *aff'd*, 349 A.2d 8 (Del. 1975); *Flowers v. District of Columbia*, 478 A.2d 1073 (D.C. 1984); *Fulton-DeKalb Hosp. Auth. v. Graves*, 314 S.E.2d 653 (Ga. 1984); *Betancourt v. Gaylor*, 344 A.2d 336 (N.J. Super. Ct. Law. Div. 1975); *Mason v. W. Pa. Hosp.*, 453 A.2d 974 (Pa. 1982); *Smith v. Gore*, 728 S.W.2d 738 (Tenn. 1987); *C.S. v. Nielson*, 767 P.2d 504 (Utah 1988); *James G. v. Caserta*, 332 S.E.2d 872 (W. Va. 1985); *Beardsley v. Wierdsma*, 650 P.2d 288 (Wyo. 1982).

28. *See, e.g.*, *Stills v. Gratton*, 127 Cal. Rptr. 652 (Cal. Ct. App. 1976).

29. *See Anthony Jackson, Action for Wrongful Life, Wrongful Pregnancy, and Wrongful Birth in the United States and England*, 17 LOY. L.A. INT'L & COMP. L.J. 535, 583 (1995).

cases of wrongful pregnancy, practitioner negligence makes the fulfillment of the parents' will impossible, while in wrongful abortion cases practitioner negligence instigates, but does not necessitate, a decision that turns out to be inconsistent with such will.

Another category of BRC, more closely related to wrongful abortion, consists of cases in which a woman (with or without a companion) seeks medical advice regarding the health of her fetus; and decides to conceive or to continue her pregnancy once the adviser maintains that the child will not be born with congenital disabilities, a statement that is later found to be incorrect. The parents' cause of action for their resulting losses is labeled *wrongful birth*, while the infant's cause of action for his own losses is termed *wrongful life*.³⁰ Wrongful birth actions are recognized in most common law jurisdictions,³¹ although they are barred in some.³² On the other hand, most jurisdictions do not allow recovery on wrongful life theory.³³

30. For more elaborate discussions of these topics, see, for example, Alexander Morgan Capron, *Tort Liability in Genetic Counseling*, 79 COLUM. L. REV. 618 (1979); Constance Frisby Fain, *Wrongful Life: Legal and Medical Aspects*, 75 KY. L.J. 585 (1987); Jackson, *supra* note 29; Thomas Keasler Foutz, "*Wrongful Life*": *The Right Not To Be Born*, 54 TUL. L. REV. 480 (1980); Kurtis J. Kearl, *Turpin v. Sortini: Recognizing the Unsupportable Cause of Action for Wrongful Life*, 71 CAL. L. REV. 1278 (1983); Douglas Edward Peck, *Azzolino v. Dingfelder: North Carolina Court of Appeals Recognizes Wrongful Birth and Wrongful Life Claims*, 63 N.C. L. REV. 1329 (1985); Carel J.J.M. Stolker, *Wrongful Life: The Limits of Liability and Beyond*, 43 INT'L & COMP. L.Q. 521 (1994); Mark Strasser, *Wrongful Life, Wrongful Birth, Wrongful Death, and the Right To Refuse Treatment: Can Reasonable Jurisdictions Recognize All But One?*, 64 MO. L. REV. 29 (1999); Harvey Teff, *The Action for "Wrongful Life" in England and the United States*, 34 INT'L & COMP. L.Q. 423 (1985); Michael A. Berenson, Comment, *The Wrongful Life Claim—The Legal Dilemma of Existence Versus Nonexistence: "To Be or Not To Be,"* 64 TUL. L. REV. 895 (1990); Timothy J. Dawe, Note, *Wrongful Life: Time for a "Day in Court,"* 51 OHIO ST. L.J. 473 (1990); and James M. Parker, Comment, *Wrongful Life: The Child's Cause of Action for Negligent Genetic Counseling in Texas*, 16 ST. MARY'S L.J. 639 (1985).

31. *Berman v. Allan*, 404 A.2d 8 (N.J. 1979); *Karlsons v. Guerinot*, 394 N.Y.S.2d 933 (App. Div. 1977); *Schirmer v. Mt. Auburn Obstetrics & Gynecologic Assoc.*, 802 N.E.2d 723 (Ohio Ct. App. 2003); *Jacobs v. Theimer*, 519 S.W.2d 846 (Tex. 1975); *Dumer v. St. Michael's Hosp.*, 233 N.W.2d 372 (Wis. 1975); *C.A. 512/81 Zeitsov v. Katz*, 40(2) P.D. 85 (Isr.); Berenson, *supra* note 30, at 899 n.14. This cause of action is also recognized in continental jurisdictions. See, e.g., BGHZ 76, 249 (F.R.G.).

32. See, e.g., MINN. STAT. § 145.424(2) (2004); 42 PA. CONS. STAT. § 8305(a) (2004). In both statutes wrongful birth actions are barred only to the extent that they are based on a claim that but for the wrongful act or omission of the defendant, the child would have been aborted.

33. See, e.g., *Kush v. Lloyd*, 616 So. 2d 415 (Fla. 1992); *Siemieniec v. Lutheran Gen. Hosp.*, 512 N.E.2d 691 (Ill. 1987); *Kassama v. Magat*, 792 A.2d 1102 (Md. 2002); *Nelson v. Krusen*, 678 S.W.2d 918 (Tex. 1984). In several jurisdictions, wrongful life claims are prohibited (in whole or in

Wrongful birth is a more accurate mirror image of wrongful abortion. The former (just like wrongful pregnancy) deals with the *non-prevention* of the birth of an *unwanted* child, whereas the latter deals with the *prevention* of the birth of a *wanted* child. In both cases the defendant's negligence does not make the fulfillment of the parent's will physically impossible, but instigates a decision that turns out to be inconsistent with such will. This parallelism is useful in determining the expected legal response to wrongful abortions and in explaining why it is insufficient.

B. The Social Costs

The social costs of wrongful abortion generally consist of two components: parental losses and the loss of potential human life. These two ingredients overlap to a certain extent, but we have decided to discuss them separately for two reasons, which will become more obvious below. First, some of the parental loss is clearly independent of the loss of potential life and must be dealt with accordingly. In other words, despite a possible overlap between the two components, they have separate independent spheres that merit appropriate legal response. Recognizing this, we thought it would be somewhat artificial to discuss parental losses twice—once under the heading of “parental losses independent of the loss of the fetus,” and once within our discussion of the loss of potential life.

Second, and more importantly, while the law seems to respond quite effectively to parental losses, it leaves the loss of potential life (exclusive of any overlap with the parental loss) unaccounted for in most (and in certain jurisdictions—all) cases of wrongful abortion. The separation of these two components helps to clarify the exact problem that this Article attempts to address, i.e., the problem of unaccountability for the loss of potential life, which may result in under-deterrence of medical advisers and an overly lenient retribution for negligent violation of an important public interest. In this Section, we analyze the various elements of each component and explain where they overlap.

As regards parental losses, one may distinguish between pecuniary and non-pecuniary losses. Both include direct and relational losses. Direct parental losses are those incurred by the parents regardless of any harm done to their potential offspring. Relational losses are those incurred by a parent on account of the

part) by statute. In some, a person cannot base an action on the claim that but for the conduct of another he or she would have been aborted. *See, e.g.*, UTAH CODE ANN. § 78-11-24 (2004). Yet in others, a person cannot base an action on the claim that but for the conduct of another he or she would not have been conceived or, once conceived, would have been aborted. *See, e.g.*, S.D. CODIFIED LAWS § 21-55-1 (Michie 2003).

damage caused to the fetus or to the other parent.³⁴ We do not (and in fact cannot, as will be seen below) assume that the fetus may claim compensation for its “own” damage. At this stage we only use the damage caused to the fetus as a determinant for of the extent of the parental loss. The various types of parental losses are summarized in the following table.

| | <i>PECUNIARY LOSSES</i> | <i>NON-PECUNIARY LOSSES</i> |
|--------------------------|---|--|
| <i>DIRECT LOSSES</i> | <p>Costs of unnecessary abortion, and any medical treatment consequent upon the abortion.</p> <p>Lost wages during abortion and post-operational recovery.</p> | <p>Maternal pain and suffering during the surgical intervention and post-operational recovery.</p> <p>Chronic pain and suffering related to the abortion.³⁵</p> <p>Psychological reactions to the aforementioned pain.³⁶</p> |
| <i>RELATIONAL LOSSES</i> | <p>Loss of child’s contribution to parents from anticipated earning capacity.</p> <p>Loss of potential child’s services.³⁷</p> <p>Costs of medical/psychological treatment related to the loss of a child.</p> | <p>Mental anguish resulting from the unnecessary loss of a wanted child (maternal/paternal).</p> <p>Paternal grief over maternal pain.</p> <p>Loss of the child’s companionship and affection.³⁸</p> |

34. See Ronen Perry, *Relational Economic Loss: An Integrated Economic Justification for the Exclusionary Rule*, 56 RUTGERS L. REV. 711, 712 (2004) (defining relational economic loss).

35. An abortion may result in serious pelvic and abdominal pain. See, e.g., Genie M. Smith et al., *Pain of First Trimester Abortion: Its Quantification and Relations with Other Variables*, 133 AM. J. OBSTETRICS & GYNECOLOGY 489 (1979); Nancy Wells, *Pain and Distress During Abortion*, 12 HEALTH CARE FOR WOMEN INT’L 293 (1991).

36. See, e.g., E.A. Walker et al., *Dissociation in Women with Chronic Pelvic Pain*, 149 AM. J. PSYCHIATRY 534 (1992).

37. As will be seen below, childrearing costs and efforts usually exceed any possible financial benefit that a child may bestow upon her parents.

38. This loss may be mitigated if the parents can have the same number of children they initially planned regardless of the abortion.

The second component of the social costs of a wrongful abortion is the loss of potential human life, which has two aspects: (1) the loss of any net benefit the child herself would have had if she had not been aborted (“the fetal loss”), and (2) any detrimental effect the loss of potential life had on the well-being of others (“the relational loss”). The fetal loss includes a pecuniary element, i.e., the value of the assets that the child could acquire during her lifetime, and a non-pecuniary element, i.e., the value of her joy of life (taking into account the vicissitudes of life).

The relational loss includes the *net* pecuniary and non-pecuniary benefit that other people could have obtained if the fetus had not been aborted. It may comprise the loss of expected taxes minus expected social welfare (which is a state relational loss); the loss of expected profits from supplying groceries, housing, commodities, and services to the unborn; the loss of expected profit from employing the unborn; and the loss of the unborn child’s love, companionship, and other net non-pecuniary contributions to the world. The relational loss also includes mental anguish incurred by any person attributable to the loss of potential life and the outrage of the public at large. One can easily observe that the relational aspect of the loss of potential life overlaps (to a certain extent) the parental relational loss. We emphasize this here since any legal response to wrongful abortion must not take double account of a single loss.

II. THE LEGAL RESPONSE TO PARENTAL LOSSES

In this Part, we shall investigate whether and to what extent the law is responsive to the first ingredient of the social costs of wrongful abortions. At the outset, we assume that in cases of wrongful abortion all jurisdictions will recognize the parents’ cause of action in tort, at least if the abortion is legal. A wrongful abortion is a mere variant of medical malpractice, and as such entitles the patient to sue her (or his) physician. Only a few American courts have tackled the issue so far, and none of them denied the parents’ right of action.³⁹ We have restricted our assumption of liability to legal abortions, given that in several jurisdictions a woman who undergoes an illegal abortion may be deprived of any right of action for abortion-related losses.⁴⁰ Nonetheless, as we noted in Section I.A, an abortion can hardly ever be considered illegal in the archetypal cases of wrongful abortion.

39. See *supra* notes 22-23.

40. Gail D. Hollister, *Tort Suits for Injuries Sustained During Illegal Abortions: The Effects of Judicial Bias*, 45 VILL. L. REV. 387, 407-47 (2000) (analyzing court decisions that denied claims by women injured during the negligent performance of illegal abortions).

The more important question is what types of damages are compensable in a “wrongful abortion” tort action. As seen in Section I.B, parental losses are partly direct and partly relational. We shall discuss tort law’s expected response to parental losses using this dichotomy. Since there is no clear indication about the legal attitude to the various parental losses in wrongful abortion cases, we shall analogize from similar factual settings.

Direct pecuniary losses include the costs of the unnecessary abortion, plus any other medical treatment related to the abortion, and lost wages during the abortion and the post-operational recovery. Here we can analogize from the judicial treatment of wrongful pregnancy and wrongful birth actions. The costs of the unnecessary abortion and any consequent treatment are parallel to the medical costs related to continued pregnancy and delivery in wrongful pregnancy and wrongful birth cases. To the same extent that medical costs that were directly necessitated by the wrongful act are recoverable by the parents in wrongful pregnancy and wrongful birth actions,⁴¹ they should be recoverable in wrongful abortion actions brought by the parents.⁴² Similarly, as the mother in wrongful pregnancy and wrongful birth cases can recover for lost wages during pregnancy, delivery, and postnatal convalescence,⁴³ the mother in a wrongful abortion case should be allowed to recover for lost wages incurred during the abortion and the post-operational recovery (both physical and mental).

Direct non-pecuniary losses include maternal pain and suffering during the surgical intervention and post-operational recovery, chronic pain and suffering related to the abortion and any related treatment, and psychological reactions to the pain. Pain and suffering resulting directly from the negligent conduct, and their psychological aftermath are clearly recoverable.⁴⁴ Accordingly, in cases of wrongful pregnancy, courts allow the mother to recover for the pain and suffering related to the unwanted pregnancy, delivery, and postnatal recovery.⁴⁵

41. See, e.g., *Flowers v. District of Columbia*, 478 A.2d 1073, 1074 (D.C. 1984) (allowing parents to recover medical expenses incurred as a result of wife’s unwanted pregnancy); *Fulton-DeKalb Hosp. Auth. v. Graves*, 314 S.E.2d 653, 654 (Ga. 1984) (same); *Smith v. Gore*, 728 S.W.2d 738, 751-52 (Tenn. 1987) (same); *C.S. v. Nielson*, 767 P.2d 504, 509-10 (Utah 1988) (same); *Beardsley v. Wierdsma*, 650 P.2d 288, 292 (Wyo. 1982) (same).

42. They should be recoverable at least to the extent that they exceed the medical expenses that were saved on account of the abortion.

43. See, e.g., *Flowers*, 478 A.2d at 1074 (finding lost wages during pregnancy and recovery from delivery to be recoverable); *Graves*, 314 S.E.2d at 654 (same); *Smith*, 728 S.W.2d at 751 (same); *C.S.*, 767 P.2d at 510 (same); *James G. v. Caserta*, 332 S.E.2d 872, 877 (W. Va. 1985) (same); *Beardsley*, 650 P.2d at 292 (same).

44. 2 DAN B. DOBBS, *LAW OF REMEDIES* § 8.1(4) (2d ed. 1993).

45. See, e.g., *Flowers*, 478 A.2d at 1074 (compensating mother for pain, suffering, and discomfort resulting from unwanted pregnancy and delivery); *Graves*, 314 S.E.2d at 654 (same);

Similarly, the mother should recover for the pain and suffering and consequent psychological effects related to the unwanted surgical intervention, and following recovery in cases of wrongful abortion.

We now turn to parental relational losses. According to the traditional common law view, the death of one human being could not give rise to a cause of action for the benefit of others.⁴⁶ Moreover, the common law did not allow recovery for relational losses even when they were not related to death (subject to a few narrowly defined exceptions).⁴⁷ However, both principles were superseded to a limited extent by wrongful death legislation, which entitles certain enumerated relatives of victims of fatal injuries—including their parents—to seek legal redress.⁴⁸ The inadequacy of this legislation as a legal response to wrongful abortions will be discussed thoroughly below. At this stage suffice it to say that in many states a wrongful death action is never available in cases of wrongful abortion; that in nearly all other states it is not available in the vast majority of wrongful abortions cases; and that in any case a wrongful death action does not normally cover the mental anguish related to the loss of a prospective child.

It may well be that the unavailability of a wrongful death cause of action in cases of wrongful abortion does not have a serious impact on the parents' economic well-being. Damages under wrongful death legislation typically include lost financial support and lost services. Theoretically, the parents of an unborn child may also lose her support and services. However, since in nearly all cases childrearing costs and efforts significantly exceed any possible financial benefit that the child may bestow upon her parents, it is highly unlikely that

Smith, 728 S.W.2d at 751 (same); *C.S.*, 767 P.2d at 510 (same); *James G.*, 332 S.E.2d at 877 (same); *Beardsley*, 650 P.2d at 292 (same).

46. *Baker v. Bolton*, 170 Eng. Rep. 1033 (1808). This rule was also followed in the United States. *See, e.g.*, *The Harrisburg*, 119 U.S. 199, 213 (1886); *Ins. Co. v. Brame*, 95 U.S. 754, 757 (1877). Note, however, that in 1970 the Supreme Court recognized a limited exception to the old rule, which was supposed to fill in an obvious gap in state and federal legislation. *Moragne v. States Marine Lines Inc.*, 398 U.S. 375, 402-03, 408-09 (1969).

47. Liability for relational economic loss was and is generally excluded in the United States. *See Perry, supra* note 34, at 725-26. Liability for relational emotional harm was once excluded, *see, e.g.*, *S. Ry. Co. v. Jackson*, 91 S.E. 28, 28 (1916), and is extremely exceptional even today, *see Dale Joseph Gilsinger, Annotation, Recovery Under State Law for Negligent Infliction of Emotional Distress Under Rule of Dillon v. Legg*, 68 Cal. 2d 728, 69 Cal. Rptr. 72, 441 P.2d 912 (1968), or *Refinements Thereof*, 96 A.L.R.5th 107 (2002); *Dale Joseph Gilsinger, Annotation, Relationship Between Victim and Plaintiff-Witness as Affecting Right To Recover Under State Law for Negligent Infliction of Emotional Distress Due to Witnessing Injury to Another Where Bystander Plaintiff Is Not Member of Victim's Immediate Family*, 98 A.L.R.5th 609 (2002).

48. *See infra* Section III.B.

parental pecuniary relational loss in fact exists.⁴⁹

Non-pecuniary losses pose more acute problems. First, the parents lose their potential child's companionship and affection. Usually, this non-pecuniary benefit outweighs any non-pecuniary detriment related to parenthood.⁵⁰ Many jurisdictions allow recovery for the net-benefit under wrongful death legislation where the victim is a minor child. Where wrongful death actions are not available in cases of wrongful abortion, and where they are available but do not cover non-pecuniary losses, a significant loss thus remains uncompensated.

Second, losing a wanted child may quite clearly result in severe emotional distress to both parents. An abortion may lead to negative psychological reactions even if it was not induced by a negligent misrepresentation (and was not performed negligently).⁵¹ Conceivably, such reactions would be intensified whenever it became clear that the abortion was induced by a negligent misrepresentation, i.e., the abortion was undergone by mistake.⁵²

We have already said that wrongful death statutes are not applicable to wrongful abortion cases in several jurisdictions, and that in other jurisdictions these statutes do not apply to the vast majority of wrongful abortion cases. But even where these statutes apply, they do not always allow recovery for the mental grief consequent on the loss of a child (or any other relative). In some states, the wrongful death statutes specifically provide for recovery for the mental anguish of a few close relatives,⁵³ while the courts in other states have allowed such damages as a matter of judicial interpretation of the relevant statute.⁵⁴ Yet in

49. Andrew Jay McClurg, *It's a Wonderful Life: The Case for Hedonic Damages in Wrongful Death Cases*, 66 NOTRE DAME L. REV. 57, 59, 64 (1990); see also *infra* notes 117-118 and accompanying text.

50. See, however, *supra* note 38, regarding the possibility of loss mitigation.

51. See E. Joanne Angelo, *Psychiatric Sequelae of Abortion: The Many Faces of Post-Abortion Grief*, 59 LINACRE Q. 69 (1992); Aliza Kolker & B. Meredith Burke, *Grieving the Wanted Child: Ramifications of Abortion After Parental Diagnosis of Abnormality*, 14 HEALTH CARE FOR WOMEN INT'L 513, 516-23 (1993); Terry Nicole Steinberg, Note, *Abortion Counseling: To Benefit Maternal Health*, 15 AM. J. L. & MED. 483, 488-92 (1989).

52. We found no reference to support this intuition, but it seems to us self-evident.

53. ARK. CODE ANN. § 16-62-102(f) (Michie 2003); COLO. REV. STAT. § 13-21-203(1)(a) (2003); DEL. CODE ANN. tit. 10, § 3724(d)(5) (2004); FLA. STAT. ANN. § 768.21(2)-(3) (2003); KAN. STAT. ANN. § 60-1904(a)(1) (2003); ME. REV. STAT. tit. 18-A, § 2-804(B) (West 2003); MD. CODE ANN., CTS. & JUD. PROC. § 3-904(d) (2003); NEV. REV. STAT. § 41.085(4) (2004); OHIO REV. CODE ANN. § 2125.02(B)(5) (West 2004); OKLA. STAT. ANN. tit. 12, § 1053(B) (2004); VA. CODE ANN. § 8.01-52(1) (Michie 2004); W. VA. CODE § 55-7-6(c)(1) (2003).

54. See, e.g., *Tommy's Elbow Room, Inc. v. Kavorkian*, 727 P.2d 1038, 1047-48 (Alaska 1986); *Mullen v. Posada Del Sol Health Care Ctr.*, 819 P.2d 985, 986 (Ariz. Ct. App. 1991); *Hopkins v. McBane*, 427 N.W.2d 85, 94 (N.D. 1988); *Sanchez v. Schindler*, 651 S.W.2d 249, 251

other jurisdictions, damages for mental anguish may not be recovered in wrongful death actions.⁵⁵ Surprisingly, this does not necessarily leave the parents without redress for their loss.

A remedy may be found in the common law of torts. We are familiar with the traditional reluctance to allow recovery for emotional distress, especially where such loss is consequent on an injury to another.⁵⁶ Nevertheless, it appears that in factual settings similar to those discussed here, courts have been more willing to permit recovery. For example, in cases of wrongful pregnancy, a few courts have awarded the parents damages for the mental anguish related to the upbringing of an unwanted child.⁵⁷ This, of course, is a highly controversial head of damages, given the common belief that the joy associated with nurturing a normal child usually outweighs the sorrows.⁵⁸ In wrongful birth cases, parents are sometimes allowed to recover for mental anguish suffered on account of their child's condition.⁵⁹

More relevant in the current context are cases of prenatal injuries resulting in stillbirth. In such cases it can hardly be said that the ensuing parental grief is overshadowed by some kind of joy (i.e., the avoidance of distress related to childrearing). In many jurisdictions, it is well established that where the *mother* was *physically injured*, and thereby lost her fetus, she can recover for mental anguish resulting from the death of the unborn child.⁶⁰ Moreover, a trend seems to exist toward abandoning the requirement that the mother suffer physical injury other than the injuries sustained by the fetus. For example, the New York Court

(Tex. 1983); *Hartnett v. Union Mut. Fire Ins. Co.*, 569 A.2d 486, 488 (Vt. 1989); *Wilson v. Lund*, 491 P.2d 1287, 1290 (Wash. 1971).

55. *See, e.g.*, *Krouse v. Graham*, 562 P.2d 1022, 1028 (Cal. 1977); *OB-GYN Assocs. of Albany v. Littleton*, 386 S.E.2d 146, 147 (Ga. 1989); *Bullard v. Barnes*, 468 N.E.2d 1228, 1232-33 (Ill. 1984); *Wardlow v. City of Keokuk*, 190 N.W.2d 439, 448 (Iowa 1971); *MacCuish v. Volkswagenwerk, A.G.*, 494 N.E.2d 390, 401 (Mass. App. Ct. 1986), *aff'd*, 400 Mass. 1003 (1987); *Williams v. Monarch Transp., Inc.*, 470 N.W.2d 751, 756 (Neb. 1991); *Small v. McKennan Hosp.*, 437 N.W.2d 194, 204 (S.D. 1989).

56. *See supra* note 47.

57. *Betancourt v. Gaylor*, 344 A.2d 336, 340 (N.J. Super. Ct. Law Div. 1975).

58. *See Mason v. Western Pa. Hosp.*, 453 A.2d 974, 976 (Pa. 1982).

59. *Berman v. Allan*, 404 A.2d 8, 15 (N.J. 1979) (holding that parents are entitled to be recompensed for the mental and emotional anguish they have suffered and will continue to suffer on account of their child's condition); *Schroeder v. Perkel*, 432 A.2d 834, 838-39 (N.J. 1981) (same); *Speck v. Finegold*, 439 A.2d 110, 112-15 (Pa. 1981) (same).

60. *See, e.g.*, *Snow v. Allen*, 151 So. 468, 470-71 (Ala. 1933); *Thomas v. Carter*, 506 S.E.2d 377, 379-80 (Ga. Ct. App. 1998); *Smith v. Borello*, 804 A.2d 1151, 1162-63 (Md. 2002); *Occhipinti v. Rheem Mfg. Co.*, 172 So.2d 186, 190 (Miss. 1965); *Graf v. Taggart*, 204 A.2d 140, 146 (N.J. 1964); *Endresz v. Friedberg*, 24 N.Y.2d 478, 484 (1969).

of Appeals very recently held:

Although, in treating a pregnancy, medical professionals owe a duty of care to the developing fetus . . . they surely owe a duty of reasonable care to the expectant mother, who is, after all, the patient. Because the health of the mother and fetus are linked, we will not force them into legalistic pigeonholes.

We therefore hold that, even in the absence of an independent injury, medical malpractice resulting in miscarriage or stillbirth should be construed as a violation of a duty of care to the expectant mother, entitling her to damages for emotional distress.⁶¹

The Court of Appeals held, however, that the physician owed no duty of care to the expecting father.⁶² On the other hand, it was held in several states that medical malpractice causing an infant stillbirth constitutes a tort against both parents and that they may recover compensatory damages for their emotional distress and mental suffering.⁶³ The Texas Court of Appeals stated that there is

[N]o compelling state interest in a gender-based denial of a father's right to recover damages for his own mental anguish from the negligently caused loss of his viable fetus, a denial which "perpetuates the myth that only a woman grieves and suffers the mental anguish caused by the loss of a baby in the womb."⁶⁴

Finally, it should be noted that in the very few wrongful abortion cases tried so far, courts recognized the mother's cause of action for her mental grief.⁶⁵

61. *Broadnax v. Gonzalez*, 809 N.E.2d 645, 648-49 (N.Y. 2004) (citations and footnotes omitted).

62. *Id.* at 649 n.3.

63. *See, e.g.*, *Tanner v. Hartog*, 696 So. 2d 705, 708 (Fla. 1997); *Giardina v. Bennett*, 545 A.2d 139, 139, 143 (N.J. 1988); *Johnson v. Ruark Obstetrics & Gynecology Assocs.*, 395 S.E.2d 85, 86, 98 (N.C. 1990).

64. *Parvin v. Dean*, 7 S.W.3d 264, 279 (Tex. Ct. App. 1999) (quoting *Krishnan v. Sepulveda*, 916 S.W.2d 478, 483 (Tex. 1995) (Gonzales, J., dissenting)). *But cf.* *Sosebee v. Hillcrest Baptist Med. Ctr.*, 8 S.W.3d 427, 436 (Tex. Ct. App. 1999) (rejecting gender-discrimination argument following *Krishnan*, 916 S.W.2d at 482).

65. *Marie v. McGreevey*, 314 F.3d 136, 140 (3d Cir. 2002); *Martinez v. Long Island Jewish Hillside Med. Ctr.*, 512 N.E.2d 538, 538-39 (N.Y. 1987). In *Martinez*, the court held that the plaintiff's claim for emotional distress, derived from agreeing to an act that was contrary to her religious beliefs, was actionable. But it was later interpreted as stating that "the breach of duty owed directly to plaintiff leading to her emotional distress is plainly compensable." *Ferrara v. Bernstein*, 81 N.Y.2d 895, 898 (1993). In the Israeli case of *Dayan*, the court held that both parents could recover for their direct non-pecuniary loss. *Kupat Holim v. Dayan*, 55(1) P.D. 765 (1999).

In conclusion, the legal response to parental losses seems reasonable, although somewhat imperfect.⁶⁶ As we shall contend below, such is not the case with the other ingredient of the social costs of wrongful abortion.

III. THE LEGAL PROTECTION OF THE PUBLIC INTEREST IN PRESERVING POTENTIAL LIFE

A. Legal Recognition of the Public Interest

In this Section, we explore whether and to what extent the law is responsive to the second ingredient of the social costs of wrongful abortions—the loss of a potential life. The state’s “important and legitimate interest in preserving and protecting . . . the potentiality of human life” was recognized by the Supreme Court in *Roe v. Wade*.⁶⁷ In *Planned Parenthood v. Casey*, the Court explained that this interest merits protection from the very beginning of the pregnancy, that is, from the moment of fertilization.⁶⁸ Yet the *Roe* and *Casey* Courts sought to balance such interest (together with the state’s interest in protecting the pregnant woman’s health) against the pregnant woman’s right to privacy. They concluded that viability marks the earliest point at which the state’s interest in fetal life is constitutionally adequate to justify a legislative ban on non-therapeutic abortions.⁶⁹ These cases demonstrate that if there were no legally recognized public interest in the protection of fetuses, there would be no legal justification for allowing state regulation of abortions, subject perhaps to the need to guarantee adequate conditions for their performance. It is also clear that the public interest in preserving potential life may be fully protected where this protection does not violate the mother’s constitutional right to privacy.⁷⁰ That is

66. The parents may not be compensated for the loss of love and society of the unborn child.

67. 410 U.S. 113, 162 (1972).

68. *Planned Parenthood of Southeastern Pa. v. Casey*, 505 U.S. 833, 846 (1992).

69. *Casey*, 505 U.S. at 846. We do not accept the interpretation of the Supreme Court of Arkansas that “the state’s interest in protecting the life of a fetus begins at viability.” *Aka v. Jefferson Hosp. Ass’n*, 42 S.W.3d 508, 517-18 (Ark. 2001). Viability only determines the turning point with regard to the balance between the state interest and the right to privacy.

70. Cf. Cari L. Leventhal, *The Crimes Against the Unborn Child Act: Recognizing Potential Human Life in Pennsylvania Criminal Law*, 103 DICK. L. REV. 173, 185-90 (1998) (arguing that the Supreme Court’s decision in *Roe* forbids the state’s protection of the unborn’s interest only when it conflicts with the protected interest of the mother); Mamta K. Shah, *Inconsistencies in the Legal Status of an Unborn Child: Recognition of a Fetus as Potential Life*, 29 HOFSTRA L. REV. 931, 966 (2001). This perception is manifested, for example, in ARK. CONST. amend. LXVIII, § 2, which declares that “[t]he policy of Arkansas is to protect the life of every unborn child from conception until birth, to the extent permitted by the Federal Constitution.”

exactly why the states can prohibit the killing of a fetus, at any stage after conception, outside the realm of legal consensual abortions.⁷¹ The same reasoning must be applied in wrongful abortion settings. In cases of this type, there is no need to balance the societal interest against the private right since they coincide. The woman would very likely have chosen not to abort had she been given accurate information. The inaccuracy of the information given by the adviser induced her to make a decision that she would otherwise not have made.⁷² Imposing a legal sanction on the negligent adviser is thus compatible with the public's interest in protecting potential life *and* the mother's right to privacy.

A legal sanction is thus required to ensure that professional advisers have an adequate incentive to abstain from giving inaccurate information that may result in loss of potential life, and that negligent advisers receive their just desert (in the retributive sense). In the following Sections, we attempt to show that the law does not currently provide such a sanction. Note that our concern is with the inadequate protection of the public interest in preserving potential life, and not with the possible failure of the law in vindicating private interests. Therefore, our analysis of private rights of action in the next Section is merely an attempt to discern whether existing tort law might ensure adequate protection of the *public* interest. Although one may conclude from our analysis that tort law does not afford adequate *compensation* to interested parties,⁷³ the lack of coverage for some of the social costs of wrongful abortion is only relevant in this Part to the extent that it may result in under-deterrence or overly lenient retribution.⁷⁴

B. Protection of the Public Interest in Civil Law

We shall assume (although it is not always the case)⁷⁵ that a fetus cannot

71. See *infra* notes 136-138 and accompanying text.

72. Cf. Daniel S. Meade, *Wrongful Death and the Unborn Child: Should Viability Be a Prerequisite for a Cause of Action?*, 14 J. CONTEMP. HEALTH L. & POL'Y 421, 445 (1998) ("If someone causes the death of a woman's unborn child . . . that person has infringed upon the woman's choice to have a baby, thus violating her right to privacy.").

73. This conclusion may become problematic if one argues that the *fetus* (through its representative) is also "under-compensated." A conclusion of this type is founded on the assumption that a fetus is a legal person (because non-persons have no right to compensation). As indicated in the Introduction, we prefer to avoid such controversial assumptions. This point will also be discussed in Subsection IV.D.1.

74. We assume that under-compensation may result in under-deterrence. This point will be discussed in detail in Section III.E.

75. In the Israeli *Dayan* case, the child lived for fifteen minutes after abortion. *Kupat Holim v. Dayan*, 55(1) P.D. 765 (1999). But this is a peculiar case.

survive an abortion. Abortion terminates the potentiality of life. Whenever a “person” dies most state legislatures recognize two types of civil actions:⁷⁶ (1) an action brought by the personal representative of the deceased for the loss suffered by the latter prior to her death (a “survival action”). The proceeds in this action are recovered for the benefit of the estate and subsequently distributed among the heirs; and (2) an action brought for the benefit of the deceased’s dependants (i.e., certain statutorily enumerated relatives) with regard to their own loss (a “wrongful death action”). The proceeds of this action accrue directly to the dependants.⁷⁷

In cases where a fetus injured *in utero* survives delivery and dies shortly

76. ALA. CODE §§ 6-5-391, -410, -462 (2004); ALASKA STAT. §§ 09.55.570, .580 (2004); ARIZ. REV. STAT. §§ 12-611 to -613 (2004); ARK. CODE ANN. §§ 16-62-101, -102 (Michie 2003); CAL. CIV. PROC. CODE §§ 377.20, .30, .34, .60-.62 (West 2004); COLO. REV. STAT. §§ 13-20-101, -21-201 to -203 (2003); CONN. GEN. STAT. §§ 52-555, -599 (2003); DEL. CODE ANN. tit. 10, §§ 3701, 3707, 3721-3725 (2004); D.C. CODE ANN. §§ 12-101, 16-2701 to -2703 (2004); FLA. STAT. chs. 46.021, 768.16-.26 (2004); GA. CODE ANN. §§ 9-2-41, 51-4-1 to -5 (2004); HAW. REV. STAT. §§ 663-3 to -8 (2003); IDAHO CODE §§ 5-311, -319 (Michie 2004); 740 ILL. COMP. STAT. 180/2 (2004); 755 ILL. COMP. STAT. ANN. 5/27-6 (2004); IND. CODE §§ 34-9-3-1, -11-7-1, -23-1-1, -23-1-2, -23-2-1 (2004); IOWA CODE §§ 611.20, 633.336 (2003); KAN. STAT. ANN. §§ 60-1801, -1901 to -1905 (2003); KY. REV. STAT. ANN. §§ 411.130, 411.140 (Banks-Baldwin 2004); LA. CIV. CODE ANN. arts. 2315.1-.2 (West 2004); ME. REV. STAT. tit 18-A, §§ 2-804, 3-817 (West 2004); MD. CODE ANN., CTS. & JUD. PROC. §§ 3-901 to -904, 6-401 (2003); MASS. GEN. LAWS ANN. ch. 228, § 1, ch. 229, §§ 2, 6 (West 2004); MICH. COMP. LAWS §§ 600.2921-2922 (2004); MINN. STAT. § 573.02 (2003); MISS. CODE ANN. §§ 11-7-13, 91-7-233 (2004); MO. REV. STAT. §§ 537.020, .080, .090 (2004); MONT. CODE ANN. §§ 27-1-323, -501, -513 (2003); NEB. REV. STAT. §§ 25-1401, 30-809, -810 (2003); NEV. REV. STAT. §§ 41.085, .100 (2004); N.H. REV. STAT. ANN. §§ 556:7, :9, :12 to :14 (2003); N.J. STAT. ANN. §§ 2A:15-3, :31-1 to -6 (2004); N.M. STAT. ANN. §§ 37-2-1, 41-2-1 to -4 (Michie 2004); N.Y. EST. POWERS & TRUSTS LAW §§ 5-4.1, 11-3.2 (McKinney 2004); N.C. GEN. STAT. §§ 28A-18-1 to -2 (2004); N.D. CENT. CODE §§ 28-01-26.1, 32-21 (2003); OHIO REV. CODE ANN. §§ 2125.01-.02, 2305.21 (West 2004); OKLA. STAT. tit. 12, §§ 1051-1055 (2004); OR. REV. STAT. §§ 30.020, 30.075 (2003); 42 PA. CONS. STAT. §§ 8301-8302 (2004); R.I. GEN. LAWS §§ 9-1-6, 10-7-1 to -13 (2004); S.C. CODE ANN. §§ 15-51-10 to -20, -40 to -42, -60 (West 2003); S.D. CODIFIED LAWS §§ 15-4-1, -2, 21-5-1 to -9 (Michie 2003); TENN. CODE ANN. §§ 20-5-101 to -113 (2004); TEX. CIV. PRAC. & REM. CODE ANN. §§ 71.001 to -.011, -.021 (Vernon 2004); UTAH CODE ANN. §§ 78-11-6, -7, -12 (2004); VT. STAT. ANN. tit. 14, §§ 1452-53, 1491-1492 (2003); VA. CODE ANN. §§ 8.01-25, 8.01-50 to -56 (2004); WASH. REV. CODE §§ 4.20.005, .010, .020, .046, .050, .060 (2004); W. VA. CODE §§ 55-7-5 to -8a (2003); WIS. STAT. §§ 895.01, .03, .04 (2003). These two types of actions are also available in other common law jurisdictions. *See, e.g.*, Fatal Injuries Act, R.S.N.S., ch. 163 (1989) (Can.); Survival of Actions Act, R.S.N.S., ch. 453 (1989) (Can.).

77. Such actions are usually brought by the personal representative of the deceased. *See, e.g.*, *Brewer v. Lacefield*, 784 S.W.2d 156 (Ark. 1990) (finding that the personal representative in bringing suit for wrongful death acts only as a trustee or conduit, and any proceeds recovered are for the benefit of the beneficiaries and not for the estate).

thereafter due to its injuries, all jurisdictions that have considered the issue allow both actions against the injurer, given that live birth turns the fetus into a legal person.⁷⁸ The question that must be addressed here is twofold. First, are these actions also available in cases of wrongful abortion under the aforementioned assumption (i.e., a fetus cannot survive an abortion)? Second, do these actions (where available) impose a sanction that is roughly equivalent to the loss of potential life?

1. Availability of Actions

Until 1949, no jurisdiction permitted wrongful death proceedings for a stillborn infant.⁷⁹ The unborn child was not regarded as an independent legal person whose "death" may give rise to a wrongful death action, mainly because a fetus was not thought to be an independent biological entity.⁸⁰ Since the decision of the Supreme Court of Minnesota in *Verkennes v. Corniea*,⁸¹ however, there seems to be a trend in favor of allowing such claims via statutory revision or judicial interpretation, subject to certain restrictions that will be discussed below.⁸² In nearly all jurisdictions where the expansion originated in judicial interpretation, the interpretative conclusions regarding the wrongful death statute were applied or are equally applicable to the survival statute.⁸³

In spite of this development, it cannot be said that survival and wrongful

78. See, e.g., *Wolfe v. Isbell*, 280 So. 2d 758 (Ala. 1973); *Callahan v. Slavsky*, 385 P.2d 674 (Colo. 1963); *Amann v. Faigy*, 114 N.E.2d 412 (Ill. 1953); *Torigian v. Watertown News Co.*, 225 N.E.2d 926 (Mass. 1967); *Stegall v. Morris*, 258 S.W.2d 577 (Mo. 1953); *Jasinsky v. Potts*, 92 N.E.2d 809 (Ohio 1950); *Hall v. Murphy*, 113 S.E.2d 790 (S.C. 1960); *Shousha v. Matthews Drivurself Serv., Inc.*, 358 S.W.2d 471 (Tenn. 1962); *Leal v. C.C. Pitts Sand & Gravel, Inc.*, 419 S.W.2d 820 (Tex. 1967); *Kalafut v. Gruver*, 389 S.E.2d 681 (Va. 1990); see also RESTATEMENT (SECOND) OF TORTS § 869(1) (1977); Timothy Stoltzfus Jost, *Rights of Embryo and Fetus in Private Law*, 50 AM. J. COMP. L. 633, 642 (2002). Note, however, that this is a relatively recent development in the common law of torts. Until 1946, American courts did not allow recovery for prenatal injuries, even in cases of live birth. *Bonbrest v. Kotz*, 65 F. Supp. 138, 140 (D.D.C. 1946), marks the turning point.

79. Robert J. Del Tufo, *Recovery for Prenatal Torts: Actions for Wrongful Death*, 15 RUTGERS L. REV. 61, 73 (1960).

80. Jost, *supra* note 78, at 642; Shah, *supra* note 70, at 934.

81. 38 N.W.2d 838 (Minn. 1949).

82. See *infra* note 98.

83. See, e.g., *Gorke v. Le Clerc*, 181 A.2d 448, 451 (Conn. Super. Ct. 1962); *Greater Southeast Cmty. Hosp. v. Williams*, 482 A.2d 394, 398 (D.C. 1984); *Amadio v. Levin*, 501 A.2d 1085, 1087 (Pa. 1985); *Cavazos v. Franklin*, 867 P.2d 674, 676-77 (Wash. Ct. App. 1994). The survival statute and the wrongful death statute are usually interwoven, so their similar interpretation is not surprising.

death actions are generally applicable to wrongful abortion cases. On the contrary, in most jurisdictions at least one of them—and usually both—will not be available in all or most cases of wrongful abortion. There are a few noteworthy reasons for this. First, although the *majority* of jurisdictions allow wrongful death actions in cases of fetal death, others (among them California, New York, and Florida) adhere to the traditional view that survival and wrongful death actions are not maintainable unless the child is born alive.⁸⁴ In those jurisdictions, a fetus not yet born is not a legal “person,” and therefore no “personal” cause of action can survive its “death.” Additionally, a wrongful death action is considered to be a derivative action, which cannot be brought unless the immediate victim had a cause of action prior to his or her death. Clearly, then, in these jurisdictions wrongful death acts do not permit recovery attributable to the wrongful “death” of a fetus before birth.⁸⁵ An attempt to challenge the constitutionality of this viewpoint failed. The Third Circuit held that wrongful death statutes that discriminate against mothers of fetuses that do not survive birth, including aborted fetuses, do not violate the Equal Protection Clause.⁸⁶ It should be noted that the traditional view is also adhered to in common law jurisdictions outside the United States.⁸⁷

Second, in at least two jurisdictions that generally allow wrongful death actions in cases of fetal death (Arkansas and Illinois), a medical adviser cannot be held liable for the wrongful abortion of a fetus. This is because the wrongful death statute mandates either that: (1) *no person* shall be liable under the statute when the death of the fetus results from a legal abortion;⁸⁸ or (2) there shall be no cause of action against a physician or a medical institution for the wrongful death of a fetus caused by a legal and consented abortion.⁸⁹

84. *Justus v. Atchison*, 565 P.2d 122, 131 (Cal. 1977); *Tanner v. Hartog*, 696 So. 2d 705, 707 (Fla. 1997); *Tebbutt v. Virostek*, 483 N.E.2d 1142 (N.Y. 1985); *see also* *Bolin v. Wingert*, 764 N.E.2d 201, 207 (Ind. 2002); *Shaw v. Jendzejec*, 717 A.2d 367, 368 (Me. 1998); *Giardina v. Bennett*, 545 A.2d 139, 139, 147 (N.J. 1988); *Nelson v. Peterson*, 542 P.2d 1075, 1077 (Utah 1975); *Lawrence v. Craven Tire Co.*, 169 S.E.2d 440 (Va. 1969). Texas and Nebraska very recently left this group. *See infra* notes 95-97 and accompanying text.

85. *See, e.g., Marie v. McGreevey*, 314 F.3d 136, 140 (3d Cir. 2002); *Giardina*, 545 A.2d at 143-44; *Graf v. Taggart*, 204 A.2d 140 (N.J. 1964).

86. *Marie*, 314 F.3d at 141-42; *see also* *Alexander v. Whitman*, 114 F.3d 1392, 1400 (3d Cir. 1997).

87. *See* JOHN SEYMOUR, *CHILDBIRTH AND THE LAW* 103, 119 (2000) (noting that in England and Canada, the courts have indicated that they will not depart from the traditional view, and in Australia, although the issue has not arisen, the courts would be likely to adopt the same view).

88. ARK. CODE ANN. § 16-62-102(a)(3) (Michie 2003).

89. 740 ILL. COMP. STAT. ANN. 180/2.2 (2004). It was held that in cases of legal abortion no liability can be imposed, not only on the physician who performed the abortion but on any other

Likewise, in Texas the wrongful death statute excludes imposition of liability for fetal death upon “a physician or other health care provider licensed in this state, if the death directly or indirectly is caused by, associated with, arises out of, or relates to a lawful medical or health care practice or procedure of the physician or the health care provider.”⁹⁰ Given that the wrongful death statute includes another provision exempting physicians who perform lawful abortions,⁹¹ it is arguable that the Texas legislature intended to exempt from liability not only physicians who perform abortions, but also any physician or health care provider whose conduct resulted (directly or indirectly) in fetal death.⁹² Consequently, it seems that the above-cited paragraph applies to negligent medical advisers in wrongful abortion cases.

The underlying rationale of these broad exemptions is not clear to us. We believe that exemption from liability for an intentional or negligent causation of harm may be granted only where there is legal justification for the harmful conduct that annuls its *prima facie* wrongfulness. We thus find a limited exemption of any physician who performs a consented legal abortion (as exists in Michigan and Nebraska) far more understandable.⁹³ Nonetheless, the fact remains that wrongful abortion is not actionable in a few jurisdictions that generally allow recovery in cases of fetal death.

Third, several jurisdictions differentiate between wrongful death and survival actions. In cases of fetal death, they allow the former but exclude the latter. In Louisiana and Iowa, this distinction was explicitly proclaimed by the judiciary.⁹⁴ In Nebraska and Texas, the legislature very recently amended the

person who proximately caused the abortion. *Light v. Proctor Cmty. Hosp.*, 538 N.E.2d 828, 829 (Ill. App. Ct. 1989).

90. TEX. CIV. PRAC. & REM. CODE ANN. § 71.003(c)(4) (Vernon 2004).

91. *Id.* § 71.003(c)(2).

92. However, there is no case law on the subject.

93. MICH. COMP. LAWS § 600.2922a(2)(b) (2004) (person who performs a consented abortion is not liable); NEB. REV. STAT. § 30-809(2)(b) (2003) (“No action for damages for the death of a person who is an unborn child shall be brought under this section against . . . [a] physician or other licensed health care provider if the death was the intended result of a medical procedure performed by the physician or health care provider and the requisite consent was given . . .”). As noted above, a similar provision exists in Texas, but it is accompanied by a general and very broad exemption for physicians who cause fetal death.

94. *See, e.g., Wartelle v. Women’s & Children’s Hosp.*, 704 So. 2d 778, 781 (La. 1997) (holding that an unborn child is a person for the purposes of wrongful death statute but not for the purposes of bringing a survival action); *Dunn v. Rose Way, Inc.*, 333 N.W.2d 830, 832-33 (Iowa 1983) (finding that a “wrongful death” civil procedure rule applies to fetal death, but a fetus is not a “person” for the purposes of survival-wrongful death hybrid statute). The discrepancy merits explanation. Wrongful death actions are the product of specific statutory provisions, which are not

wrongful death statute to encompass fetal death,⁹⁵ superseding long and well established authorities denying both wrongful death and survival actions in cases of this sort.⁹⁶ We believe that in doing so (i.e., amending only the wrongful death statute), the legislature in both states intended to uphold the application of the old rule to survival actions.⁹⁷

Fourth and most significant, in the vast majority of jurisdictions that allow survival and wrongful death actions in cases of fetal death, these actions are maintainable only where the fetus was *viable* at time of its death.⁹⁸ In Georgia

identical in all common law jurisdictions. Whenever the courts conclude that an unborn child is a "person" whose death entitles its dependants to claim for their own loss, they do so through interpretation of the local wrongful death statute. Such interpretation of a *sui generis* act does not have to alter the status (more accurately non-status) of the unborn fetus for other legal purposes, unless otherwise stated by the legislature. The survival statute may thus be interpreted differently from the wrongful death statute, although in most states it is not.

95. NEB. REV. STAT. § 30-809 (2003); TEX. CIV. PRAC. & REM. CODE ANN. § 71.001 (Vernon 2004).

96. See, e.g., *Egbert v. Wenzl*, 260 N.W.2d 480, 481-82 (Neb. 1977) (holding that a child born dead cannot maintain an action at common law for injuries received by it while in its mother's womb, and consequently the personal representative cannot maintain the action under a wrongful death statute); *Krishnan v. Sepulveda*, 916 S.W.2d 478, 481 (Tex. 1995) (same).

97. NEB. REV. STAT. § 25-1401 (2003); TEX. CIV. PRAC. & REM. CODE ANN. § 71.021 (Vernon 2004). In our opinion, although the Texas wrongful death statute applies generally to fetal death it does not apply to wrongful abortions. This means that in Texas neither wrongful death nor survival actions are currently available in cases of wrongful abortion.

98. *Espadero v. Feld*, 649 F. Supp. 1480, 1484 (D. Colo. 1986) (applying Colorado law); *Wade v. United States*, 745 F. Supp. 1573, 1579 (D. Haw. 1990) (applying Hawaii law); *Gentry v. Gilmore*, 613 So. 2d 1241, 1244 (Ala. 1993); *Summerfield v. County of Maricopa Super. Ct.*, 698 P.2d 712 (Ariz. 1985); *Aka v. Jefferson Hosp. Ass'n*, 42 S.W.3d 508 (Ark. 2001); *Gorke v. Le Clerc*, 181 A.2d 448, 451 (Conn. Super. Ct. 1962); *Worgan v. Greggo & Ferrara, Inc.*, 128 A.2d 557 (Del. Super. Ct. 1956); *Volk v. Baldazo*, 651 P.2d 11 (Idaho 1982); *Dunn v. Rose Way, Inc.*, 333 N.W.2d 830 (Iowa 1983); *Shelton v. DeWitte*, 26 P.3d 650 (Kan. 2001); *Rice v. Rizk*, 453 S.W.2d 732, 735 (Ky. 1970); *Kandel v. White*, 663 A.2d 1264, 1268 (Md. 1995); *Remy v. MacDonald*, 801 N.E.2d 260, 265 (Mass. 2004); *Pehrson v. Kistner*, 222 N.W.2d 334 (Minn. 1974); *Strzelczyk v. Jett*, 870 P.2d 730, 733 (Mont. 1994); *White v. Yup*, 458 P.2d 617, 623-24 (Nev. 1969); *Wallace v. Wallace*, 421 A.2d 134, 137 (N.H. 1980); *Miller v. Kirk*, 905 P.2d 194, 197 (N.M. 1995); *DiDonato v. Wortman*, 358 S.E.2d 489 (N.C. 1987); *Hopkins v. McBane*, 359 N.W.2d 862, 865 (N.D. 1984); *Werling v. Sandy*, 476 N.E.2d 1053, 1056 (Ohio 1985); *Evans v. Olson*, 550 P.2d 924, 925 (Okla. 1976); *Libbee v. Permanente Clinic*, 520 P.2d 361 (Or. 1974); *Coveleski v. Bubnis*, 634 A.2d 608, 610 (Pa. 1993); *Miccolis v. AMICA Mut. Ins., Co.*, 587 A.2d 67, 71 (R.I. 1991); *Crosby v. Glasscock Trucking Co.*, 532 S.E.2d 856, 857 (S.C. 2000); *Vaillancourt v. Med. Ctr. Hosp., Inc.*, 425 A.2d 92, 94-95 (Vt. 1980); *Moen v. Hanson*, 537 P.2d 266, 267-68 (Wash. 1975); *Kwaterski v. State Farm Mut. Auto. Ins. Co.* 148 N.W. 2d 107, 111-12 (Wis. 1967); see also TENN. CODE ANN. § 20-5-106(c) (2004); Joseph McReynolds, *Childhood's*

and Mississippi, they are maintainable only if the fetus was either viable or quick.⁹⁹ Only a handful of jurisdictions allow the action for an injury caused at any point during gestation.¹⁰⁰ The viability requirement is indeed very significant in the current context: It means that any negligently instigated abortion of a non-viable fetus will not lead to liability. Since pre-viability abortions constitute the majority of legal abortions, the viability requirement weakens the protection of the public interest in the potentiality of life to a very considerable extent, although its protection would not violate any constitutional right.¹⁰¹

2. *Extent of Liability*

We now turn to the extent of liability in those jurisdictions that allow wrongful death and survival actions in cases of fetal death. Theoretically, wrongful death statutes apply to relational losses whereas survival statutes apply to the personal losses of the decedent (here the fetus), although in some states this distinction is not clear. The dichotomy of wrongful death/survival statutes therefore parallels our dichotomy of relational/fetal losses, and may be useful in

End: Wrongful Death of a Fetus, 42 LA. L. REV. 1411, 1412 (1982) (discussing the viability requirement in the early 1980s). Some of the cases cited herein do not explicitly hold that wrongful death and survival actions cannot be maintained where the fetus was non-viable at the time of death. However, in light of the historical background discussed above, holding—as these cases do—that the death of a fetus may yield rights of action *if* the fetus was *viable*, and emphasizing the fact of viability, is equivalent to maintaining the traditional rule with regard to non-viable fetuses, at least for the time being.

99. Citron v. Ghaffari, 542 S.E.2d 555 (Ga. Ct. App. 2000); 66 Fed. Credit Union v. Tucker, 853 So. 2d 104, 114 (Miss. 2003) (“Quickening is the period prior to viability when the mother first feels the fetus move in the womb, normally between the sixteenth and eighteenth week of pregnancy.”); Sandra L. Smith, Note, *Fetal Homicide: Woman or Fetus as Victim? A Survey of Current State Approaches and Recommendations for Future State Application*, 41 WM. & MARY L. REV. 1845, 1855 (2000).

100. 740 ILL. COMP. STAT. ANN. 180/2.2 (2004); MICH. COMP. LAWS § 600.2922a (2004); NEB. REV. STAT. § 30-809 (2003); S.D. CODIFIED LAWS § 21-5-1 (Michie 2003), *interpreted in* Wiersma v. Maple Leaf Farms, 543 N.W.2d 787, 791 (S.D. 1996); TEX. CIV. PRAC. & REM. CODE ANN. § 71.001 (Vernon 2004); Danos v. St. Pierre, 402 So. 2d 633, 638-39 (La. 1981); Connor v. Monkem Co., 898 S.W.2d 89, 92 (Mo. 1995) (en banc); Farley v. Sartin, 466 S.E.2d 522, 533 (W. Va. 1995). According to our foregoing analysis, Illinois and Texas do not allow wrongful death or survival actions in cases of wrongful abortion, whereas Louisiana and Nebraska do not allow survival actions. This leaves only four jurisdictions that allow both types of actions in cases of a wrongful abortion of a non-viable and non-quick fetus (Michigan, Missouri, South Dakota, and West Virginia).

101. Eighty-eight percent of all abortions performed in the United States occur during the first six to twelve weeks of pregnancy, i.e., before viability and quickening. See Katz, *supra* note 21.

determining whether these statutes constitute an adequate legal response to wrongful abortions.

We start with recovery for *relational* pecuniary losses in *wrongful death* actions. The primary head of damages in wrongful death actions is “loss of support.” The exact method employed to calculate and distribute damages for such loss depends on the specific wording of each statute. However, the various methods of calculation may be subsumed under two general categories.¹⁰²

The loss-to-survivors method (which is employed by most jurisdictions) measures economic loss by the loss of support to recognized dependants.¹⁰³ Most states that use this method also recognize as an element of damages the loss of a prospective inheritance based on the probability that the decedent would have accumulated an estate out of her earnings and would have left it to her surviving beneficiaries.¹⁰⁴ This method does not guarantee an appropriate response to the loss of potential life. In principle, the legally recognized dependants of the unborn child may recover for loss of support based on his or her anticipated earning power and probable contribution to existing legally-recognized dependants.¹⁰⁵ However, it can hardly be said that an unborn child would have supported any of her existing legally-recognized dependants or that any of them would have inherited her assets. Consequently, damages for loss of support constitute only a small fraction of the economic ingredient of the loss of potential life, i.e., some of the relational economic losses (those actually incurred by recognized dependants) plus a very small portion, if any, of the fetal loss (where it can be proved that the survivors are likely to inherit the unborn child).

The loss-to-the-estate method (which is employed by very few jurisdictions) bases the measurement of economic loss on the projected lifetime earnings of the

102. DOBBS, *supra* note 44, § 8.3(3); STUART M. SPEISER ET AL., RECOVERY FOR WRONGFUL DEATH AND INJURY § 1:9 (3d ed. 1992); Cindy Domingue-Hendrickson, Note, *New Mexico Adopts Hedonic Damages in the Context of Wrongful Death Actions: Sears v. Nissan (Romero v. Byers)*, 25 N.M. L. REV. 385, 388 (1995).

103. See *infra* note 106 for exceptions.

104. See, e.g., Kulawik v. ERA Jet Alaska, 820 P.2d 627, 634 (Alaska 1991) (including prospective inheritance as element of loss-to-the-beneficiary); Denver & R.G.R. Co. v. Spencer, 61 P. 606, 609 (Colo. 1900) (same); Reynolds v. Willis, 209 A.2d 760, 762 (Del. 1965) (same); Gonzalez v. N.Y. City Hous. Auth., 572 N.E.2d 598, 600-01 (N.Y. 1991) (same); Yowell v. Piper Aircraft Corp., 703 S.W.2d 630, 632-33 (Tex. 1986) (same).

105. See, e.g., Carey v. Lovett, 622 A.2d 1279, 1291 (N.J. 1993) (“When parents sue for the wrongful death of a child, their damages may include the pecuniary value of the child’s help with household chores, the pecuniary value of the child’s anticipated financial contributions, and the pecuniary value of the child’s companionship, including his or her advice and guidance, as the parents grow older.”). Note, however, that under New Jersey law wrongful death actions are not permitted in cases of fetal death.

deceased.¹⁰⁶ Under loss-to-the-estate statutes, the fact that there are no beneficiaries who have sustained a pecuniary loss does not preclude recovery. There are three variations of the loss-to-the-estate method.

In the net earnings method, loss-to-the-estate is calculated by determining the deceased's probable lifetime earnings and then deducting the expenses the decedent would have had in maintaining herself,¹⁰⁷ and in some states also income tax.¹⁰⁸ Amounts that would have been expended to support other family members are not deducted. In that case, damages cover a portion of the aggregate relational pecuniary loss, and, quite surprisingly,¹⁰⁹ the pecuniary element of the fetal loss.

For the net savings method, loss-to-the-estate is measured by the present value of the amount the decedent would have saved and left as an estate had he or she survived to a normal life expectancy.¹¹⁰ Unlike the net earnings method, this formula requires deduction of amounts the decedent would have expended to support her dependents (and, of course, income tax).¹¹¹ Under this method, wrongful death proceeds are equal to the pecuniary element of the fetal loss. Income-based relational economic losses (familial, business, etc.) are not redressed at all.¹¹²

Next, in the gross income method, loss-to-the-estate is measured by the present value of the decedent's gross future earnings. Under this theory, no deductions are made for either the decedent's personal living expenses or the amount that would have been expended to support her dependents.¹¹³ The proceeds in those jurisdictions are equal to any income-based relational loss, plus the pecuniary element of the fetal loss.¹¹⁴

106. GA. CODE ANN. §§ 51-4-1, -2 (2004), *construed in* W. & Atl. R.R. Co. v. Michael, 165 S.E. 37, 41 (Ga. 1932); IOWA CODE § 633.336 (2003); KY. REV. STAT. § 411.130 (2004), *construed in* Luttrell v. Wood, 902 S.W.2d 817, 819 (Ky. 1995); N.H. REV. STAT. ANN. § 556:12 (2003); R.I. GEN. LAWS § 10-7-1.1 (2004).

107. DOBBS, *supra* note 44, § 8.3(4); *see also, e.g.*; R.I. GEN. LAWS § 10-7-1.1 (2004); Kennett v. Delta Air Lines, Inc., 560 F.2d 456, 458, 461 (1st Cir. 1977) (applying New Hampshire law); Varney v. Taylor, 448 P.2d 164, 167 (N.M. 1968).

108. *See, e.g.*, Turcotte v. Ford Motor Co., 494 F.2d 173, 185-86 (1st Cir. 1974) (applying Rhode Island law).

109. Given that wrongful death statutes were historically intended to provide compensation for relational losses only.

110. *See, e.g.*, Pagel v. Notbohm, 188 N.W.2d 314, 315 (Iowa 1971).

111. Adams v. Deur, 173 N.W.2d 100, 105 (Iowa 1969).

112. *See supra* note 109 and accompanying text.

113. Brock v. Wedincamp, 558 S.E.2d 836, 839-41 (Ga. Ct. App. 2002); Miller v. Jenkins, 412 S.E.2d 555, 556 (Ga. Ct. App. 1991).

114. *See supra* note 109 and accompanying text.

In addition to loss of support, several jurisdictions have statutorily provided for recovery for loss of services to a few enumerated relatives.¹¹⁵ In other jurisdictions, courts have interpreted more general damages provisions in their wrongful-death statutes as allowing such recovery.¹¹⁶ Note, however, that only a few existing relatives are entitled to recover for loss of support and services, and it is very likely that the financial support and services lost by those relatives in cases of wrongful abortion are overshadowed by the maintenance costs and services that they would have bestowed upon the child had she been born alive.¹¹⁷ One court held that damages for loss of support and services are not available at all in cases of a wrongful death of a viable fetus because of their speculative nature.¹¹⁸

Relational, non-pecuniary losses pose an even more serious problem from the deterrence standpoint. These losses may consist of (1) loss of companionship and affection and (2) mental anguish and grief. Most jurisdictions permit recovery for the loss of companionship, society, love, and affection incurred by the statutory dependants. In several states, these heads of damages are explicitly recognized in the wrongful death statute,¹¹⁹ while in others they were recognized by the courts through statutory interpretation.¹²⁰ Whenever a minor child is

115. *E.g.*, ALASKA STAT. § 09.55.580(c)(3) (2004); DEL. CODE ANN. tit. 10, § 3724(d)(3) (2004); FLA. STAT. ch. 768.21(1) (2004); MASS. GEN. LAWS ANN. ch. 229, § 2 (2004); MO. REV. STAT. § 537.090 (2004); N.C. GEN. STAT. § 28A-18-2(b)(4)(b) (2004); OHIO REV. CODE ANN. § 2125.02(B)(2) (West 2004); OR. REV. STAT. § 30.020(2)(d) (2003); R.I. GEN. LAWS § 10-7-1.1 (2004); VA. CODE ANN. § 8.01-52(2) (2004); W. VA. CODE § 55-7-6(c)(1) (2003).

116. *E.g.*, *Muckler v. Buchl*, 150 N.W.2d 689, 697-98 (Minn. 1967).

117. *McClurg*, *supra* note 49, at 57-60 (1990) (noting that as childrearing expenses are usually higher than the pecuniary benefits bestowed by a child, the relational pecuniary loss is usually negative).

118. *DiDonato v. Wortman*, 358 S.E.2d 489, 494 (N.C. 1987); *Greer v. Parsons*, 416 S.E.2d 174, 174, 176 (N.C. 1992).

119. *E.g.*, ALASKA STAT. § 09.55.580(c)(4) (Michie 2004); ARK. CODE ANN. § 16-62-102(f) (Michie 2003); COLO. REV. STAT. § 13-21-203(1)(a) (2003); FLA. STAT. ch. 768.21(2)-(3) (2004); HAW. REV. STAT. § 663-3 (2003); KAN. STAT. ANN. § 60-1904(a)(2) (2003); KY. REV. STAT. ANN. § 411.135 (Banks-Baldwin 2004); ME. REV. STAT. tit. 18-A, § 2-804(b) (West 2004); MD. CODE ANN., CTS. & JUD. PROC. § 3-904(d) (2003); MASS. GEN. LAWS ANN. ch. 229, § 2 (West 2004); MICH. COMP. LAWS § 600.2922(6) (2004); MO. REV. STAT. § 537.090 (2004); NEV. REV. STAT. § 41.085(4) (2004); N.C. GEN. STAT. § 28A-18-2(b) (2004); OHIO REV. CODE ANN. § 2125.02(B)(3) (West 2004); OKLA. STAT. tit. 12, § 1053(B) (2004); OR. REV. STAT. § 30.020(2)(d) (2003); R.I. GEN. LAWS § 10-7-1.2 (2004); VA. CODE ANN. § 8.01-52(1) (2004); W. VA. CODE ANN. § 55-7-6(c)(1) (Michie 2003); WIS. STAT. § 895.04(4) (2003); WYO. STAT. ANN. § 1-38-102(c) (Michie 2003).

120. *E.g.*, *Krouse v. Graham*, 562 P.2d 1022, 1025-28 (Cal. 1977); *Elliott v. Willis*, 442 N.E.2d 163, 167-68 (Ill. 1982); *Gravley v. Sea Gull Marine, Inc.*, 269 N.W.2d 896, 901 (Minn. 1978);

killed, these losses are probably more significant than pecuniary losses, given that in our times a child confers upon her family non-pecuniary rather than pecuniary benefits. The same is true in cases of fetal death, including wrongful abortion cases. Most jurisdictions that allow wrongful death actions in cases of fetal death permit recovery for the loss of companionship and affection of the unborn child.¹²¹ However, at least one state supreme court held that these losses are irrecoverable in cases of a wrongful death of a fetus due to their speculative nature.¹²² More importantly, while a living human being may bestow non-pecuniary benefits upon various persons (relatives, neighbors, friends, colleagues, etc.), only a small fraction of the aggregate loss—that incurred by a few existing and legally-recognized dependants—is covered.

As mentioned above, a few states have adopted legislation that specifically authorizes recovery of damages for mental anguish or grief arising from the death of a loved one; in others, the statute has been construed to include compensation for such loss.¹²³ Yet in many states, the wrongful death statute either specifically restricts recovery to pecuniary losses or has been construed to exclude liability for mental anguish.¹²⁴ Moreover, assuming that wrongful abortion may aggrieve many people apart from the statutory dependants, the aggregate distress will not be accounted for even if wrongful death statutes are interpreted to encompass mental anguish.¹²⁵

In sum, the damages claimable under the wrongful death statutes reflect merely a fraction of actual relational losses. Wrongful death statutes compensate a few relatives for *certain* losses. Even if the legally recognized dependants were

Dickey v. Parham, 331 So. 2d 917, 918 (Miss. 1976); Swanson v. Champion Int'l Corp., 646 P.2d 1166, 1170 (Mont. 1982); Spangler v. Helm's N.Y.-Pittsburgh Motor Express, 153 A.2d 490, 492 (Pa. 1959); Flagtwet v. Smith, 367 N.W.2d 188, 189-91 (S.D. 1985).

121. See, e.g., Burnham v. Miller, 972 P.2d 645, 647 (Ariz. Ct. App. 1998); Volk v. Baldazo, 651 P.2d 11, 14 (Idaho 1982); Seef v. Sutkus, 583 N.E.2d 510, 511-12 (Ill. 1991); Dunn v. Rose Way, Inc., 333 N.W.2d 810, 831-32 (Iowa 1983); Hopkins v. McBane, 427 N.W.2d 85, 92 (N.D. 1988). However, according to our foregoing analysis, although Illinois allows wrongful death actions in cases of fetal death, it excludes liability in cases of wrongful abortion. Cf. Fraternal Order of Eagles v. Ill. Cas. Co., 364 N.W.2d 218, 221 n.2 (Iowa 1985) (finding that when parents sue for the wrongful death of a child, loss of society, companionship, etc. is recoverable); Williams v. Monarch Transp., Inc., 470 N.W.2d 751, 755 (Neb. 1991) (same); Sanchez v. Schindler, 651 S.W.2d 249, 251 (Tex. 1983) (same); Clymer v. Webster, 596 A.2d 905, 914-15 (Vt. 1991) (same).

122. *DiDonato*, 358 S.E.2d at 494; *Greer*, 416 S.E.2d at 176.

123. See *supra* notes 53-54.

124. See *supra* note 55.

125. The law governing negligent infliction of emotional harm usually sets very restrictive guidelines regarding liability for emotional harm resulting from an injury to another. See *supra* note 47.

compensated for their entire loss, the wrongdoer would not bear relational losses incurred by those who were not legally recognized dependants (e.g., the state, purveyors of goods and services, potential friends, potential neighbors, potential spouses and cohabitants, and the public at large).¹²⁶ Moreover, even those jurisdictions that allow wrongful death actions for fetal deaths do not necessarily compensate legally recognized dependants for their entire loss.

Similarly, most survival statutes do not cover the entire *fetal* loss. Theoretically, a survival action, being an action inherited by the deceased, should cover any loss incurred by her, including pecuniary and non-pecuniary losses. We saw in Section I.B that an important component of the social costs of wrongful abortion is the value of the assets that the unborn child could have acquired during her lifetime. This element equals the loss of the child's earning power minus her costs of maintenance and any contribution made by her to other persons (who may sometimes recover for their loss in a wrongful death action). Most jurisdictions allow the estate in a survival action to recover for such loss.¹²⁷ However, in several states no recovery is allowed for future lost earnings in a survival action.¹²⁸ In those states, the pecuniary element of the fetal loss remains unaccounted for.

The non-pecuniary element once again raises a more acute problem. Most jurisdictions do not allow compensation for the loss of the ability to enjoy life in cases of wrongful death (either in a survival action or in a loss-to-the-estate based wrongful death action).¹²⁹ This seems to be the majority view, subject to only a

126. Relational losses are generally irrecoverable at common law, subject to the wrongful death statutory exception, and a few common law exceptions that are inapplicable here. *See supra* note 47.

127. *See, e.g.,* Weil v. Seltzer, 873 F.2d 1453, 1462 (D.C. Cir. 1989); Loden v. Getty Oil Co., 359 A.2d 161, 163-64 (Del. 1976); Payne v. Eighth Judicial Dist. Ct., 60 P.3d 469, 472 (Mont. 2002); Amadio v. Levin, 501 A.2d 1085, 1088 (Pa. 1985); Cavazos v. Franklin, 867 P.2d 674, 678 (Wash. Ct. App. 1994).

128. DOBBS, *supra* note 44, § 8.3(2) n.1; *see, e.g.,* Greene v. Texeira, 505 P.2d 1169, 1172-73 (Haw. 1973); Flowers v. Marshall, 494 P.2d 1184, 1190-91 (Kan. 1972); Jones v. Flood, 716 A.2d 285, 290 (Md. 1998); Prunty v. Schwantes, 162 N.W.2d 34, 38 (Wis. 1968).

129. Domingue-Hendrickson, *supra* note 102, at 386-87 (noting that courts generally have not held tortfeasors liable for the value of the decedent's life in wrongful death actions); Erin A. O'Hara, Note, *Hedonic Damages for Wrongful Death: Are Tortfeasors Getting Away with Murder?*, 78 GEO. L.J. 1687, 1691-92 (1990) (stating that in most jurisdictions the tortfeasor is not held liable for depriving the victim of the enjoyment of life); *see also, e.g.,* Sterner v. Wesley Coll., Inc., 747 F. Supp. 263, 272-74 (D. Del. 1990) (applying Delaware law); Nichols v. Estabrook, 741 F. Supp. 325, 328-29 (D.N.H. 1989) (applying New Hampshire law); O'Leary v. U.S. Lines, 111 F. Supp. 745, 747 (D. Mass. 1953) (applying Massachusetts law); Phillips v. E. Me. Med. Ctr., 565 A.2d 306, 309 (Me. 1989); Smallwood v. Bradford, 720 A.2d 586, 594-95 (Md. 1998); Smith v. Whitaker, 734 A.2d 243, 251 (N.J. 1999); Willinger v. Mercy Catholic Med. Ctr., 393 A.2d 1188,

few deviations.¹³⁰ An important element of the fetal loss thus remains uncompensated.

In conclusion, tort law is not responsive to the lion's share of the loss of potential human life. In many states, wrongful death and survival statutes are not applicable to cases of wrongful abortion; in others, they are not applicable to the vast majority of wrongful abortions (i.e., wrongful abortions of non-viable fetuses). Even where applicable, they do not make the tortfeasor accountable for the entire social costs of the wrongful abortion. We acknowledge that the gap between the social cost of human death and the scope of civil liability imposed under wrongful death and survival legislation is not distinctive of fetal death.¹³¹ It arises whenever a person is wrongfully killed. However, it seems more acute where the immediate victim is a fetus or a minor child.¹³² More importantly, the

1190-91 (Pa. 1978); *Spencer v. A-1 Crane Serv., Inc.*, 880 S.W.2d 938, 943-44 (Tenn. 1994); *Otani ex rel. Shigaki v. Broudi*, 92 P.3d 192, 198 (Wash. 2004); *Schwantes*, 162 N.W.2d at 38. In other states, it was held that loss of enjoyment of life is a component of pain and suffering, for which compensation depends on the awareness of the victim. *See, e.g.*, *Pitman v. Thorndike*, 762 F. Supp. 870, 872-73 (D. Nev. 1991); *Poyzer v. McGraw*, 360 N.W.2d 748, 753 (Iowa 1985); *Leiker ex rel. Leiker v. Gafford*, 778 P.2d 823 (Kan. 1989); *Anderson v. Neb. Dep't of Soc. Servs.*, 538 N.W.2d 732, 740-41 (Neb. 1995); *Nussbaum v. Gibstein*, 536 N.E.2d 618 (N.Y. 1989); *First Trust Co. v. Scheels Hardware & Sports Shop, Inc.*, 429 N.W.2d 5, 13-14 (N.D. 1988); *Mo. Pac. R.R. v. Lane*, 720 S.W.2d 830, 834 (Tex. Ct. App. 1986); *Bulala v. Boyd*, 389 S.E.2d 670, 677-78 (Va. 1990). Obviously, the decedent's loss of enjoyment of life cannot be recovered under a loss-to-survivors wrongful death statute. *See, e.g.*, *Russell, Inc. v. Trento*, 445 So. 2d 390, 392 (Fla. Dist. Ct. App. 1984); *Southlake Limousine & Coach, Inc. v. Brock*, 578 N.E.2d 677, 679-80 (Ind. Ct. App. 1991).

130. A few states allow recovery for loss of enjoyment of life (LOEL) under a loss-to-the-estate wrongful death statute. *See Choctaw Maid Farms, Inc. v. Hailey*, 822 So. 2d 911, 922-23 (Miss. 2002) (finding right to recover for LOEL); *Romero v. Byers*, 872 P.2d 840, 846 (N.M. 1994) (same). Two states permit recovery in a survival action. *See* ARK. CODE ANN. § 16-62-101(b) (Michie 2003) ("In addition to all other elements of damages provided by law, a decedent's estate may recover for the decedent's loss of life as an independent element of damages."); *Durham v. Marberry*, No. CV-2002-24, 2004 Ark. LEXIS 179, at *17-18 (Mar. 25, 2004); *Ozaki v. Ass'n of Apartment Owners*, 954 P.2d 652, 667-68 (Haw. Ct. App. 1998). Loss-of-life damages are also available in survival actions brought under 42 U.S.C. § 1983 (2000) in cases of constitutional deprivation resulting in death. *See, e.g.*, *Lewis ex rel. Bass v. Wallenstein*, 769 F.2d 1173, 1190 (7th Cir. 1985).

131. As opposed to the possibility of no liability at all for the lost value of life that is distinctive of fetal death or at least fetal death through wrongful abortion.

132. A fetus, like a minor child, does not normally have dependants. "Future dependants" cannot sue for their own loss (since they are not yet dependants or are not in existence at the time of death). Yet the prospects of marriage and procreation may reduce the immediate victim's compensation for loss of earnings. Consequently, the loss of potential relational advantages is not internalized by the injurer.

gap between the social cost and actual liability (where liability is imposed) clearly intensifies the problem of under-deterrence discussed in this article. In a comprehensive analysis of the inability of tort law to make a negligent adviser accountable for the social costs of wrongful abortion, it simply cannot be ignored. Whether and to what extent this gap should be dealt with in other contexts is a question that lies outside the scope of this article.¹³³

C. Protection of the Public Interest in Criminal Law

Criminal law is another possible avenue for the protection of potential life. In earlier times, the killing of a fetus was not in itself a criminal offense, given that the fetus was not regarded as an independent human being entitled to legal protection.¹³⁴ The courts in the United States, England, Canada, and Australia have generally held that a person who injured a fetus (resulting in a stillbirth) could not be guilty of murder or manslaughter.¹³⁵ In the modern era, however, criminal law has been modified in several jurisdictions to protect fetuses. This was done in three different ways. In most states where a revision took place, a new criminal offense (usually termed “feticide,” “homicide of unborn child,” “murder of an unborn child,” “manslaughter of an unborn child,” or the like) was established by the legislature.¹³⁶ A small number of state legislatures, along with the federal government, have expanded the definition of traditional homicide offenses to encompass the killing of a fetus.¹³⁷ In a few other states, the courts

133. On the one hand, it may be argued that by not imposing the social cost of taking life on the person whose conduct caused death, tort law does not adequately protect living persons. On the other hand, in most jurisdictions living persons are better protected by criminal law than the unborn.

134. Tara Kole & Laura Kadetsky, *The Unborn Victims of Violence Act*, 39 HARV. J. ON LEGIS. 215, 216 (2002); Mary Lynn Kime, *Hughes v. State: The “Born Alive” Rule Dies a Timely Death*, 30 TULSA L.J. 539, 539-41 (1995); Clarke D. Forsythe, *Homicide of the Unborn Child: The Born Alive Rule and Other Legal Anachronisms*, 21 VAL. U. L. REV. 563, 571 (1987).

135. SEYMOUR, *supra* note 87, at 137.

136. ARIZ. REV. STAT. § 13-1103(A)(5) (2004); FLA. STAT. ch. 782.09 (2004); GA. CODE ANN. § 16-5-80 (2004); 720 ILL. COMP. STAT. ANN. 5/9-1.2, -2.1, -3.2, 5/12-3.1, -4.4 (2004); IND. CODE §§ 35-42-1-1(4), -42-1-6 (2004); LA. REV. STAT. ANN. §§ 14:2(7), (11), 14:32.5-.8 (West 2004); MICH. COMP. LAWS § 750.322 (2004); MINN. STAT. ANN. §§ 609.2661-.2665, .268 (2003); MISS. CODE ANN. § 97-3-37 (2004); NEV. REV. STAT. § 200.210 (2004); N.D. CENT. CODE §§ 12.1-17.1-01 to -06 (2003); OHIO REV. CODE ANN. § 2903.09(A) (West 2004); 18 PA. CONS. STAT. §§ 2601-2607 (2004); R.I. GEN. LAWS § 11-23-5 (2004); S.D. CODIFIED LAWS §§ 22-16-1, -1.1 (Michie 2003); WASH. REV. CODE § 9A.32.060(1)(b) (2004); WIS. STAT. §§ 940.01, .02, .04 to .06 (2003).

137. ARK. CODE ANN. § 5-1-102(13)(B) (Michie 2003); CAL. PENAL CODE § 187 (West 2004); MISS. CODE ANN. § 97-3-37(1) (2004); MO. REV. STAT. § 1.205.2 (2004), *construed in* State v. Holcomb, 956 S.W.2d 286 (Mo. Ct. App. 1997); OHIO REV. CODE ANN. § 2901.01(B)(1)(a)(ii)

modified the time-honored “born alive” rule of the common law to make the third-party killing of a viable fetus a crime under general homicide statutes.¹³⁸ Nonetheless, criminal law currently seems incapable of dealing with the societal problem of wrongful abortions. At this stage, we do not contend that it should, only that at present it does not.

First, many jurisdictions still adhere to the traditional view that the killing of a fetus does not constitute a criminal offense.¹³⁹ In some of these jurisdictions, the criminal statute explicitly defines “person,” “individual,” or “human being” as one who is born and alive.¹⁴⁰ In others, where no explicit statutory exclusion of the fetus is present, courts have held that the aforementioned terms do not encompass fetuses. The number of jurisdictions that have not criminalized fetal homicide is much larger than the number of jurisdictions that do not allow wrongful death actions in cases of fetal death. The former group consists of nearly all jurisdictions that do not allow wrongful death actions (California being an exception),¹⁴¹ plus many others that do.¹⁴²

Second, although several jurisdictions have criminalized the killing of a fetus at any stage of gestation,¹⁴³ many others did not go this far. In most

(West 2004); TENN. CODE ANN. § 39-13-214 (2004); TEX. PENAL CODE ANN. § 1.07(26) (Vernon 2004); UTAH CODE ANN. § 76-5-201 (2004). A similar reform took place on the federal level. The Unborn Victims of Violence Act, 18 U.S.C.A. § 1841 (West Supp. 2004), provides that if a person intentionally kills or attempts to kill an unborn child, that person shall be punished for intentionally killing or attempting to kill a human being. The same act provides that whoever engages in conduct that violates any of the provisions of law listed therein and thereby causes (unintentionally) the death of or bodily injury to a child, who is in utero at the time the conduct takes place, is guilty of a separate offense.

138. *Commonwealth v. Morris*, 142 S.W.3d 654, 660 (Ky. 2004); *Commonwealth v. Cass*, 467 N.E.2d 1324, 1329 (Mass. 1984); *Hughes v. State*, 868 P.2d 730, 734 (Okla. Crim. App. 1994); *State v. Horne*, 319 S.E.2d 703, 704 (S.C. 1984).

139. SEYMOUR, *supra* note 87, at 140 (noting that the “born alive” rule remains part of homicide law in many jurisdictions).

140. ALA. CODE § 13A-6-1(2) (2004); ALASKA STAT. § 11.41.140 (Michie 2004); COLO. REV. STAT. § 18-3-101(2) (2003); HAW. REV. STAT. ANN. § 707-700 (Michie 2003); MONT. CODE ANN. § 45-2-101(28) (2003); NEB. REV. STAT. § 28-302(2) (2003); OR. REV. STAT. § 163.005(3) (2003).

141. *See, e.g., In re A.W.S.*, 440 A.2d 1144, 1145 (N.J. Super. Ct. 1981); *People v. Vercelletto*, 514 N.Y.S.2d 177, 179 (Ulster County Ct. 1987); *People v. Joseph*, 496 N.Y.S.2d 328, 328-29 (Orange County Ct. 1985).

142. *See, e.g., State v. Anonymous*, 516 A.2d 156, 157 (Conn. Super. Ct. 1986); *State v. Green*, 781 P.2d 678, 682 (Kan. 1989); *Williams v. State*, 550 A.2d 722, 726 (Md. 1988); *State v. Beale*, 376 S.E.2d 1, 4 (N.C. 1989); *State v. Oliver*, 563 A.2d 1002, 1003-05 (Vt. 1989); *State ex rel. Atkinson v. Wilson*, 332 S.E.2d 807, 809-10 (W. Va. 1984).

143. 18 U.S.C.A. § 1841(d) (West Supp. 2004); ARIZ. REV. STAT. § 13-1103(A)(5) (2004); 720 ILL. COMP. STAT. ANN. 5/9-1.2(b), -3.2(c) (West 2004); LA. REV. STAT. ANN. §§ 14:2(7), (11)

jurisdictions that have abandoned the “born alive” rule, either through explicit legislation or judicial interpretation, the killing of a fetus is a criminal offense only if the fetus was viable¹⁴⁴ or quick,¹⁴⁵ or has reached a certain stage of development prior to quickening.¹⁴⁶

Third, in some of the states that have criminalized fetal homicide, the killing of a fetus constitutes a criminal offense only when committed with intent to kill the fetus.¹⁴⁷ In California, malice aforethought is required.¹⁴⁸ Only a few states have criminalized negligent causation of fetal death.¹⁴⁹ An intention to kill the fetus is usually absent in the paradigmatic wrongful abortion case discussed here. The adviser usually intends to convey information to the prospective parents, not kill their potential offspring.

Fourth, and closely related to the previous point, in some states fetal homicide leads to criminal liability only if caused by a physical injury to the mother, which would be murder if the death of the mother had occurred.¹⁵⁰ In other words, the fatal injury to the fetus must be inflicted with intent to kill the mother. Of course, the act of the adviser in a wrongful abortion setting is not an intentional attempt to kill the pregnant woman.

(2004); MINN. STAT. ANN. § 609.266 (West 2003); MISS. CODE ANN. § 97-3-37 (2004) (as amended recently); MO. ANN. STAT. § 1.205.1 (West 2004); 18 PA. CONS. STAT. ANN. § 3203 (West 2004); S.D. CODIFIED LAWS § 22-1-2(50A) (Michie 2003); TEX. PENAL CODE ANN. § 1.07 (26) (Vernon 2004); UTAH CODE ANN. § 76-5-201 (2004); WIS. STAT. § 940.04 (2003).

144. See, e.g., OHIO REV. CODE ANN. § 2901.01(B)(1)(a)(ii) (West 2004); TENN. CODE ANN. § 39-13-214 (2004); *Commonwealth v. Cass*, 467 N.E.2d 1324, 1329 (Mass. 1984); *Hughes v. State*, 868 P.2d 730, 734 (Okla. Crim. App. 1994); *State v. Horne*, 319 S.E.2d 703, 704 (S.C. 1984).

145. FLA. STAT. ch. 782.09 (2004); GA. CODE ANN. § 16-5-80 (Michie 2004); MICH. COMP. LAWS § 750.322 (2004); NEV. REV. STAT. § 200.210 (2004); R.I. GEN. LAWS § 11-23-5 (2004); WASH. REV. CODE ANN. § 9A.32.060(1)(b) (West 2004).

146. ARK. CODE ANN. § 5-1-102(13)(B) (Michie 2003) (twelve weeks gestation); *People v. Davis*, 872 P.2d 591, 599 (Cal. 1994) (seven to eight weeks).

147. FLA. STAT. ch. 782.09 (2004); GA. CODE ANN. 16-5-80 (2004); IND. CODE § 35-42-1-6 (West 2004); MICH. COMP. LAWS § 750.322; NEV. REV. STAT. § 200.210 (2004); R.I. GEN. LAWS § 11-23-5 (2004); WASH. REV. CODE § 9A.32.060(1)(b) (2004); WIS. STAT. § 940.04 (2003).

Under 18 U.S.C.A. § 1841 (West Supp. 2004), there are two types of crimes against unborn children, neither of which is relevant in our context. The first type focuses on unintentional killing or injuring a fetus while committing a violent crime against its mother. *Id.* § 1841(a)(1). The second type focuses on *intentional* killing or attempting to kill a fetus. *Id.* § 1841(a)(2)(C).

148. CAL. PENAL CODE § 187 (West 2004).

149. See, e.g., 18 PA. CONS. STAT. ANN. §§ 2603, 2605 (West 2004); N.D. CENT. CODE § 12.1-17.1-04 (2003).

150. ARIZ. REV. STAT. § 13-1103(A)(5) (2004); FLA. STAT. ANN. § 782.09 (West 2004); GA. CODE ANN. § 16-5-80 (2004); MICH. COMP. LAWS § 750.322 (2004); MISS. CODE ANN. § 97-3-37 (2004); S.D. CODIFIED LAWS § 22-16-1.1 (Michie 2003).

Fifth, it can hardly be said that the adviser in a wrongful abortion case actually “kills” or “terminates the life” of the fetus for the purposes of criminal law. The adviser does not perform the act that terminates the pregnancy. The act of “killing” is usually performed by another person, who cannot be regarded as the adviser’s agent.¹⁵¹ Even if we could attribute the act to the adviser (which we do not believe is possible), most fetal homicide statutes explicitly grant immunity from any criminal liability to medical staff involved in a consented abortion (at least where the abortion is legal).¹⁵² This bars the possibility of any legal reaction to wrongful abortions on the criminal level.

Sixth, even where the killing of a fetus may constitute a criminal offense, “prosecutors are reluctant to charge individuals with these crimes.”¹⁵³ Such reluctance is present even in the most outrageous cases of fetal homicide.¹⁵⁴ It would thus be unrealistic to expect criminal law to make a significant contribution to the prevention of unintentional and indirect “killing” of fetuses, as in wrongful abortion settings, even if it formally applied to such cases (an evidently dubious assumption).

D. Protection of the Public Interest Through Disciplinary Proceedings

As we have just shown, criminal legislation in most states does not generally cover liability for a negligent professional misrepresentation leading to an unwanted and unnecessary abortion. Another possible response to this kind of medical malpractice is a disciplinary action. Professional codes grant state medical boards the authority to regulate and discipline physicians, nurses, and other persons involved in the practice of medicine, *inter alia* by imposing sanctions for professional misbehavior.¹⁵⁵ Is this legal channel effective and

151. It is quite probable that the adviser whose advice induces the woman to undergo an abortion is not the physician who actually performs the abortion.

152. 18 U.S.C.A. § 1841(c)(1) (West Supp. 2004); ARK. CODE ANN. § 5-1-102(13)(B) (Michie 2003); CAL. PENAL CODE § 187(b) (West 2004); 720 ILL. COMP. STAT. ANN. 5/9-1.2(c), -3.2(d) (West 2004); IND. CODE ANN. § 35-42-1-6 (West 2004); LA. REV. STAT. ANN. § 14:32.5(A) (West 2004); MISS. CODE ANN. § 97-3-37(3) (2004); N.D. CENT. CODE § 12.1-17.1-07 (2003); OHIO REV. CODE ANN. §§ 2901.01(B)(2)(a), 2903.09(C)(1) (West 2004); 18 PA. CONS. STAT. ANN. § 2608(a) (West 2004); S.D. CODIFIED LAWS § 22-16-1.1 (Michie 2003); TENN. CODE ANN. § 39-13-214 (2004); UTAH CODE ANN. § 76-5-201(1)(b) (2004).

153. JEAN REITH SCHROEDEL, IS THE FETUS A PERSON? A COMPARISON OF POLICIES ACROSS THE FIFTY STATES 132-33 (2000).

154. *Id.*

155. For a general survey of the main characteristics of disciplinary proceedings in the State of New York, see Paul Bennett Marrow, *Professional Misconduct: New York's Unified System for Professional Misconduct*, 29 WESTCHESTER COUNTY B. J. 15 (2002).

adequate as a deterrent against wrongful abortions?

We think that here too the answer is in the negative. First, the aim of disciplinary proceedings, in the context of medical malpractice as elsewhere, is usually thought to be one of protecting the public from incompetent professionals, rather than directly punishing improper professional behavior for the sake of deterrence or retribution.¹⁵⁶ The boards do not view their role as replacing or even complementing the more direct forms of controlling medical malpractice, i.e., criminal and civil law.¹⁵⁷ This view of the role of disciplinary actions is reflected in the provisions of most professional codes and in the actual practice of medical boards. Many codes do not authorize the imposition of fines at all (not to mention imprisonment), and the ones that do allow them only in relatively small sums.¹⁵⁸ In addition, fines are typically available for more

156. This common view is reflected in materials on the Federation of State Medical Boards website. Fed'n of State Med. Bds., What Is a State Medical Board?, at <http://www.fsmb.org/consumer.htm> (last visited Aug. 31, 2004) [hereinafter FSMB website] ("Medical boards may review malpractice reports to proactively identify practitioners who may be a hazard to the public by detecting a pattern of inappropriate actions. . . . Medical boards focus on protecting the public, not on punishing physicians."); see also N.C. Med. Bd. (NCMB), Topics of Interest About the Board and Its Work, at <http://www.ncmedboard.org/Clients/NCBOM/Public/PublicMedia/topics.htm> (last visited Apr. 6, 2005) ("Disciplinary action by the Board is primarily intended to protect the public by preventing a practitioner from doing harm (or further harm) to patients. The Board does not focus on punishing problem practitioners, though that may certainly be one effect of its action when a practitioner loses his or her license or is otherwise sanctioned by the Board."). The distinction between punishment and protection has been emphasized by the courts in similar contexts. See, e.g., *Attorney Grievance Comm'n v. Ashworth*, 851 A.2d 527, 536 (Md. 2004); *Bar Ass'n v. Marshall*, 307 A.2d 677, 682 (Md. 1973); see also *In re Bennethum*, 161 A.2d 229, 236 (Del. 1960); *In re Sabbath*, 662 S.W.2d 511, 512 (Mo. 1984) (en banc); *Nardi's Case*, 444 A.2d 512, 513 (N.H. 1982); *In re Willis* 552 A.2d 979, 982 (N.J. 1989); *Cleveland Bar Ass'n v. Feneli*, 712 N.E.2d 119, 121 (Ohio 1999).

157. See, e.g., NCMB, An Introduction to the North Carolina Medical Board, at <http://www.ncmedboard.org/Clients/NCBOM/Public/PublicMedia/intbro.htm> (last visited Apr. 6, 2005) ("Complaints to the Board should not be seen as an alternative to appropriate legal action when that is called for.").

158. The fines are comparatively small, the typical sum being \$1,000, \$5,000, or \$10,000 at most. For example, CAL. BUS. & PROF. CODE § 2670 (West 2004) authorizes the board to impose a maximum penalty of \$1,000 or six months' imprisonment for any violation of its provisions. N.Y. PUB. HEALTH LAW § 230-a(7) (McKinney 2004) permits the imposition of fines up to \$10,000 for each act of misconduct as defined by the statute. TEX. OCC. CODE ANN. § 165.003 (Vernon 2004) limits the fine (termed "administrative penalty") to \$5,000. However, there is one notable exception. CAL. BUS. & PROF. CODE § 2242.1 allows a medical board to impose a civil fine of not more than \$25,000 for each offense of illegally providing prescriptions for dangerous drugs. On February 10, 2003, the California medical board imposed an unprecedented total amount of \$48

“technical” offenses such as practicing without a valid license or outside the jurisdiction, rather than for providing deficient services. Finally, fines are rarely imposed in disciplinary proceedings, even where possible.¹⁵⁹ These facts echo the general view that a disciplinary action is directed against the physician’s license, not his or her pocket.¹⁶⁰

Even more important, disciplinary actions are generally unavailable, and, even where available, not regularly taken in cases of ordinary or one-time negligence, such as the medical negligence discussed in this Article. Rather, an action would usually be taken only against a physician whose conduct raises serious doubts as to his or her competence or personal integrity.¹⁶¹

A further limitation on the deterrent effect of disciplinary actions is of a more pragmatic nature, but nevertheless serious. Medical boards have limited resources and staff and consequently lack the power to create sufficient

million against six doctors who were found guilty of such conduct. Damon Adams, *California Fines Out-of-State Doctors for Prescribing*, AMNEWS, Mar. 3, 2003, at <http://www.ama-assn.org/amednews/2003/03/03/prsb0303.htm>.

159. For instance, in 2004 the NCMB imposed nine different kinds of sanctions against 154 defendants (130 of which were physicians), with not one judgment including a fine. NCMB, ANNUAL BOARD ACTION REPORT 2004 §§ 1, 3 (2005), <http://www.ncmedboard.org/Clients/NCBOM/Public/Board/2004annualreport.pdf> (last visited Apr. 6, 2005). The minor role of fines in disciplinary proceedings is also manifest in the fact that in the list of sanctions recommended by the FSMB it appears almost at the end, after ten other sanctions. A GUIDE TO THE ESSENTIALS OF A MODERN MEDICAL PRACTICE ACT § 9.A (10th ed., 2003), http://www.fsmb.org/Policy Documents and White Papers/tenth_edition_essentials.htm.

160. Position of Federation of State Medical Boards on Partial-Birth Abortion Ban Acts, <http://www.fsmb.org/Policy Documents and White Papers/partial birth acts.htm> (last visited Aug. 31, 2004) (“The purpose of a medical board hearing is to determine whether a violation of the Medical Practice Act has occurred that indicates the need for disciplinary action against a physician’s license in the interest of public protection.”); see also FSMB website, *supra* note 156 (“The board is charged with the responsibility of evaluating when a physician’s professional conduct or ability to practice medicine warrants modification, suspension or revocation of the license to practice medicine.”).

161. E.g., N.Y. EDUC. LAW § 6530 (McKinney 2004), which defines in length the various forms of professional misconduct, does not mention ordinary negligence as a form of professional misbehavior, but refers only to aggravated forms of negligence such as “gross negligence on a particular occasion” or “negligence on more than one occasion.” Similarly, CAL. BUS. & PROF. CODE § 2234 (West 2004) defines “unprofessional conduct” as including “gross negligence” or “repeated negligent acts” but not ordinary negligence. See also Kara M. McCarthy, *Doing Time for Clinical Crime: The Prosecution of Incompetent Physicians as an Additional Mechanism To Assure Quality Health Care*, 28 SETON HALL L. REV. 569, 584 (1997); Gregory G. Peters, *Reallocating Liability to Medical Staff Review Committee Members: A Response to the Hospital Corporate Liability Doctrine*, 10 AM. J.L. & MED. 115, 119 (1984).

incentives against malpractice.¹⁶² Due to budget restraints, and possibly lack of zeal in the members of medical boards to investigate and prosecute their colleagues, even grossly negligent practitioners may escape board sanctions.¹⁶³ According to a report by the Health Research Group of Public Citizen, a national non-profit public interest organization, only 2864 serious disciplinary actions were taken in 2002 by medical boards nationwide.¹⁶⁴ Given that in the same year a total of 805,372 licensed physicians practiced medicine in the United States, it is doubtful that this channel can be relied on as a means to prevent wrongful abortions, even if it were more attuned to the needs of punishment and deterrence.

In sum, although at first glance disciplinary proceedings might seem a promising means for deterrence of medical advisers, the problems identified above make them practically ineffective in this regard.¹⁶⁵

E. Interim Conclusion: An Interest in Search of Protection

The preceding analysis reveals a significant and disturbing anomaly in American law. On the one hand, wrongful abortions infringe the states' eminent interest in the preservation of potential life without any constitutional justification or necessity. On the other hand, the law in most jurisdictions does little (if anything) to prevent the loss of potential life in wrongful abortions and to punish those responsible for such loss. The legal response to the loss of potential life in wrongful abortions is at best scarce, and quite often absent.

We have shown that, at least in the context of wrongful abortion, tort law does not effectively protect the state interest in preserving potential life.¹⁶⁶ First, a wrongful abortion will *not* give rise to a civil action for the loss of potential life in most cases. This is because: (1) in some jurisdictions a fetus is not a "person" whose death gives rise to wrongful death and survival actions (or at least to

162. See, e.g., Ethics and Quality of Care: Report of the American Medical Association and the Federation of State Medical Boards, http://www.fsmb.org/Policy Documents and White Papers/ethics_&_quality_care.htm (last visited Aug. 31, 2004).

163. Cf. McCarthy, *supra* note 161, at 584 ("The reality is, however, that the revocation, or the suspension, of medical licenses for incompetence is extremely rare."); Peters, *supra* note 161, at 119 ("Possibly because of a lack of funds and personnel, incompetent or negligent physicians escape board sanctions.").

164. The Health Research Group, Pub. Citizen, Ranking of State Medical Board Serious Disciplinary Actions in 2002 (HRG Publication #1658), <http://www.citizen.org/publications/release.cfm?ID=7234> (last visited Aug. 31, 2004).

165. Cf. McCarthy, *supra* note 161, at 588-89 ("[S]tate licensing boards do little to assure the optimal level of quality care for patients.").

166. See *supra* Section III.B.

survival actions); (2) in others, liability is explicitly precluded in cases of abortion or death caused by medical treatment; and (3) in nearly all jurisdictions where liability is allowed, it depends on the viability of the fetus in time of its wrongful death (a serious limitation where the death is caused by a legal—albeit negligently instigated—abortion). Secondly, even if all jurisdictions allowed wrongful death and survival actions in cases of wrongful abortion, the extent of damages would not correspond to the societal value of potential life.¹⁶⁷ As we saw above, these deficiencies of tort law¹⁶⁸ are not corrected by criminal law (which is inapplicable to cases of wrongful abortion)¹⁶⁹ or disciplinary proceedings (which have an insignificant deterrent effect, at least in our context).¹⁷⁰

We have seen that tort law protects, somewhat imperfectly, the interest of the parents in not being given incorrect professional information that may lead to abortion. One may argue that the parents' right of action—standing alone—provides a good incentive for the prevention of wrongful abortions and a fair sanction upon negligent inducers of abortions, and that further penalties are not required to achieve the goals of deterrence and retribution. However, we find this argument unconvincing for at least four reasons.

First, the parents' claim is at most for their *own* personal loss, not for the full value lost to society on account of the professional negligence. A significant consequence of the wrongful act remains unaccounted for. According to traditional economic theory, efficient deterrence requires full internalization of the social costs of one's conduct, and this usually means full compensation to all victims.¹⁷¹ Although in certain cases liability for a mere fraction of the social costs of a negligent conduct, coupled with extra-legal sanctions,¹⁷² may provide an efficient incentive for potential wrongdoers, this cannot be assumed where the social cost significantly outweighs the expected liability, as is currently the case with wrongful abortions. It is quite probable that liability for a relatively small fraction of the social costs of negligent conduct will not guarantee efficient deterrence, even with the help of extra-legal sanctions.¹⁷³ We cannot say with

167. This deficiency is not special to cases of fetal death.

168. The aforementioned characteristics of the law of torts may be regarded as "deficiencies," at least from an efficient deterrence standpoint.

169. See *supra* Section III.C.

170. See *supra* Section III.D.

171. See STEVEN SHAVELL, *ECONOMIC ANALYSIS OF ACCIDENT LAW* 127-28 (1987).

172. E.g., reputational harm.

173. As indicated above, efficient deterrence requires full internalization, and this usually means full compensation. Partial compensation where extra-legal sanctions exist may be economically justified only to the extent that these extra-legal sanctions create or transfer value to people other than the wrongdoer. See Robert Cooter & Ariel Porat, *Should Courts Deduct Nonlegal Sanctions*

certainty that this would be true in all cases of wrongful abortion, but we are convinced that in many cases it would, given the gap between the social costs of wrongful abortion and the extent of tort liability under contemporary law.

Second, in considering the actual effect of potential tort liability toward the parents in cases of wrongful abortion, one cannot ignore significant "counter-incentives." These exist at least in two paradigmatic cases of wrongful abortion. Whenever an adviser provides the mother with information regarding the possible perils related to the continuance of the pregnancy, he knows that one kind of mistake is a lot more costly than another: If he mistakenly tells the mother that there is no risk, and the mother is seriously or fatally injured, he may expect an onerous personal injury action by the mother (and perhaps a loss of consortium claim by the father), or survival and wrongful death actions by *her* estate. If, on the other hand, he mistakenly tells the mother that the pregnancy is fraught with substantial risks, and the pregnancy is terminated, he may expect some liability for pain due to the abortion, mental anguish, and relatively limited pecuniary losses. In many cases, the adviser would rather make a mistake of the second type, which is usually a lot cheaper than the first. The protection of the state interest in the preservation of potential life is thereby enervated even more. Similarly, where advice is sought regarding the possibility of congenital disabilities, a mistaken diagnosis of the non-existence of disabilities may result in an onerous wrongful birth action (and in a few jurisdictions an additional wrongful life action). The adviser will have to bear not only losses related to pregnancy and delivery (pecuniary and non-pecuniary) but also damages associated with the disease, defect, or handicap suffered by the child,¹⁷⁴ and sometimes even for ordinary childrearing expenses.¹⁷⁵ Once again, the rival

from Damages?, 30 J. LEGAL STUD. 401, 405, 409 (2001).

174. See, e.g., ME. REV. STAT. ANN. tit. 24, § 2931 (West 2004), *construed in* Thibeault v. Larson, 666 A.2d 112, 115 (Me. 1995) (allowing damages associated with disease, defect, or handicap); Keel v. Banach, 624 So. 2d 1022, 1030 (Ala. 1993) (holding that parents are entitled to recover for the extraordinary expenses they incur because of the child's unhealthy condition, including: (1) hospital and medical costs, (2) costs of medication, and (3) costs of education and therapy for the child); Turpin v. Sortini, 643 P.2d 954, 965-66 (Cal. 1982) (same); Ramey v. Fassoulas, 414 So. 2d 198, 200-01 (Dist. Ct. App. Fla. 1981) (same); Blake v. Cruz, 698 P.2d 315, 320 (Idaho 1984) (same); Schroeder v. Perkel, 432 A.2d 834, 841-42 (N.J. 1981) (same); Jacobs v. Theimer, 519 S.W.2d 846, 849-50 (Tex. 1975) (same); Harbeson v. Parke-Davis, Inc., 656 P.2d 483, 492-93 (Wash. 1983) (same); James G. v. Caserta, 332 S.E.2d 872, 882-83 (W. Va. 1985) (same).

175. Cf. Ochs v. Borrelli, 445 A.2d 883, 885 (Conn. 1982) (allowing recovery for ordinary childrearing expenses in a wrongful pregnancy case); Burke v. Rivo, 551 N.E.2d 1, 5-6 (Mass. 1990) (same); Girdley v. Coats, No. 17117, 1991 Mo. App. LEXIS 1065, at *7-8 (Mo. Ct. App. July 3, 1991); Lovelace Med. Ctr. v. Mendez, 805 P.2d 603, 611-12 (N.M. 1991) (same);

incentives are not balanced.¹⁷⁶

Third, we strongly believe that the occurrence of wrongful abortions is extremely hard for the parents to detect. Detection of a wrong requires both suspicion by an interested party and availability of some evidence to the same interested party. Wrongful abortion cases raise severe problems on both levels. With regard to suspicion, we assume that if the parents had no reason to suspect that the adviser's statement was incorrect when they decided to have an abortion, they would not normally start suspecting after the abortion was carried out. Their decision implies that they utterly trusted their adviser. And if they did, something exceptional must occur before they infer that anything went wrong. After all, there is no apparent external manifestation of the wrong committed. With regard to the availability of evidence, the parents may face a serious obstacle at least in one of the paradigmatic wrongful abortion cases. Whenever the abortion is instigated by advice concerning the bodily integrity and health of the fetus, it may be very hard to test its accuracy given that the primary evidence, the fetus itself, is disposed of shortly after the abortion.¹⁷⁷ The fact that only a small percentage of wrongful abortions may be accounted for diminishes whatever deterrent effect the parents' right of action might have had.¹⁷⁸

Marciniak v. Lundborg, 450 N.W.2d 243, 245 (Wis. 1990) (same).

176. It is clear that the financial burden that may be imposed on an adviser that mistakenly informs the pregnant woman that the fetus is healthy (causing the birth of an unwanted child) is much more onerous than the one currently imposed on an adviser that erroneously advises the woman that the fetus is disabled (causing an unnecessary abortion). We suspect that advisers may consequently prefer to make an error of the second type (at least in cases of uncertainty).

177. *See, e.g.*, ARK. CODE ANN. § 20-17-802(a) (Michie 2003) (ordering the disposition of fetal remains and all parts thereof after an abortion); FLA. STAT. ch. 390.0111(7) (2004) (same); GA. CODE ANN. § 16-12-141.1 (2004) (same); MASS. GEN. LAWS ANN. ch. 111, § 202 (West 2004) (same); MINN. STAT. ANN. § 145.1621(4) (2003 West) (same); N.D. CENT. CODE § 14-02.1-09 (2003) (same). *But cf.* ARIZ. REV. STAT. § 36-2152(G)(1) (2004) (stating that fetal tissues should be preserved in cases of pregnancy resulting from incest to assist criminal investigation); MINN. STAT. § 145.1621(4) (2003) (allowing laboratory tests of fetal tissues for purposes of criminal investigation).

178. One may argue that allowing the parents to recover for their own loss in cases of wrongful abortion gives them an incentive to inquire whether their adviser was negligent, and makes detection fairly probable. However, an incentive to inquire may be effective only if there is an initial suspicion. We strongly believe that if a pregnant woman decided to undergo an abortion following negligent medical advice (believed to be accurate), she would not become suspicious after the termination of the pregnancy unless something exceptional happened. Again, if she suspected that the adviser's statement was false, she would not have undergone an abortion based on this statement in the first place. The fact that she may sue a negligent adviser would not, by itself, undermine her trust in her doctor after the abortion. In addition, an incentive to inquire does not solve the evidentiary problem mentioned above.

Moreover, even if an adviser whose statement was found to be incorrect had to bear the aggregate social costs of his or her conduct, and not merely parental losses, this might not be sufficient to guarantee efficient deterrence given that the problems of absence of suspicion and non-availability of a primary evidence would still exist. Lastly, the “counter-incentives” (e.g., personal injury and wrongful birth actions) do not raise similar problems: Personal injuries and congenital disabilities are obvious and easy to prove. This important dissimilarity strengthens the imbalance (which would have existed even without it) between the rival incentives that may affect the adviser’s statement.

Fourth, from a *retributive* perspective, the sanction provided by the parents’ claim does not match the wrong committed by the adviser on a qualitative level, and perhaps not on a quantitative level either. On the qualitative level, the sanction provided by the parents’ claim is imposed for the wrong committed against them and not for the infringement of the state interest in protecting potential life. This sends a disturbing message to the general public: The law recognizes valuable interests, but does not protect them fully and directly. It only aims at rectifying a few incidental side-effects of their infringement. On the quantitative level, the sums obtainable in the parents’ action are rather small compared with the societal value of the interest that has been negligently destroyed. It is thus probable (although not preordained) that the severity of the monetary sanction will not accord with the gravity of the wrong committed.¹⁷⁹

IV. RETHINKING THE LEGAL RESPONSE TO WRONGFUL ABORTION

A. Introduction

From our analysis of the special problems raised in the paradigmatic cases of wrongful abortion and of the current status of the law in this area we conclude that the incentives to avoid negligence are probably insufficient to guarantee the required level of deterrence. At present, in most American states a negligent physician or counselor may induce an unjustified and unwanted abortion without

179. Retributive justice does not require that the sanction imposed on the wrongdoer be equivalent to the harm she caused (as in the ancient *lex talionis*). It merely insists on proportionality between the gravity of the wrong and the severity of the sanction in light of the various features of the wrong (including the wrongdoer’s state of mind). See, e.g., Tony Honoré, *The Morality of Tort Law—Questions and Answers*, in *PHILOSOPHICAL FOUNDATIONS OF TORT LAW* 87 (David G. Owen ed., 1995). Consequently, from a retributive perspective, the sanction imposed on the negligent adviser does not have to equal the social costs incurred by her conduct. However, imposing a rather lenient sanction for a wrong that resulted in harsh consequences seems inconsistent with the idea of proportionality.

being exposed to any substantial risk of being held accountable for causing the loss of potential life, which is a significant component of the costs that such negligence inflicts upon society: not by civil law, not by criminal law, nor by disciplinary action. This reality poses a significant concern for those who recognize the need to prevent through deterrence (*ex ante*) and impose a fair sanction (*ex post*) for the considerable harm caused to society by a wrongful abortion.¹⁸⁰

In the remainder of this Article, we examine and evaluate three possible legal solutions to these concerns. After introducing the distinction between attaining deterrence through criminal punishment and attaining it through civil compensation, we first discuss the possibility of extending criminal liability to cover negligent inducement of abortion. Pointing to the problems this suggestion raises, we move on to examine the prospect of enhancing deterrence by expanding civil liability to cover the elements of the loss of potential life not currently compensable under tort law. After discussing the conceptual obstacles and pragmatic difficulties this possibility entails, we finally present our proposal for reform: a discretionary civil fine, which we find the most attractive legal solution to our problem, in terms of both fairness and efficiency.

Although we believe our recommendation is superior to the two other solutions discussed, we do not consider any of them totally implausible. Hence the critical analysis of all three solutions should not be viewed merely as a means to justify the concrete law reform we propose here. It is rather an attempt to provide decision-makers of varying ideological inclinations with a theoretical framework upon which they may debate and evaluate the appropriate solution to be adopted in their jurisdiction. Any concrete solution to the problem identified in this article should, and probably will be influenced by public opinion and the views of judges and legislators in any jurisdiction in which the matter may be subject to examination. This does not reduce the value of our framework, which we believe is applicable and relevant to any legal system that values the potentiality of human life and seeks reasonable means to ensure that it receives meaningful protection.

B. Two Deterrence Techniques: Punishment v. Compensation

According to the great utilitarians of the eighteenth and nineteenth centuries,

180. As explained above, we believe this claim to be valid even after taking into account the risk to the physician of being held liable for parental losses caused by the wrongful abortion. See *supra* Section III.E. Another possible concern may be the lack of compensation to the various victims of wrongful abortions, and perhaps even the fetus itself. However, as stated above, this article does not focus on this concern and does not attempt to resolve it. See *supra* notes 73-74 and accompanying text.

as well as their contemporary predecessors in the law and economics school, the essential purpose of criminal law and tort law is the same: to discourage undesirable forms of conduct that are detrimental to society in terms of the aggregate welfare.¹⁸¹ However, it is undeniable that even though both legal regimes impose burdens on wrongdoers and offenders, the typical technique employed by each of them is different. Generally, criminal punishment addresses wrongful conduct directly, by focusing on the defendant's acts and reacting to their inherent anti-social nature, while tort liability addresses wrongful conduct by reacting to the actual harm caused by it in the specific case. In the context of the criminal law, once a pattern of conduct defined as wrongful is identified, it is punished. The degree of the punishment is typically determined with reference to the gravity of the risk created by the defendant's act and to his or her mental state of mind, rather than to the extent of the actual injuries any specific individual has suffered as a result.¹⁸² By contrast, in the context of a civil action, once a conduct defined as wrongful or tortious is detected, it is reacted to indirectly by means of forcing the wrongdoer to repair any loss of welfare suffered by legally-recognized victims.

Which of the two traditional methods—direct punishment or compensation—is a more appropriate means to resolve the problem of inadequate protection of the public interest in preserving potential life in the context of wrongful abortions? Through which enforcement mechanism—private or public—should the law strive to secure the appropriate level of deterrence (and just retribution)? In the following Subsections we focus our inquiry on these two questions.

181. Jeremy Bentham argued that any legal sanction, be it defined as "compensation" or as "punishment," is an evil inflicted upon the defendant and therefore belongs to the "penal law." JEREMY BENTHAM, *THE LIMITS OF JURISPRUDENCE DEFINED* (Charles Warren Everett ed., 1945), reprinted in CLARENCE MORRIS, *THE GREAT LEGAL PHILOSOPHERS* 274-77 (1959). His disciple, John Austin, completely rejected the distinction between the purposes of the criminal and the civil law, and provocatively claimed that "the difference between civil injuries and crimes, can hardly be found in any difference between the ends or purposes of the corresponding sanctions." JOHN AUSTIN, *LECTURES ON JURISPRUDENCE* 520 (photo. reprint 1998) (Robert Cambell ed., 1879). The idea that civil law is no less aimed at promoting the goal of deterrence than criminal law is a fundamental tenet of mainstream law and economics scholarship today. See, e.g., RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* 209, 220 (5th ed. 1998). However, it is widely recognized by the judiciary as well. See, e.g., *Consorti v. Armstrong World Indus.*, 72 F.3d 1003, 1010 (2d Cir. 1995) ("[I]t is an aim of tort law to deter wrongful conduct . . ."); *Wash. Metro. Area Transit Auth. v. Johnson*, 726 A.2d 172, 176 (D.C. 1999) ("[O]ne aim of tort law is to deter negligent (and certainly reckless) behavior . . .").

182. This is not to say that in determining the reaction to criminal offenses criminal law ignores the specific consequences of the offense in the case at hand.

C. Deterrence Through Criminal Punishment

Criminal law is no doubt the most powerful legal means of shaping human behavior. With its typical threats of harsh stigma and incarceration, it creates significant incentives to refrain from certain acts defined by it as offenses. As shown in Section III.C above, in most jurisdictions medical advisers are not exposed to the risk of being held criminally liable for negligently causing the loss of potential life. Should this situation change? Should criminal legislation be amended the better to protect society's interest in preventing the pointless destruction of the potentiality of human life?

On the one hand, an argument in support of criminalizing such conduct should not strike us as wholly implausible. The majority of American jurisdictions recognize the value of potential human life as deserving the fierce protection of criminal law against intentional and unjustified interference, at least from viability onwards.¹⁸³ Moreover, in several jurisdictions, criminal law protects fetuses even from negligence, and sometimes even prior to viability,¹⁸⁴ although as explained above this protection does not encompass the possibility of wrongful abortion. Indeed, by penalizing acts of violence that result in the death of or bodily injury to an unborn child at any stage of development, as if the injury or death occurred to its mother, the Unborn Victims of Violence Act seems to represent a tendency to expand the legal protection of unborn children via criminal law.¹⁸⁵

If fetal life is such a valuable public interest, equivalent or almost equivalent to the public interest in protecting actual human life, may it not be argued that it deserves protection even from negligent interference, and even before viability? If negligent destruction of human life *after* birth can be criminalized, why can it not be justified to criminalize negligence when for some reason the potentiality of human life has been destroyed or endangered *before* the moment of birth?¹⁸⁶ Even under the assumption that the preservation of fetal life is of somewhat less importance to society than preserving human life after birth,¹⁸⁷ this difference

183. See *supra* Section III.C.

184. See *supra* notes 143, 149.

185. See 18 U.S.C.A. § 1841(a)(2), (d) (West Supp. 2004).

186. The arbitrariness of drawing the line between liability and non-liability based only upon the moment of birth has been recognized in case law as a reason to abandon the "born alive rule" in the context of civil liability for wrongful death. See, e.g., *Baldwin v. Butcher* 184 S.E.2d 428, 434-45 (W. Va. 1971).

187. This assumption seems to be in line with criminal statutes that set more lenient sanctions for the killing of fetuses, than those imposed for the killing of living persons. See, e.g., 18 U.S.C.A. § 1841(a)(2)(D) (West Supp. 2004) ("Notwithstanding any other provision of law, the death penalty shall not be imposed for an offense under this section.").

could be addressed by making the punishment for involuntary feticide more lenient than that for involuntary homicide. For example, instead of incarceration, a criminal fine may be imposed on the negligent offender so as to avoid disproportionate punishment that could lead to over-deterrence. In light of the problem of under-deterrence identified in Part III, it may seem appropriate to reconsider the traditional reluctance to criminalize wrongful abortions.¹⁸⁸ Denouncing ordinary negligence leading to a needless abortion as criminal would no doubt exert strong pressure on medical advisers to avoid negligence and would raise their level of care.¹⁸⁹

On the other hand, difficulties arise that we believe make the proposal of criminalization unwarranted. First, criminalizing negligent conduct is clearly problematic in terms of retributive justice. The negligent adviser has not advertently decided to disregard the rights and autonomy of the patient and her fetus. Therefore, at least according to subjectivist theories of criminal liability, she does not deserve to be convicted and punished as a criminal.¹⁹⁰ Criminalizing negligent inducement of abortion would expose physicians and genetic counselors acting in good faith and for (what they perceive as) the benefit of the biological mother (and sometimes of the unborn child as well) to the risk of being incarcerated, or at least severely stigmatized by the criminal conviction, even when their fault was only a one-time act of ordinary negligence.¹⁹¹

Second, from a social welfare perspective, and without underestimating the social value of potential human life (at any stage of development), there may be valid reasons for society to deny the life of a fetus the same protection afforded to actual human life, especially prior to viability.¹⁹² And if the social value of

188. Indeed, concerns of under-enforcement in the field of medical malpractice have led at least one scholar to argue that criminal liability should be imposed on physicians for negligent conduct resulting in death. See McCarthy, *supra* note 161, at 619.

189. See Leslie Yalof Garfield, *A More Principled Approach to Criminalizing Negligence: A Prescription for the Legislature*, 65 TENN. L. REV. 875, 914 (1998) ("While there is little retributive effect at the criminal level for defendants whose conduct involves ordinary negligence, punishment will communicate clearly to the community that such conduct is intolerable.").

190. A clear proponent of this view is Alan Brudner, who believes that "true crimes" (crimes against personality) should depend on a subjective state of mind. Alan Brudner, *Agency and Welfare in the Penal Law*, in ACTION AND VALUE IN CRIMINAL LAW 21 (Stephen Shute et al. eds., 1993). For a recent attempt to justify criminalizing negligence, see Kyron Huigens, *Virtue and Criminal Negligence*, 1 BUFF. CRIM. L. REV. 431 (1998).

191. As mentioned in Section III.D, this kind of negligence would not usually attract even a disciplinary action.

192. The fact that society does not view fetal life as equal in importance to human life, especially before viability, is apparent from the very recognition in *Roe* of the woman's right to abort her fetus at will in the early stages of her pregnancy. It is also evident in the relatively weak

potential life is indeed lower than the social value of actual life, it may be inefficient for society to criminalize wrongful abortions, even if imposing criminal liability for negligent manslaughter is justified. In the end, such a criminal reform would produce a sharp rise in the costs of providing medical or genetic advice to pregnant women and drastically reduce the supply of these services.¹⁹³ Imposing criminal liability might also have an adverse effect on the quality and quantity of medical services provided by pregnancy counselors. This could happen if advisers, in order to minimize the risk of negligent misrepresentation, communicated less frequently with their patients, and in vaguer and less meaningful terms. Assuming the services given by advisers to be a legitimate and valuable service to potential parents and society as a whole, this steep reduction might lead to a net social loss.¹⁹⁴

Unfortunately, these two problems do not seem solvable by means of reducing the criminal penalty imposed on the medical adviser for his or her negligence to only a comparatively moderate fine. The reason lies in a general weakness of the criminal sanctioning system, namely the inflexibility of criminal punishment. This inflexibility originates in the harsh and enduring stigma attached to any criminal conviction and, to a lesser extent, even to a mere indictment. A criminal record may prejudice the employment prospects of an

protection of fetuses in criminal law. *See supra* notes 140-146 and accompanying text. The attempt to provide a rational basis for this discrimination between fetal life and human life has occupied the minds of many able scholars in recent decades. *See, e.g.*, TRACIE MARTIN, INTERESTS IN ABORTION: A NEW PERSPECTIVE ON FOETAL POTENTIAL AND THE ABORTION DEBATE 59 (2000) (concluding that "the morally relevant positive causal factors for personhood are an appropriate genetic structure, sentience, consciousness, and a capacity to interact with the social world"); MICHAEL TOOLEY, ABORTION AND INFANTICIDE 87-157, 419-20 (1983) (arguing that what makes something a "person" is the fact of its being a subject of non-momentary interests, which in turn requires possessing, *inter alia*, a sense of time and a capacity for consciousness); L.W. SUMNER, ABORTION AND MORAL THEORY 142-54 (1981) (suggesting that sentience—the ability to feel and suffer harm—constitutes the most appropriate criterion according to which society's moral attitude to fetuses and to abortions should be determined).

193. Genetic counselors and physicians specializing in gynecology and obstetrics are much more exposed to the risk of negligently causing fetal death than other potential injurers by the nature of their occupation. This fact would make a criminal prohibition on involuntary feticide a much stronger threat for advisers than for any other class of potential offenders, who may only occasionally negligently endanger a woman's pregnancy (such as negligent drivers, burglars, violent people, etc.).

194. A borderline case would be the one for criminalizing grossly negligent wrongful abortions. Here the same considerations against criminalization apply, but with less force—the need for deterrence is significantly stronger. In any event, the analysis of such cases is outside the scope of this Article, mainly because the reactions of civil law, criminal law, and disciplinary law to gross negligence are entirely different.

offender for many years, and may alienate him from relatives, friends and society in general. It may lead to the revocation of certain civil rights such as the right to sit in a jury or participate in the democratic process, and in the case of a physician, may severely injure one's reputation and status within the profession.¹⁹⁵ A mere conviction would therefore impose on the negligent adviser a very significant burden, even when it is not coupled with imprisonment or a substantial fine. Such a heavy burden that may be justified in cases of intentional wrongdoing is much less deserved (in terms of retribution) and much less needed (in terms of deterrence) in cases of unintentional wrongdoing, especially those of ordinary negligence.¹⁹⁶

Moreover, any criminal liability would be imposed on the negligent adviser not instead of, but in addition to his or her civil liability toward the parents of the unborn child.¹⁹⁷ This fact only exacerbates the problems discussed above, since a threat additional to the risk of being held criminally liable would exist in the form of a damages award. As pointed out earlier, this financial threat, which is accompanied by the moderate social stigma attached to any finding of negligence, especially in the field of medical practice, may not in itself guarantee sufficient incentives to take due care and may be too lenient from a retributive perspective.¹⁹⁸ However, criminalization of wrongful abortions would increase the burden imposed on negligent advisers to an extent that would probably transform the legal response to wrongful abortions from one that is too weak to one that is too strong in terms of both fairness and social welfare.¹⁹⁹

Lastly, criminalization of wrongful abortions would be inconsistent with prevailing principles of criminal law. According to a well established common law doctrine, criminal negligence (as opposed to negligence in tort law) refers to gross negligence, and not merely ordinary negligence.²⁰⁰ Since "wrongful"

195. Commentators seem to be in agreement regarding the harshness of the criminal stigma. See, e.g., Abbe Smith, *Promoting Justice Through Interdisciplinary Teaching, Practice, and Scholarship: The Difference in Criminal Defense and the Difference It Makes*, 11 WASH. U. J.L. & POL'Y 83, 112 (2003).

196. Again, we set aside cases of grossly negligent wrongful abortions. See *supra* note 194.

197. In some states, additional claims may be brought by the estate of the unborn child, or its dependants, to recover some elements of the loss of potential life. However, as demonstrated above, these claims cover only a part of this loss, and are usually inapplicable to cases of fetal death before viability.

198. See *supra* Section III.E.

199. Theoretically, these concerns exist whenever a criminal act unlawfully injures a private interest protected by civil law. However, since criminal liability is usually based on intentional wrongdoing, the fear of over-deterrence and unfair penalty is much less significant in most of these cases.

200. *State v. Weiner*, 194 A.2d 467, 470 (N.J. 1963) (defining criminal negligence as involving

abortions are typically induced by acts of ordinary negligence, criminalizing them would amount to a radical reform that may produce instability and incoherence within the criminal law.²⁰¹

D. Deterrence Through Compensation

Having rejected the proposal to increase deterrence directly through expansion of criminal responsibility, we may now proceed to examine the other most common legal vehicle for behavior control, namely tort law's traditional compensation mechanism. It may be argued that the best way to resolve the problem of under-deterrence exposed in this Article would simply be to make the negligent adviser accountable for the entire loss he or she has caused. This would consist not only of parental losses but any social loss originating in the wrongful abortion, including all the elements of the loss of potential life which at present are non-compensable under most tort law regimes.²⁰² Put differently, the adviser would be forced to internalize, through damage awards to his or her victims, any externality his or her negligence has imposed on society.²⁰³ Such a move would presumably remove the existing disproportion between the social costs of wrongful abortions and the legal burden imposed in reaction to them. It would also eliminate, or at least considerably reduce, the imbalance between the legal reaction to wrongful abortions on the one hand, and the legal reactions to personal injuries or wrongful birth on the other.

Unfortunately, as will be demonstrated below, tort law's compensation mechanism cannot provide an adequate solution to the problem. This is mainly because as opposed to a criminal sanction, which is *defendant-oriented*, civil law's typical sanction of compensatory damages is *plaintiff-oriented*.²⁰⁴ This

utter disregard for the safety of others); see also Garfield, *supra* note 189, at 890 n.56.

201. This does not preclude the possibility of imposing an administrative fine outside the realm of criminal law. This possibility is discussed in Section IV.E.

202. For a discussion of these elements, see *supra* Part I.

203. ROBERT COOTER & THOMAS ULEN, *LAW AND ECONOMICS* 290 (3d ed. 2000) ("The economic purpose of tort liability is to induce injurers to *internalize* [externalities that are not internalized through private agreements] . . . by making the injurer compensate the victim.").

204. This basic difference between the two sanctions derives from the fact that criminal law technique responds to anti-social conduct independently of its detrimental consequences, while tort law responds to those very consequences, and attempts to repair the harm actually suffered. See *supra* Section IV.B. A sanction is "plaintiff-oriented" if it is designed with an eye to its effects on the plaintiff, and "defendant-oriented" if it is designed with an eye to its effects on the defendant. For a similar use of these terms see, for example, David W. Leebron, *The Right to Privacy's Place in the Intellectual History of Tort Law*, 41 CASE W. RES. L. REV. 769, 809 (1991) (distinguishing between "the plaintiff-oriented goal of compensation and the defendant-oriented goal of deterrence").

characteristic of the civil sanction is manifested in the two most basic principles governing its award. The first principle is that the availability of damages depends on—and is limited by—the existence of a recognized victim, i.e., an aggrieved party who has suffered compensable loss through the defendant's wrongful conduct. The second principle is that the scope of compensatory damages is determined—hence limited—by the extent of the actual injury the wrongdoer has inflicted on his or her victims.²⁰⁵ In the following Subsections, we shall argue that each of these two limiting principles poses serious difficulties, which in our context make the expansion of civil liability problematic as a vehicle for securing appropriate levels of deterrence.

1. Lack of a Recognized Victim

Any proposal to expand the civil liability of negligent advisers to cover additional elements of the loss of potential life not at present compensable faces a major obstacle: the lack of a recognized legal entity that could be viewed as the sufferer of these losses.

Unlike losses suffered by the unborn child's parents (which we labeled "parental losses"), other losses ensuing from wrongful abortions are suffered by a variety of entities (the fetus, the state, businesses, and other persons) none of which is generally recognized by the law as a victim deserving of compensation. As we shall show shortly, this fact may impede traditional tort law's ability to force the negligent physician to fully internalize the social costs of his or her conduct.

a. Fetal Loss

In Section I.B we defined fetal loss as the social cost caused by preventing a healthy and desired fetus from being born, and thus denying it the ability to acquire pecuniary and non-pecuniary benefits from birth to death. However, as pointed out, under most tort law regimes the typical fetus in a wrongful abortion case is not considered a legal person at the time of its wrongful death. Hence, it

205. Many tort law scholars view these principles as reflecting the principle of corrective justice, while others view them as mere means to achieve desired social goals such as compensation, deterrence, economic efficiency, distributive justice, loss spreading, or any combination of them. Whatever the correct view, the fact that these principles govern and limit the operation of positive tort law is rarely disputed by contemporary scholars. See, e.g., COOTER & ULEN, *supra* note 203, at 291 ("We discuss the traditional theory [of tort law] because the essential elements of a tort as stipulated by it [i.e., harm to the plaintiff, breach of duty on the part of the defendant, and causal link between the breach and the harm] serve as building blocks in the economic model of tort liability.").

may not recover any damages for the loss of life it would have enjoyed had it not been wrongfully aborted. Surely, no other entity may be entitled to claim damages for this very loss, since no one other than the fetus has suffered it. If that is the case, and no person in the legal sense has suffered this fetal loss, how can civil law possibly assist in forcing the negligent adviser to internalize it? At first sight this may seem an insurmountable obstacle. But we can think of at least two distinct ways whereby a legal system may bypass this doctrinal obstacle, and thus force injurers to internalize fetal losses.

First, the law may adopt a straightforward technique and explicitly recognize the fetus at any stage of gestation as a legal person for the sake of extracting compensatory damages from a wrongdoer (other than the biological mother) for the loss of its potential life.²⁰⁶ As we have seen, many jurisdictions allow the estate of an unborn child to bring suit in its own name (a survival action) for the losses it suffered prior to and upon its death.²⁰⁷ True, in the majority of jurisdictions that recognize such claims, they are limited to the death of viable fetuses, and even then, as demonstrated above, not all elements of the fetal loss are compensable.²⁰⁸ However, a system keen to making a negligent person who wrongfully caused fetal death accountable for the social costs of his or her undesirable conduct may seek to abandon, or at least relax these traditional limitations. It might allow a non-viable fetus to recover for the loss of its life potential either by legislative reform or by a more liberal judicial approach to interpreting the term "person" in the relevant statutory provisions.

Changes in this direction have been evinced in recent years in several jurisdictions in which an unborn child has been declared, either through legislative action or by means of judicial interpretation, a legal "person" from the very moment of its conception.²⁰⁹ Could such a move affect the constitutional

206. It is beyond the scope of this Article to discuss the doctrinal problems that may arise in a damages claim for the loss of fetal life due to a wrongful abortion. We are quite confident that most of these problems (mainly estimation of the pecuniary and non-pecuniary value of lost life, and legal causation) are not unique to our context, and may be addressed by reference to the same principles that are applied in other claims originating in wrongful death, wrongful life, and wrongful birth.

207. See *supra* Subsection III.B.1.

208. See *supra* notes 98, 129 and accompanying text.

209. In South Dakota, for example, in the context of a claim for wrongful death, the Supreme Court ruled in 1996 that S.D. CODIFIED LAWS § 21-5-1 clearly meant to include non-viable children in the term "unborn child." Furthermore, the majority of the Court expressed their view that the very concept of viability was outmoded in tort law and was a purely arbitrary milestone from which to reckon a child's legal existence. *Wiersma v. Maple Leaf Farms*, 543 N.W.2d 787, 791 (S.D. 1996). For another sharp judicial criticism of the viability test see, for example, *Farley v. Sartin*, 466 S.E.2d 522, 533 (W. Va. 1995):

right of pregnant women to choose abortion at the early stages of their pregnancy? On the one hand, it may be submitted that the call to relax the viability requirement in tort law may be seriously considered even by states in which a comparatively liberal abortion regime prevails. The argument would probably be that recognition of a non-viable fetus as a person for the purpose of granting it the protection of tort law against interference by third parties would not undermine the right to abort, as recognized in *Roe* and its progeny. That is so mainly because protecting the fetus from being killed (or injured) by third parties, not in the course of a consented abortion, fulfills the mother's true will (i.e., to give birth) and thereby reinforces her procreative autonomy. Recognizing the *legal* personhood of a fetus for the sake of its protection by tort law does not amount to recognizing its *constitutional* personhood, and therefore cannot affect the mother's constitutional right.²¹⁰ As for the third party, her constitutional rights are not violated by obliging her to refrain from wrongfully causing fetal death.²¹¹ Not recognizing the duty of a negligent third party to the non-viable fetus would grant that party unjustified immunity from civil liability for negligently harming a fetus, an immunity that only the mother of the fetus should possess.²¹² Even if such immunity were justified in the context of our discussion

In our judgment, justice is denied when a tortfeasor is permitted to walk away with impunity because of the happenstance that the unborn child had not yet reached viability at the time of death. The societal and parental loss is egregious regardless of the state of fetal development. Our concern reflects the fundamental value determination of our society that life—old, young, and prospective—should not be wrongfully taken away.

See also *Gentry v. Gilmore*, 613 So. 2d 1241, 1246 (Ala. 1993) (Maddox, J., dissenting). In Missouri, MO. REV. STAT. § 1.205.2 (2004) explicitly provides that “the laws of this state shall be interpreted and construed to acknowledge on behalf of the unborn child at every stage of development, all the rights, privileges, and immunities available to other persons, citizens, and residents of this state.” This legislation is wide in its scope and applies both to criminal and civil law. In California, a fetus is considered a “person” for the purpose of homicide offenses from the end of the embryonic stage, *People v. Davis*, 872 P.2d 591 (Cal. 1994), but not for the purpose of a civil action, *Justus v. Atchison*, 565 P.2d 122 (Cal. 1977).

210. This point is eloquently explained by Rosen, *supra* note 5.

211. A similar point is made by Meade, *supra* note 72, at 444-45. Indeed, the same separation had been actually implemented in the context of the criminal law. As we have seen, many states have criminalized the intentional or even negligent killing of a fetus when committed by a third party, but not when committed by the biological mother of the fetus or on her behalf, during a legal abortion procedure or other medical treatment intended to protect the mother's life. See, e.g., CAL. PENAL CODE § 187(b)(1)-(3) (West 2004); TEX. PENAL CODE ANN. § 19.06(1)-(4) (Vernon 2004).

212. This logic is echoed in the majority opinion in *Wiersma*, 543 N.W.2d at 791 (“If we accept [the defendant's] argument, someone could fatally injure an unborn child by a nonconsensual, wrongful act and still avoid civil liability because the child was not yet viable. This would, ironically, give the tortfeasor the same civil rights as the mother to terminate a pregnancy.”).

with respect to criminal proceedings,²¹³ we believe that it should not extend to the civil sphere.

On the other hand, although a distinction between *legal* personhood and *constitutional* personhood may be drawn, it might be difficult to maintain in the specific context discussed here. It may be argued that a widespread recognition of the legal personhood of fetuses at any stage of development reflects a change of value in American society that necessitates a reassessment of the delicate constitutional balance established in *Roe* and modified by *Casey*. This is a possibility that supporters of abortion-rights might be concerned about.

Assuming now that certain legal systems would be reluctant to assign legal personhood to the unborn child at any stage of gestation, a different technique should be considered for better internalization of the social costs of wrongful abortions. Such a goal may be achieved through recognition of the state as the sufferer of the fetal loss, and accordingly granting it the right to bring a civil suit for the pecuniary and non-pecuniary elements of that loss.²¹⁴ For any person who is interested in improving the legal protection of potential human life without supplying ammunition to either side of the abortion debate, this solution may seem preferable to the one discussed above, since it does not grant legal personhood to the unborn child.

However, such a proposition raises two major difficulties. Both originate in the fact that, unlike the ordinary context in which the state claims damages in a civil suit, here it would be demanding compensation for the loss caused neither to an asset in its possession nor to its legally recognized economic interest (such as a contractual or other obligatory right). Rather, it would be seeking indemnification for the social loss manifest in the destruction of a fetus, whose existence has never before been recognized as the state's private or personal interest. It may therefore be argued that notwithstanding society's undisputed interest in preserving the potentiality of life latent in a fetus, this interest is a pure public interest. As such, it should be protected exclusively through public law devices (criminal law, administrative law) rather than through a private action in torts, which by definition requires the plaintiff to prove that his or her *private* right has been violated.²¹⁵ Furthermore, leaving this formal line of argument

213. See *supra* Section IV.C.

214. The various elements of fetal loss are discussed *supra* Section I.B.

215. JOSEPH CHITTY, A TREATISE ON PLEADING AND PARTIES TO ACTIONS 1 (1867), cited with approval by *In re African-American Slave Descendants Litig.*, 304 F. Supp. 2d 1027, 1045 (N.D. Ill. 2004) ("The action for a tort must in general be brought in the name of the person whose *legal* right has been affected, and who was *legally* interested in the property at the time the injury thereto was committed; for he is impliedly the party injured by the tort, and whoever has sustained the loss is the proper person to call for compensation from the wrongdoer."); see also *Tyler v. Judges of Ct. of Registration*, 179 U.S. 405, 407 (1900).

aside, it may be claimed that no theoretical or moral basis exists that is capable of justifying such an extension of the existing rights of the state.

While both objections seem to be valid, we do not view them as insurmountable obstacles to the adoption of the aforementioned proposal. Let us start with the more substantive objection. We submit that the theoretical basis of the state's claim for compensation in this case lies in both the necessity to vindicate the social value of potential life and deter wrongful interference with it, and in the state's unique position as the classic representative of the public interest. Under the assumption that a non-viable fetus is not a legal person, the loss of fetal life caused by a wrongful abortion may not be attributed to the fetus itself. Nor may it be attributed to the unborn child's parents, who as private individuals maintain the right to claim damages only for their own private losses. The state, being the ultimate representative of the public interest, differs in this context from any other individual, including the parents. From a moral point of view, society as a whole (the state being merely its legal representative) may, in certain cases, rightfully demand recognition as the residual victim of any wrongful injury which cannot be viewed as the private and personal loss of any specific individual. As such, the state should be regarded as a direct, rather than an indirect victim of such losses, and should be entitled to compensation for them, as if it were their direct bearer. Fetal loss, which denies the unborn child the benefits and pleasures of life without enabling it to recover anything, is a clear example of such a case. The recognition of the state's right to be compensated for this loss would force the negligent adviser to take into account not only the risk of causing harm to the fetus's parents, but the loss to the fetus itself as well.²¹⁶

It may still be contended that the task of responding to anti-social conduct has in modern times traditionally been assigned mainly to criminal law. However, not every moral or social wrong should attract criminal liability. Some wrongs, especially those committed unintentionally, do not usually justify the imposition of a criminal sanction. Assuming negligence causing fetal death to be one such wrong, the social value of fetal life may not be vindicated at all, absent the possibility to impose civil liability on the party negligently causing it.²¹⁷

216. Arguably, this construction may also apply to certain relational losses originating in the loss of the fetus, as long as these losses are not attributable to any specific person in the legal sense. However, to recover damages for any such loss, the extent of the loss must be approximately calculable. As we shall see, most relational losses originating in the loss of potential life do not lend themselves to any such calculation. *See infra* note 238 and accompanying text.

217. As one court once put it in another context, "[i]f a child . . . has no right of action . . . we have a wrong inflicted for which there is no remedy." *Montreal Tramways v. Leveille*, [1933] 4 D.L.R. 337, 345 (Can.), *cited with approval in Bonbrest v. Kotz*, 65 F. Supp. 138, 141 (D.D.C. 1946). These words were said in support of recognizing the right of a child born with physical

Indeed, although comparatively rare, a number of contexts exist in which the law in modern times, in order to solve a unique enforcement problem, grants a private cause of action to an entity which is not the actual victim of the wrong complained of. A famous example is the common law doctrine of *parens patriae* (literally "parent of the country"), under which the government may represent all of its citizens and act on their behalf as a trustee of the public in a private suit involving a matter of sovereign interest.²¹⁸ This old doctrine has been utilized in modern times to allow the state to file civil suits for infringement of social interests that would otherwise remain unprotected.²¹⁹ It may be argued that this

defects to claim compensation for prenatal injury that had caused these defects. Although different, this situation resembles ours in that when the case was tried it was not clear whether an act injuring a fetus could be considered negligent for the sake of a damage claim filed by the child subsequent to her birth. By upholding this possibility, the court constructed a new cause of action, and established a new wrong, in order to vindicate society's need to deter acts injurious to it. Disciplinary action is another possible legal response to medical malpractice, but as shown earlier, it is doubtful whether it can be relied upon in the paradigmatic cases on which this Article focuses. See *supra* Section III.D. The possibility of a wholly new administrative enforcement mechanism is considered in Section IV.E.

218. *Late Corp. of the Church of Jesus Christ of Latter-Day Saints v. United States*, 136 U.S. 1, 57 (1890) ("This prerogative of *parens patriae* is inherent in the supreme power of every State [I]t is a most beneficent function, and often necessary to be exercised in the interests of humanity, and for the prevention of injury to those who cannot protect themselves."). In the past, the doctrine has been applied mainly for the protection of juveniles and incompetent persons. See, e.g., Neil Howard Cogan, *Juvenile Law, Before and After the Entrance of "Parens Patriae,"* 22 S.C. L. REV. 147 (1970). For a general survey of the historical origins and modern development of the doctrine, see George Curtis, *The Checkered Career of Parens Patriae: The State as Parent or Tyrant?*, 25 DEPAUL L. REV. 895 (1976).

219. In *Alfred L. Snapp & Son, Inc. v. Puerto Rico ex rel. Barez*, 458 U.S. 592 (1982), the Commonwealth of Puerto Rico brought suit in its capacity as *parens patriae* against a number of private American employers for discriminating against Puerto Rican migrant farmworkers by subjecting them to burdensome working conditions and improperly terminating their employment. These acts allegedly violated the relevant federal statutes and regulations and injured the Puerto Rican economy. Recognizing the right of Puerto Rico to demand a declaratory judgment and an injunction, the Supreme Court clarified that the "concept does not involve the State stepping in to represent the interests of particular citizens who, for whatever reason, cannot represent themselves . . . if the State is only a nominal party without a real interest of its own—then it will not have standing under the *parens patriae* doctrine." *Id.* at 600. However, when the state seeks to protect a "quasi-sovereign" interest, i.e., an interest "that the State has in the [physical or economic] well-being of its populace," as had been the case at hand, the court may, in appropriate cases, apply the doctrine of *parens patriae* to vindicate that interest. *Id.* at 601. Further guidelines for the implementation of the *parens patriae* doctrine were developed by lower courts. For an elaborate analysis and a useful survey of cases where this doctrine was applied, see *Massachusetts v. Bull HN Info. Sys., Inc.*, 16 F. Supp. 2d 90, 96-98 (D. Mass. 1998).

doctrine enables the courts to recognize the state's right to protect the social interest in preserving potential life through a civil action in torts.

Another judicial technique implemented at times to overcome problems of insufficient protection of an important social value is the somewhat artificial expansion of the category of victims recognized as entitled to compensation under a given statute. A good example in this context is the judicial recognition of employees as victims of anti-trust violations when they were wrongfully discharged by their employers because of their cooperation with the anti-trust authorities. The language of the relevant provision of the federal anti-trust legislation limits the right to recover treble damages under the statute to persons "injured in their business or property" by anti-trust violations.²²⁰ Yet some courts have adopted an extremely liberal interpretation of this provision, and have extended it to include these "indirect victims" in order to improve enforcement in this field.²²¹ Admittedly, this technique does not apply directly to our situation, since in our case no general statutory cause of action allows victims of acts injuring fetuses to be compensated. However, these examples may serve as a source of inspiration to the legislature when contemplating the creation of a new civil cause of action to the state.

The public interest in preserving and vindicating potential life may indeed receive recognition through a specific legislative effort, which will grant the state the right to claim compensation for fetal loss in cases where no other legal entity may do so. In recent years, this vehicle has been increasingly adopted in the context of environmental law in order to protect natural resources.²²² For instance, in order to remedy a serious problem of under-enforcement in this field,

220. 15 U.S.C. § 15 (2000).

221. For a comparative survey of a line of relevant cases, see Sean P. Gates, *California Antitrust: Standing Room for the Wrongfully Discharged Employee?*, 47 HASTINGS L.J. 509 (1996). Other examples of this sort exist in other contexts as well. In *Trafficante v. Metropolitan Life Insurance Co.*, 409 U.S. 205 (1972), the Supreme Court recognized the right of a white resident of an apartment complex to sue the owner for loss of interracial associations under Title VII of the Civil Rights Act of 1968, even though the discriminatory rental practices were not directed at him, and were not alleged to have caused him any economic loss. In recent years, this liberal approach to the interpretation of the Civil Rights Act has been applied in the context of discrimination in the workplace to allow white claimants to sue employers for discriminating against their black co-workers. See Joseph C. Feldman, *Standing and Delivering on Title VII's Promises: White Employees' Ability To Sue Employers for Discrimination Against Nonwhites*, 25 N.Y.U. REV. L. & SOC. CHANGE 569 (1999).

222. See, e.g., Scott Kerin, *Alaska Sport Fishing Association v. Exxon Corporation Highlights the Need To Take a Hard Look at the Doctrine of Parens Patriae when Applied in Natural Resource Damage Litigation*, 25 ENVTL. L. 897 (1995).

the federal Oil Pollution Act (OPA)²²³ created new civil causes of action that would have been hard to construct under ordinary principles of tort law. Under OPA, a person in charge of a facility from which oil is discharged is liable in damages for various social costs caused by such discharge to natural resources.²²⁴ Inter alia, the statute nominates public trustees and grants them unique standing to recover damages from the polluter for any “injury to, destruction of, loss of, or loss of use of, natural resources.”²²⁵ The nature of the interest protected by the OPA is obviously different from the interest in preserving fetal life. However, its enforcement technique and rationale bear a striking resemblance to those suggested with regard to fetal death. In both cases, a public authority is allowed to recover, in the name of the public interest, damages for an injury caused not to its own property or even to its own economic welfare, but to an object (a natural resource in one case, a fetus in the other) in which the public holds no recognized proprietary interest, but which nevertheless is regarded by the legal system as valuable and deserving legal protection.²²⁶

To conclude, either by extending the concept of a “direct victim” under the conventional analysis of tort law (thus making the negligent adviser liable for having violated a duty of care towards the state) or by creating a specific

223. 33 U.S.C. §§ 2701-2761 (2000).

224. *Id.* § 2702(a)-(b).

225. 33 U.S.C. § 2702(b)(2)(A). Section 2702(b)(2)(C) goes even further, allowing any person injured by the harm to a natural resource to recover damages for loss of subsistence use of that resource, “without regard to the ownership or management of the resources.” Section 2702(b)(2)(E) even allows damages for loss of profits and earning capacity. For comprehensive surveys of the OPA legislation and case law see, for example, J.T. Smith II, *Natural Resource Damages Under CERCLA and OPA: Some Basics for Maritime Operators*, 18 TUL. MAR. L.J. 1 (1993); and Steven R. Swanson, *OPA 90 + 10: The Oil Pollution Act of 1990 After Ten Years*, 32 J. MAR. L. & COM. 135 (2001).

226. One may contemplate a similar development in the field of animal rights. To date, unlawful infliction of pain on an animal is not considered a violation of a private right of any person. Therefore, no one is entitled to compensation for the pain and suffering experienced by the animal. In our view, it is not unimaginable to acknowledge the right of a public authority nominated by statute to demand compensation for such “private” losses in cases where absent such recognition society’s interest in protecting the autonomy and the bodily integrity of certain animals would not be sufficiently vindicated. See, e.g., Robert Garner, *Political Ideology and the Legal Status of Animals*, 8 ANIMAL L. 77, 87 (2002) (“If we accept an animal welfare position, whereby animals matter morally but not as much as humans, the harm principle can be adapted to take into account the fact that harm inflicted on animals which can be shown to serve significant human benefits, is regarded as legitimate, but that harm which is unnecessary to further human interests is ruled out.”). Similar views were expressed in a recent symposium concerning the legal status of chimpanzees. Symposium, *Ten Years of Animal Law at Lewis & Clark Law School: The Evolving Legal Status of Chimpanzees*, 9 ANIMAL L. 1 (2003).

statutory cause of action, recognition of the state's right to receive compensation for the loss of welfare manifested in the wrongful prevention of the realization of potential human life does not seem to us theoretically impossible.²²⁷ Whether such techniques will guarantee an efficient level of deterrence is a different question, to be examined in Subsection IV.D.2 below.²²⁸

Regardless of everything said so far, we admit that one major theoretical obstacle might still inhibit some jurisdictions from adopting any of the propositions offered above. Traditionally, and to a great extent even today, judges and commentators have viewed tort law not only as a vehicle to promote deterrence, but also, if not primarily, as a means to provide compensation to real people for real losses they have suffered through the commission of a wrong.²²⁹ If a legal system does not genuinely accept the idea that a fetus—viable or non-viable—actually suffers loss by being wrongfully denied the opportunity to be born or, alternatively, that the state actually suffers loss by being deprived of one of its future members, it is doubtful that such a legal system will be keen to use the legal phenomenon known as “tort law” or even other civil law mechanisms (such as a statutory cause of action for compensation) when there is no perceived real victim that will truly benefit from such a compensatory award. Therefore, notwithstanding our efforts to overcome the major obstacle discussed in this Subsection (i.e., the lack of a recognized legal entity that could be viewed as the sufferer of the fetal loss), for some jurisdictions it may simply be too high.

227. We assume here that an estimation of the pecuniary and non-pecuniary losses of the potential life of a person, though difficult, is a task the civil courts are capable of carrying out in this context, as well as in other contexts where damages are awarded for this loss (in survival and wrongful death actions).

228. A final remark concerns the interrelations between this action for compensation and the one discussed in the previous Subsection (the fetus as a victim). As we have seen, most states recognize a fetus as a “person” in the context of wrongful death and survival statutes only after reaching viability. On the other hand, the suit in the name of the public interest discussed here is independent of the fetus's stage of development. This is so because society's loss of welfare reflected in the loss of potential life exists independently of whether the fetus is viewed by society as human enough to suffer its own loss of potential life. The scope of application of this cause of action is therefore much wider than the first. However, in cases of overlap, it should be clear that the private action of the fetus should bar the private or public action of the state for repairing the same loss, so as not to allow double recovery.

229. *Bennis v. Michigan*, 516 U.S. 442, 469 n.216 (1996) (Stevens, J., dissenting) (“Tort law is tied to the goal of compensation (punitive damages being the notable exception).”); *Oden v. Chemung County Indus. Dev. Agency*, 661 N.E.2d 142, 145 (N.Y. 1995) (supporting the proposition that “just compensation is the main end toward which tort law is directed”).

b. Relational Losses

As pointed out earlier, apart from fetal losses, which are the direct consequence of a wrongful abortion, many people may suffer pecuniary and non-pecuniary losses following fetal death.²³⁰ First, close relatives may incur various losses. As we saw above, some of the parental relational loss is compensable under traditional principles of tort law, whereas the recoverability of other parental relational losses and relational losses incurred by other existing relatives (siblings, grandparents, etc.) depends on the availability of a wrongful death action and its statutory scope. At present, the inapplicability of most wrongful death statutes to deaths of non-viable fetuses (or any fetal death), together with the very modest list of recognized dependants, and the fact that not all types of relational losses are covered, make these statutes ineffective as means to internalize familial relational losses.

As stated above, an argument could be made for recognizing a fetus as a legal person for the sake of bringing suit against the negligent adviser for fetal loss. If this recognition of the fetus as a "person" from the start of the pregnancy is applied to wrongful death statutes as well, an action may be brought by its statutory dependants against the negligent adviser.²³¹ Furthermore, even if the fetus is not recognized as a legal person, it is possible to allow its existing legally recognized dependants to sue for their losses. After all, wrongful death statutes were intended to compensate certain relatives for the loss of a valuable relation. The losses that these statutes were intended to redress do not depend on the legal status of the direct victim of the wrong, but on its value to the survivors.²³² If they can prove this value, we see no reason to deny them the right to be compensated for its loss. Just as a person should be entitled to compensation for the wrongful destruction of her property (e.g., objects, plants, animals), she should be allowed to claim compensation for the loss of a relationship, as long as her relationship is recognized by society as important enough to deserve legal protection. True, the common law has been traditionally hostile to the idea of compensating relational losses.²³³ However, just as this hostility has not prevented the enactment of wrongful death statutes, it should not necessarily bar the legislature from granting the dependants of a wrongfully aborted fetus the

230. See *supra* Section I.B.

231. For an endorsement of such an argument, see Meade, *supra* note 72.

232. For example, let us assume a brother or sister of the fetus would benefit economically and mentally from the birth of the fetus in due time. It is hard to see why the decision of whether to compensate the sibling for this loss should depend on whether or not the fetus had reached viability on its death.

233. See *supra* note 47.

right to bring suit, even in cases where the fetus was not viable on its death.²³⁴

The two other features of wrongful death statutes that diminish their ability to effectuate internalization of familial relational losses (the limited statutory list of recognized dependants and imperfect compensation) are not unique to cases of fetal death. They must be dealt with, if at all, within a broader consideration of statutory reform.

Fetal death may also give rise to many other relational losses. These include: (1) loss of income to the state (and other governmental authorities) from future taxation of the revenues and other taxable activities of the unborn child during a normal lifespan, minus any benefit that would have been directly conferred by the state upon that child; (2) widespread losses of economic and non-economic benefits to various persons and economic bodies that could have interacted with the unborn child during her lifetime, had she not been wrongfully aborted; and (3) widespread outrage and sorrow suffered by people who learn about the occurrence of wrongful abortions.²³⁵ Can civil liability be expanded to guarantee internalization of these categories of relational loss?

In answering this question, we think that the first loss (of taxes) is distinguishable from the two other groups of relational loss. This loss may be incurred by a few specific entities, namely the state or other tax authorities. In addition, assuming that a civil court may reach a reasonable estimate of the prospective revenues of an unborn child during her lifetime, calculating the state relational loss should seem equally possible, at least with regard to income tax. Given these two characteristics of the loss of future income tax to the state, the fear of unlimited and indeterminate liability, which is one of the primary policy arguments against tort liability for pure economic loss,²³⁶ does not seem to apply. Hence, it does not seem farfetched to contemplate a statutory provision or even a judicial decision recognizing the right of the state's treasury to be compensated for this loss in a case of wrongful abortion (as in any other case of personal injury or wrongful death). Just like the right of existing dependants to recover damages in cases of wrongful death of a non-viable fetus, the right of the state in this case should not depend on the legal status of the fetus itself. Whether the fetus is a legal person or not, the dependants and the state alike deserve compensation for the loss *they* have suffered due to its wrongful death, as long as such loss may be proved with reasonable certainty.

234. Technically, this could be done by amending wrongful death statutes so as to make compensation available for the death of a fetus at any stage of gestation.

235. We mentioned these various losses in Section I.B.

236. *See, e.g.,* *Ultramares Corp. v. Touche*, 174 N.E. 441, 444 (N.Y. 1931) (allowing claims for pure economic loss may expose the wrongdoer to liability "in an indeterminate amount for an indeterminate time to an indeterminate class").

In any case, the fact that the state relational loss is not currently compensable under tort law does not necessarily undermine tort law's ability to effectuate its internalization. An internalization of this loss will be achieved if the estate of the unborn child is compensated for her loss of income without the deduction of expected taxes.²³⁷ Allowing the estate to recover lost income measured by gross earnings is therefore an alternative means for internalization of state relational loss by the tortfeasor. From an economic perspective, it does not matter who receives the award. What matters is that the injurer bears the cost.

However, the situation with regard to the two other types of relational loss mentioned above is wholly different. Although the occurrence of these widespread losses is most probable, the range and identity of persons suffering them is unknown and therefore their extent is also impossible to estimate, even roughly.²³⁸ Given the speculative nature of these relational losses, their estimation would be wholly arbitrary, thus useless in terms of internalizing the social costs of wrongful abortions. Therefore, although in theory one may contemplate granting the state a right to recover for such unidentified relational losses, it is doubtful that in practice such right could or should be recognized.

2. The Inability of Civil Law To Warrant Accurate Internalization

Having discussed the problems arising from the need to identify a recognized legal victim, and having offered some initial guidelines for their solution, we may now examine whether expanding civil liability according to the guidelines offered in the previous Subsection would provide a satisfactory solution to the current lack of sufficient legal protection of the public interest in preserving potential life.

Our answer to this question is, unfortunately, negative. Expanding civil liability to cover fetal loss and other measurable relational losses (such as loss to dependants other than the parents) will certainly raise the level of deterrence exerted on potential negligent advisers. However, as we shall contend below, there is no reason to assume that this rise will provide, even approximately, the required amount of deterrence.

237. When the tax is not deducted, the injurer is forced by the legal system to internalize this social cost, in addition to the cost reflected in the plaintiff's private loss. For a survey of the different approaches applied by case law in this context, see John E. Theuman, Annotation, *Propriety of Taking Income Tax into Consideration in Fixing Damages in Personal Injury or Death Action*, 16 A.L.R.4th 589 (1982).

238. This derives from the fact that except for a small number of close relatives, no person can prove with reasonable certainty either that he or she would have interacted with the unborn child had it not been wrongfully aborted, or the extent to which he or she would have benefited from such interaction.

According to the classical economic theory of civil deterrence, the goal of civil liability rules is to oblige injurers to pay *ex post* for the entire social costs of their acts, thereby making them internalize *ex ante* the risks inherent in their behavior. Thus, it is argued, the right to receive compensatory damages for a wrong committed maintains an optimal level of deterrence and prevents injurers from acting inefficiently.²³⁹ However, many scholars, among them legal economists, have pointed out the unrealistic assumptions underlying this theoretically ideal model.²⁴⁰ In the context of our discussion, it is enough to point out a few important obstacles that undermine the ability of the courts, in assessing compensatory damages for fetal losses and relational losses, to internalize properly the social costs of wrongful abortions.

First, due to problems of information, a court's ability to estimate with reasonable proximity the extent of the losses caused by any wrongful act is in many cases fairly limited.²⁴¹ The problems of information and assessment are

239. Although refined by an extensive legal literature, this assumption is still adhered to by many contemporary law and economics scholars. See, e.g., Robert D. Cooter, *Three Effects of Social Norms on Law: Expression, Deterrence, and Internalization*, 79 OR. L. REV. 1, 16 (2000) (stating that "deterrence typically requires the injurer to internalize the harm that he caused"); Keith N. Hylton, *Punitive Damages and the Economic Theory of Penalties*, 87 GEO. L.J. 421, 421 (1998) (same); A. Mitchell Polinsky & Steven Shavell, *Punitive Damages: An Economic Analysis*, 111 HARV. L. REV. 870, 873 (1998) (same). The principle is assumed to be valid under the liability regimes of both negligence and strict liability. However, its validity under the former is subject to a few conditions. Polinsky & Shavell, *supra*, at 878-87. For a more technical presentation, see Steven Shavell, *Strict Liability Versus Negligence*, 9 J. LEGAL STUD. 1 (1980).

240. For powerful criticisms see, for example, Robert L. Rabin, *Deterrence and the Tort System*, in SANCTIONS AND REWARDS IN THE LEGAL SYSTEM: A MULTIDISCIPLINARY APPROACH 79, 94 (M. L. Friedland ed., 1989) ("[T]he efficacy of tort law as a deterrence strategy is open to serious question [T]he system almost certainly needs to be broadly reassessed."); Israel Gilead, *Tort Law and Internalization: The Gap Between Private Loss and Social Cost*, 17 INT'L REV. L & ECON. 589 (1997) (arguing that the gap between private loss and the social cost of unintentional wrongdoing prevents tort law from assuring proper internalization); Stephen D. Sugarman, *Doing Away with Tort Law*, 73 CAL. L. REV. 555 (1985) (arguing that the costs of the tort system outweigh its benefits in monetary terms); Izhak Englard, *The System Builders: A Critical Appraisal of Modern American Tort Theory*, 9 J. LEGAL STUD. 27, 56 (1980) ("[A]ttempt[ing] to achieve a pure economic goal within the traditional framework of tort law . . . appears a hopeless endeavor.").

241. Coase observed in his writings that proposals suggesting the creation of systems of internalization "are the stuff that dreams are made of." R. H. COASE, *THE FIRM, THE MARKET, AND THE LAW* 185 (1988). In addition, it is most unclear whether, and to what extent, potential wrongdoers are able to estimate the scope and extent of the risks created by their behavior. For an analysis of this problem see, for example, Howard A. Latin, *Problem-Solving Behavior and Theories of Tort Liability*, 73 CAL. L. REV. 677, 682-88 (1985); Sugarman, *supra* note 240, at 565-69.

most acute with regard to non-pecuniary losses and relational losses in general, many of which are indeterminate in scope and size. In the context of our discussion, this concern is most manifest with regard to the non-pecuniary elements of the fetal loss (loss of enjoyment of life) and to the widespread relational losses deriving from fetal death. This concern makes the prospect of internalization problematic. In some cases it may lead to over-deterrence, while in others it may lead to under-deterrence, due to either under-estimation or over-estimation of the monetary value of the loss sustained by every victim (the fetus, the state, a dependant, etc.).

Second, under most tort law regimes there will always remain injuries (mainly widespread relational injuries, and sometimes direct non-pecuniary injuries as well) that will not be compensated by the legal system. While part of this problem (such as judicial reluctance to award compensation for some non-pecuniary losses) may be soluble to some extent, a major part of it may not be resolved, due to the problems of information discussed above. This reality may result in under-deterrence of potential wrongdoers.²⁴²

Third, in its original form, the model of economic deterrence ignores problems of under-enforcement, which enable many wrongdoers totally or partially to escape liability. As pointed out earlier, in the context of wrongful abortion this problem is most acute due to problems of detection, which clearly lead to under-deterrence.²⁴³ The common solution to this problem offered by the law and economics literature is to multiply the damage award reflecting the full loss caused by the defendant by the reciprocal of the probability of liability.²⁴⁴ However, at least in the context of a wrongful abortion, this solution may be unsatisfactory since the ratio of undiscovered wrongful abortions is unknown and very difficult to assess. Absent an empirical study of the frequency of wrongful abortions, implementing the multiplier method would require a "guess," leading once again to under-deterrence or over-deterrence. Furthermore, this solution totally ignores the fact that under prevailing judicial practices, the defendant's civil liability (as well as the plaintiff's right to compensation) must be based on

242. This insight seems to have been first developed in the law and economics literature in an attempt to justify the award of punitive damages. See Dorsey D. Ellis, Jr., *Fairness and Efficiency in the Law of Punitive Damages*, 56 S. CAL. L. REV. 1, 28-31 (1982).

243. See *supra* Section III.E. In addition, in some cases the plaintiffs (the parents, the fetus's estate, or the state) may fail to prove that the abortion was negligent, even when it actually was.

244. For example, if the loss equals \$1 million, and if the probability of liability is 1:4 (0.25), the original award of \$1 million must be multiplied by the reciprocal of 0.25, i.e., four, to reflect the fact that only one out of four times are wrongdoers brought to justice and found liable. Hence, the appropriate damage award should be fixed at \$4 million. This model was first introduced in Robert D. Cooter, *Punitive Damages for Deterrence: When and How Much?*, 40 ALA. L. REV. 1146, 1148 (1989). It was adopted and elaborated further in Polinsky & Shavell, *supra* note 239, at 874.

hard evidence rather than on mere speculation about the probability of other wrongs committed in other cases (and not proven before the court). Therefore, we do not see how courts can implement such a recommendation, without revolutionizing the basic principles of civil liability.²⁴⁵

Fourth, the phenomenon of liability insurance, common in cases of medical malpractice, significantly reduces the deterrent effect of civil liability and undermines the ability of compensatory damage awards to guarantee complete internalization.²⁴⁶ Disregarding other factors, the problem of insurance will clearly lead to under-deterrence of wrongful abortion claims.

Fifth, the traditional economic theory of internalization does not take into account the extra-legal incentives operating on potential wrongdoers to avoid negligence, such as inner morality and public censure.²⁴⁷ Given these incentives, which in the case of a physician who provides medical services are significant²⁴⁸ (though not in themselves sufficient), it is reasonable to assume that an award of damages covering the full social cost of the negligent act would eventually provide at least in some cases—more deterrence than is actually needed.²⁴⁹ However, once again, the monetary equivalent of such informal incentives is very hard for a court to assess, and therefore this concern of under-/over-deterrence may not be easily resolved.²⁵⁰

245. Indeed, to the best of our knowledge, no civil court has yet agreed to implement this suggestion or even seriously considered it.

246. John G. Fleming, *Is There a Future for Tort?*, 44 LA. L. REV. 1193, 1197 (1984) (“[T]he admonitory effect of an adverse judgment is today largely diffused by liability insurance which protects the injurer from having to pay the accident cost . . .”). The “bonus-malus” system of adjusting premiums to the insured’s liability record may reduce this problem to a certain extent, but there is no reason to assume that it totally cures it. For an analysis of the influence of insurance on the ability of the tort system to deter wrongdoing, see Sugarman, *supra* note 240, at 573-81.

247. For an attempt to analyze these factors and their possible impact on the economic theory of deterrence see Cooter, *supra* note 239. An extensive inquiry into the influence of social norms on human conduct and the need to use legal enforcement mechanisms has been undertaken in ERIC A. POSNER, *LAW AND SOCIAL NORMS* (2000).

248. The legal incentive for a physician to refrain from acting negligently is significant because doctors (like many other professionals) are usually very sensitive to the stigma of an adverse judgment and because they are assumed to have a moral obligation to their patients.

249. This point has been made by Robert Cooter and Ariel Porat, who argued that “deducting nonlegal sanctions typically reduces social costs by improving the incentives of wrongdoers and victims.” Cooter & Porat, *supra* note 173, at 420.

250. Robert Cooter and Ariel Porat suggested that in order to take into account this factor courts should ideally deduct the monetary value of non-legal sanctions from any award of compensatory damages. However, they admitted that “[t]he precise extent of the typical deduction is unknown because so little research measures nonlegal sanctions.” *Id.* Moreover, established principles of tort law require the tortfeasor to pay full compensation to his victim, regardless of any incentive

In our view, these major difficulties make it presumptuous, if not naïve, to assume that expanding the scope of civil liability of negligent advisers to cover fetal loss, as well as some of the relational losses discussed above, will provide, even approximately, the right additional amount of deterrence needed to correct the deterrence failure and the incentive imbalance identified in Part III. Some of the problems identified above enhance the probability of under-deterrence (the second, third, and fourth); the others either enhance the probability of over-deterrence (the fifth) or equally enhance the probability of both (the first). So it would be very hard to speculate in advance about whether, on average, expanding the civil liability of a negligent adviser who has induced a wrongful abortion will lead to more or to less than the required amount of deterrence.

These problems are serious. Yet they could be resolved, or at least significantly relaxed, if the civil courts had the flexibility to adjust the size of the compensatory award to the need (or lack of need) for deterrence, the existence of which would be determined in the specific circumstances of the case adjudicated. However, under entrenched principles of civil liability, the extent of the defendant's liability is determined solely with reference to the plaintiff's legally recognized loss. If such loss has been proven, the defendant is liable for the full extent of it and the court holds no power to either remit or augment the damage award in order to achieve better the goal of deterrence (or punishment).²⁵¹ This characteristic of civil liability creates a problem of inflexibility, different from the one characterizing the criminal sanction. While the latter's inflexibility lies in its harsh stigmatizing effect,²⁵² the former's inflexibility lies in its complete subordination to the compensatory measure.²⁵³ This traditional mode of reaction,

operating on him or her (legal or non-legal). Therefore, similar to the concern of under-enforcement, we can hardly envisage a court deducting non-legal sanctions from damages awards for the sake of improving deterrence, as Cooter and Porat proposed.

251. Nevertheless, many years ago one commentator argued that civil courts actually do take into account the need to admonish the defendant (or the plaintiff) while assessing compensatory damages. Ralph S. Bauer, *The Degree of Moral Fault as Affecting Defendant's Liability*, 81 U. PA. L. REV. 586 (1933); see also Ralph S. Bauer, *The Degree of Defendant's Fault as Affecting the Administration of the Law of Excessive Compensatory Damages*, 82 U. PA. L. REV. 583 (1934). Recently, a similar contention has been sounded with regard to the award of restitutionary remedies. Andrew Kull, *Restitution's Outlaws*, 78 CHI.-KENT L. REV. 17, 18 (2003) ("[R]estitution does punish, but it punishes negatively: not by imposing liability on disfavored parties. . . but by denying a restitutionary claim (or counterclaim) to which the disfavored party would otherwise be entitled.").

252. See *supra* note 195 and accompanying text.

253. This subordination is usually taken for granted in the law and economics literature analyzing the deterrent effect of compensatory damages. When economists have put forward recommendations that contradict this subordination (a good example being the reciprocal principle,

being *plaintiff oriented* rather than *defendant oriented*,²⁵⁴ prevents standard civil liability regimes from properly adjusting their remedial response to the need to maintain efficient levels of deterrence.²⁵⁵

To conclude, even if the proposal to expand civil liability of negligent advisers is adopted, it is submitted that such an expansion may not provide the legal response required to maintain the desired level of deterrence.

E. The Proposed Solution: A Discretionary Civil Fine

Is there any legal mechanism that may resolve the problem of under-deterrence of wrongful abortions in a better way than either the criminal law or the civil law? Our answer is simple: Yes. A statutory discretionary civil fine, imposed within the framework of a civil suit brought by a parent of the wrongfully aborted fetus, will do the job. The authority to impose this civil penalty, and the maximum and minimum sums to be imposed, would be specifically provided by statute. The monies of the fine should in our view be divided, according to the provisions of the statute, between the parent(s) and the state's treasury. The fine would be imposed only after it had been proven before the court, by clear and convincing evidence, that the defendant's negligent misrepresentation to the unborn child's mother led to its wrongful abortion, and only after the defendant had been given a reasonable opportunity to convince the court that the imposition of the fine was inappropriate in the circumstances.

A number of guidelines directed us to this solution. First, we were looking for a sanction whose recognition and implementation would provide the public interest in preserving the potentiality of human life, at every stage of the pregnancy, with *much greater protection* than the one currently available in most American legal jurisdictions.

Second, we were looking for a sanction which although punitive and deterrent, would be considerably *less severe than any criminal sanction*, including a criminal fine, so as to relieve us from the concerns of over-detering and unduly punishing physicians.²⁵⁶

Third, and most important, we were looking for a sanction which, unlike the traditional criminal and civil sanctions, would be *flexible* and *adjustable* to the

discussed *supra* note 244), their suggestions have been usually ignored by judicial practice.

254. See *supra* note 204 and accompanying text.

255. Indeed, law and economics scholars today explicitly admit that "Internalization . . . is not the proper goal when perfect compensation is impossible in principle or in practice . . . or when enforcement errors systematically undermine liability. In these circumstances, law's proper goal is deterrence . . . punishments are calibrated to deter those actors who prefer to do the act in spite of its price." COOTER & ULEN, *supra* note 203, at 434.

256. In this regard see *supra* Section IV.C.

concrete circumstances of any case in which it might be imposed. Such flexibility should manifest itself in the ability to influence the extent of the burden imposed on the negligent adviser and the very decision on whether or not in the specific circumstances its imposition is required at all (given the deterrent effect of other sanctions imposed for the same act).

Fourth, we were looking for a sanction that would meet directly and expressly, rather than indirectly and implicitly, society's need for symbolic vindication of the value of potential life (rather than other interests, such as the parents' physical and emotional integrity) from interference by third parties other than the fetus' biological mother.

Fifth, we prefer a legal strategy that would not give academic ammunition to either of the opposing camps in the abortion debate, and that might be acceptable to both.

Sixth, and finally, we were looking for a legal mechanism whose administration would be comparatively cheap and efficient, but that at the same time would not violate the defendant's constitutional right to "due process of law."

Given these guidelines, we believe it is not difficult to see the advantages of our proposal. In the following Subsections, we wish to emphasize these advantages and to address some possible objections and concerns our proposal may raise. Structurally, we shall start the discussion by pointing out the significant advantages of a civil fine over the two solutions examined in the two previous sections of this part. Next, having discussed the possible drawbacks of the civil fine, we examine the possibility of overcoming these weaknesses with the aid of a parallel public enforcement mechanism (civil or administrative). After showing the pros and cons of this suggestion, we explain why we think it should be rejected. However, towards the end of this section we put forward the possibility of expanding disciplinary law—one form of an administrative enforcement mechanism—to make possible the imposition of disciplinary fines on medical advisers whose negligence resulted in unwarranted abortions.

1. The Advantages of a Discretionary Civil Fine

The use of civil penalties in different common-law and statutory formats to achieve the goal of deterrence is far from new in American jurisprudence, and its constitutionality had been affirmed many times by the Supreme Court.²⁵⁷

257. See, e.g., *Tull v. United States*, 481 U.S. 412, 423 (1987) (holding that the district court may aim to deter violations of the Clean Water Act by basing penalties on economic impact). Well-known examples of federal legislation providing for civil penalties are the anti-trust statute known as the Clayton Act, the Racketeer Influenced and Corrupt Organizations Act (RICO), and the False

Nonetheless, our proposal is innovative mainly because it advocates implementing a statutory civil penalty upon a finding of a violation not of a public-statutory norm, but of a private law duty, namely the duty of care imposed by tort law on the physician towards the pregnant woman (and her partner).²⁵⁸

A discretionary civil fine possesses several characteristics that in our view make it superior to the proposals hitherto examined. First, granting a court adjudicating the parents' case against the negligent adviser the authority to impose a civil penalty would enable it, in the appropriate cases, to couple the standard compensatory award (for causing parental losses), which in many cases would not provide sufficient deterrent incentives,²⁵⁹ with an additional punitive sanction. This additional sanction would impose on the negligent adviser a significant monetary burden, whose full deterrent and retributive effects would not be diminished by the prospect of liability insurance. Contrary to ordinary compensatory judgments, such insurance should be—and usually is—disallowed with respect to civil (as well as criminal) penalties.²⁶⁰ In addition, the civil fine

Claims Act (FCA). The Clayton Act allows any person allegedly injured in his or her business or property from a violation of anti-trust laws to claim treble damages for his or her loss plus litigation costs. 15 U.S.C. § 15(a) (2000). RICO includes a similar provision. 18 U.S.C. § 1964(c) (2000). Under the *qui tam* provisions of the FCA, 31 U.S.C. §§ 3729-3730 (2000), a private citizen can recover treble damages in a civil action brought on behalf of the government against a party that allegedly made a false claim for payment against the United States. The lion's share of the penalty is paid to the government, but the plaintiff is entitled to a share of twenty-five to thirty percent of the proceeds, or, if the action was proceeded by the state, to a share of fifteen to twenty-five percent. Punitive damages are a judge-made civil penalty. Unlike most statutory fines, they are not subject to a stringent cap. "However, in practice, few awards exceeding a single-digit ratio between punitive and compensatory damages . . . will satisfy due process." *State Farm Mut. Auto. Ins. Co. v. Campbell*, 538 U.S. 408, 425 (2003). Note that punitive damages are imposed only for "conduct that is outrageous, because of the defendant's evil motive or his reckless indifference to the rights of others." RESTATEMENT (SECOND) OF TORTS § 908 (1977). As such, they do not apply to wrongful abortions, as defined in Section I.A.

258. The reform proposed here is also unusual because it will enable the courts to impose a punishment for an act of negligence (as opposed to intentional misconduct). Punishing negligence is unusual not only in criminal law but also in the realm of civil fines. However, the imposition of penalties—criminal or civil—for acts of negligence is not unprecedented in American law and has resisted constitutional attacks in various contexts. For a useful survey of this area, see A. Dale Irlie III, Comment, *Parental Delinquency: Should Parents Be Criminally Liable for Failing To Supervise Their Children?*, 74 U. DET. MERCY L. REV. 93, 106-10 (1996).

259. See *supra* Section III.E.

260. Many liability insurance policies exclude coverage for civil fines. This is done in order to prevent the phenomenon of "moral hazard," which is most acute where the conduct leading to liability is intentional. Even if insurance companies allowed coverage for civil fines, such coverage would probably be deemed contrary to public policy, because it would mitigate the punitive effect

proposed would impose on the adviser a non-monetary burden in the form of a moderate stigma, which necessarily attaches to any civil fine or penalty.

Second, though not at all insignificant, the stigma attached to a civil fine is far less severe and harmful to the negligent physician than the one attached to a criminal fine imposed following a criminal conviction.²⁶¹ Hence, the specter of an exaggerated chilling effect on pregnancy advisers and genetic counselors, as well as a disproportionate punishment in terms of just desert, would be avoided.²⁶²

Third, and quite important, the discretionary nature of the civil fine proposed would enable the civil court to use it if, *and only if*, the court was convinced that the need to vindicate the social worth of potential life and to deter its wrongful termination in the future would be frustrated absent its imposition. Put differently, unlike the traditional civil and criminal sanctions, the imposition of a civil fine would not automatically attach to a finding of liability, but would be

of the sanction, and frustrate, at least to some extent, the goals of deterrence and retribution. *See, e.g.*, I.R.C. § 6672(a) (2000) (civil fine provision); *Mortenson v. Nat'l Union Fire Ins. Co.*, 249 F.3d 667, 672 (7th Cir. 2001) (“[I]nsurance against the section 6672(a) penalty, by encouraging the nonpayment of payroll taxes, is against public policy, so falling . . . under the rule in Illinois as elsewhere that forbids certain types of insurance as being against public policy because of the acute moral hazard that the insurance creates.”); *see also In re Tex. E. Transmission Corp.*, 870 F. Supp. 1293, 1338 (E.D. Pa. 1992) (“[T]he inability to enforce its laws by the assessment of civil penalties may well hamper its ability to force compliance without resorting to criminal or other more severe sanctions.”); *Blair v. Anik Liquors*, 510 A.2d 314 (N.J. Super. Ct. Law Div. 1986) (holding that indemnification for fines resulting from violation of Alcoholic Beverage Control laws violates public policy).

261. *See supra* note 195 and accompanying text; *cf.* Comment, *Criminal Safeguards and the Punitive Damages Defendant*, 34 U. CHI. L. REV. 408, 411 (1967) (“The money judgment assessed against the punitive damages defendant hardly seems comparable in effect to the criminal sanctions . . . [T]he [criminal] fine, unlike punitive damages, still carries the full weight of stigma associated with criminal conviction.”).

262. One may argue that given the exceptional use of civil fines as a legal response to negligence, the negligent adviser will be hit with an excessive stigma that may lead to over-deterrence. However, we do not believe that allowing courts to impose civil fines for negligent inducement of abortion may result in over-deterrence due to the resulting stigma. First, it is reasonable to assume that the stigma attached to any fine imposed for an act of negligence would, by the very nature of the act, be much weaker than the one attached to any fine imposed for intentional wrongdoing (e.g., punitive damages). Second, if the court in a specific case believes the fine might lead to over-deterrence or undue punishment (due to the consequent stigma) it may legitimately decide to refrain from imposing it. After all, the proposed fine is discretionary. Even if this happens frequently, the awareness among professionals of the risk of being punished with a civil fine may be sufficient to raise their level of care.

subject to the wide discretion of the court.²⁶³ This feature is most important as it enables the court to refrain from complementing the parent's damage award with any additional sanction, when such a supplement seems unnecessary given the extent of the defendant's liability towards the parents. Similarly, the court may decide not to resort to a civil fine if the defendant's deviation from the standard of care was slight. Special attention should also be paid to any other liabilities incurred (or expected to be incurred) by the physician for the same act following any criminal, disciplinary, or civil action.²⁶⁴

Fourth, and closely related to the previous point, like other criminal and civil fines, but unlike the award of civil damages, the amount of the civil penalty we propose would be flexibly determinable by the court *ad hoc* (subject, perhaps, to a statutory cap and some general legislative guidelines). It would be fixed with direct reference to the goals of deterrence and retribution, taking into account all considerations that may be relevant for the realization of these two goals.²⁶⁵ Central considerations would be the degree of the physician's negligence; the extent to which the physician's overall conduct reflects serious attempts to minimize error; any evidence of prior medical malpractice; the stage of gestation at which the negligence occurred; the physician's economic situation, her reputation and professional status; and any other legal sanction (civil, criminal, administrative, or disciplinary) that was or may be imposed for the same act.²⁶⁶

263. Cf. *United States v. Reader's Digest Ass'n.*, 662 F.2d 955, 967 (3d Cir. 1981) ("[A]ny penalty actually imposed by a district court would be subject to the limitation of judicial discretion."); *AFL-CIO v. FEC*, 628 F.2d 97, 100 (D.C. Cir. 1980) ("[T]he District Court is vested with discretionary authority in the imposition of civil penalties.").

264. As argued in the previous parts of this Article, in most cases of wrongful abortion we believe that none of these procedures would be available. However, they may become available if any of the proposals examined in this article is adopted. For example, the extent of tort liability depends, *inter alia*, on the reach of wrongful death and survival actions in the relevant jurisdiction. Naturally, if the ideas discussed in Section IV.D (regarding the possible extension of tort law to cover most of the social cost of wrongful abortion) are implemented in the future, the need for a civil fine may abate.

265. Concerns of just desert (retribution) may legitimately influence the court's decision, and in the paradigmatic case of wrongful abortion would probably reduce, rather than enhance, the final size of the penalty. See *supra* notes 190-191 and accompanying text. As in other contexts where punishment is inflicted, this result seems to us not only inevitable, but perfectly legitimate, as long as deterrence concerns are given their due weight, side by side with concerns of just retribution.

266. A legislative attempt to outline the relevant considerations in the award of a civil penalty was made by Congress in 33 U.S.C. § 1319(d) (2000), which provides:

In determining the amount of a civil penalty the court shall consider the seriousness of the violation or violations, the economic benefit (if any) resulting from the violation, any history of such violations, any good-faith efforts to comply with the applicable requirements, the economic impact of the penalty on the violator, and such other matters

Another important factor would be the probability of detecting the adviser's negligence. Although the relevant data may not be available to the courts, it may be collected on behalf of the legislature, and taken into account in determining the upper limit of the fine or the statutory guidelines for its imposition. The size of the fine to be imposed in a concrete case, as well as the maximum penalty to be defined by the statute authorizing its imposition, need not be very high. This follows from the fact that in the paradigmatic wrongful abortion case the defendant's fault reflects ordinary negligence, and is perpetrated with no bad faith on his or her part.²⁶⁷ Given the adverse effect of the stigma attached to both the finding of civil liability toward the parents and to the imposition of the extra-compensatory civil fine, the sum needed to deter potential wrongdoers from encouraging wrongful abortions should be relatively low, especially if the compensatory award imposed on the physician is not trivial.²⁶⁸

Fifth, like the typical sanctions of the criminal law, the fine proposed here would satisfy the social need to vindicate the value of potential life more directly and more explicitly, hence more effectively. Rather than influencing the behavior of negligent advisers indirectly, as would the expansion of ordinary civil liability for compensatory damages discussed in Section IV.D, the imposition of the fine proposed, or merely enactment of the statute allowing its use in cases of wrongful abortion, would have a significant symbolic value, which in and of itself may have an educational value and thus a preventive effect.²⁶⁹

as justice may require.

For a survey of case law implementing this provision, see Erin Belka & Sarah Kern, *Assessing Civil Penalties in Clean Water Act Citizen Suit Cases*, 10 HASTINGS W.-NW. J. ENV. L. & POL'Y 71 (2003).

267. As noted earlier, in cases of gross negligence the imposition of the fine would probably be unnecessary, given the possibility of awarding punitive damages against the physician for his or her reckless indifference to the safety of the pregnant woman and her fetus. In addition, the prospect of being exposed to suspension or even revocation through disciplinary proceedings is much higher in this case, as well as that of being charged with involuntary manslaughter (or even homicide) in jurisdictions where such offenses are applicable to wrongful abortion cases.

268. The exact size of the penalty needed to achieve an optimal level of deterrence with regard to a certain type of misconduct is always a puzzle. We do not pretend to solve this well-known difficulty in this Article, or to point out the way for legislators and judges to do so. We only assert our belief that a flexible system of penalties which is defendant-oriented (focusing on the gravity of the wrong from a wide social perspective) is generally a better means for attaining the "correct result" in terms of deterrence and retribution than a relatively inflexible system of penalties that is plaintiff-oriented (focusing only on the negative effects of the wrong on the welfare of an individual).

269. Although frequently neglected in the academic literature, education is one of the most important goals of punishment. See, e.g., ALFRED C. EWING, *THE MORALITY OF PUNISHMENT* 73-125 (1970); WALTER H. MOBERLY, *THE ETHICS OF PUNISHMENT* 78 (1968); Jean Hampton, *The Moral*

Sixth, compared with the two other solutions presented above, the costs of administering this new sanction would be minimal. Unlike a *criminal* fine, the proposed method would not necessitate the instigation of a wholly separate and costly legal process by the state, either criminal or administrative.²⁷⁰ The imposition of the civil fine would be considered within the context of an already existing civil suit, and only after a finding of negligence on the part of the defendant. This factor makes the proposed mechanism much more economic even compared with the alternative examined in Section IV.D (expanding civil liability), since the adoption of the latter would similarly necessitate the instigation of separate civil proceedings (by the unborn child's estate, by the state, and/or by other persons).²⁷¹ The most significant extra cost required in order to employ the proposed sanction would be the additional effort by the parties to convince the court that the defendant's negligence had or had not been proven with the more stringent evidentiary standard of clear and convincing evidence, which we regard as required in light of the punitive character of the proposed sanction.²⁷² However, it is submitted that in most litigated cases, parties are willing to make the best effort to prove their case, so this additional burden should not be too substantial. Moreover, the additional procedure would most probably take place during the presentation of the parties' claim for damages, so it would not ordinarily require a significant extension of the proceedings.

Seventh, as opposed to punitive damages and several statutory civil penalties that go entirely to the plaintiff's pocket, if our proposal is adopted, a substantial part of the fine would be paid to the state's treasury.²⁷³ Given that the interest to be protected and vindicated by this penalty (namely society's interest in

Education Theory of Punishment, 13 PHIL. & PUB. AFF. 208, 212 (1984).

270. The fact the fine is imposed within a civil suit instigated and conducted by a private party is also useful because it relieves the legal system of the need to provide for procedural guarantees that are typical of a criminal process. Nevertheless, being punitive in nature, the proposed procedure requires certain procedural guarantees. *See infra* Subsection IV.E.2.a.

271. In addition, unlike a typical civil process, the assessment of a fine does not require any technical work of calculating and proving the various heads of damage.

272. This evidentiary standard is usually applied in cases where the alleged wrongful conduct was fraudulent. *See, e.g.*, 19 U.S.C. § 1592(e)(2) (2000) ("[I]f the monetary penalty is based on fraud, the United States shall have the burden of proof to establish the alleged violation by clear and convincing evidence."); *Fairchild v. Comm'r*, 462 F.2d 462, 463 (3d Cir. 1972) (same). In cases of negligence, the ordinary standard of "preponderance of the evidence" is usually applied. Nevertheless, we think that the punitive character of most civil penalties justifies a more stringent evidentiary standard.

273. The exact division of the fine between the state and the plaintiffs does not seem essential to us, as long as it achieves its underlying goals.

preserving potential life) is mainly a public interest,²⁷⁴ this arrangement seems only natural and will avoid the undesirable consequences of bestowing too large a windfall on the parents.²⁷⁵ However, to encourage the parents of the unborn child to bring suit, and to make the extra effort to prove the adviser's fault with clear and convincing evidence, it is submitted that a reasonable part of the award should go into the parents' pockets if they succeed in providing such evidence. This allocation is crucial if the statute does not give the parents, as "private attorney[s] general," an explicit right—similar to that given in other private enforcement statutes—to recover for the full legal costs of bringing and conducting the suit.²⁷⁶

However, granting the parents the right to share the proceeds of the fine may be justified even if they are indemnified for their legal costs (including attorneys' fees) when they succeed in proving the defendant's fault with clear and convincing evidence. On the level of efficiency, expected legal expenses in case of failure plus other costs and burdens (monetary and non-monetary) that are not regarded as "legal expenses," and are thus non-recoverable, may outweigh the expected liability. This is especially true where, as in our case, non-recoverable detection costs are high, litigation is an extremely harrowing experience for the plaintiffs, and there is reasonable likelihood that the action will either fail or result in partial compensation and a relatively small award. These factors may eliminate or reduce the parents' incentive to bring suit in the first place. We do not ignore the parents' non-monetary incentive to bring those responsible for the loss of their future offspring to justice, and to see them punished by the official authorities of the state. This incentive may in some cases encourage the parents to put time and effort into establishing their civil cause of action, even if the expected monetary benefit is lower than the expected cost and trouble. However, in other cases, the parents' drive to retaliate will not be strong enough to outweigh the totality of the abovementioned factors, so without an additional pecuniary reward for their effort the public interest may be frustrated.

274. This does not mean that it is impossible to recognize private interests (of the fetus itself, the state, or others) in the materialization of the life potential of the fetus. The appropriate legal means for vindicating those interests would be a civil action. This possibility was discussed in Section IV.D.

275. The concerns over the creation of windfalls to plaintiffs have usually been dealt with in the context of punitive damages, where the plaintiff is normally entitled to receive the entire punitive award. These concerns have led several states to adopt legislation that curtails the plaintiff's share, and allocates parts of the punitive award to the state, or to another nominated public authority.

276. The FCA, for example, provides that "[a]ny such person [substantially contributing to the prosecution of the Act] shall also [apart from his share in the fine] receive an amount for reasonable expenses . . . plus reasonable attorneys' fees and costs. All such expenses, fees, and costs shall be awarded against the defendant." 31 U.S.C. § 3730(d)(1) (2000).

Moreover, under the proposed method the parents are those responsible for the vindication of the public interest in preserving potential life. Their efforts may save a substantial amount of public resources that would have otherwise been spent in an alternative public procedure. In terms of fairness, it does not seem counter-intuitive or farfetched to suggest that they are entitled to a substantial reward for the public service they rendered society.²⁷⁷

Eighth, and finally, our proposal attempts to protect the public interest in preserving potential life without recognizing or making use of private rights (other than those of the parents). In particular, it is not based on the assumption that a fetus is a legal person, and therefore may not be used to question the relatively entrenched opinion of the Supreme Court that a fetus is not a constitutional person prior to viability. Consequently, the proposed scheme does not jeopardize pregnant women's right to terminate their pregnancies, to the extent that it is recognized under the Constitution. At the same time, this Article does not express any opinion about the appropriate boundaries of the right-to-abort. After all, prevention of *wrongful* abortions seems to be a mutual interest of both the pro-life and pro-choice movements.

2. Drawbacks of the Proposal and Possible Ways To Resolve Them

Like any other legal tool, the civil fine is not a perfect mechanism and has its own points of weakness. What are they? Is there a way to prevent them or at least reduce their negative influence?

a. The Conceptual Problem

First, we should consider the general conceptual problem of civil punishment. Is it legitimate to punish a defendant within the context of a civil action, without affording her the procedural safeguards available to the criminal defendant? We believe this first concern should not be given much weight, and in any event should not lead to the rejection of our proposal. Since the days of Blackstone, the role of punishment, at least in theory, seems to have been allocated to the criminal law and excluded from private law.²⁷⁸ However, time-honored exceptions to this principle are recognized in the form of punitive

277. A similar argument was used, quite convincingly in our view, to justify the award of the punitive damages to the plaintiff. *See, e.g., Neal v. Newburger Co.*, 123 So. 861, 863-64 (Miss. 1929); Gregory S. Pipe, *Exemplary Damages After Camelford*, 57 MOD. L. REV. 91, 99 (1994).

278. 3 WILLIAM BLACKSTONE, COMMENTARIES *2 ("Wrongs are divisible into two sorts or species: *private wrongs*, and *public wrongs*. The former are an infringement or privation of the private or civil rights belonging to individuals . . . the latter are a breach and violation of public rights and duties, which affect the whole community . . .").

damages, multiple damages, civil penalties, forfeitures, and other forms of civil and administrative punishment.²⁷⁹ To be sure, the question of what exact procedural safeguards a defendant facing a punitive-civil sanction should be granted has never been easy.²⁸⁰ It should be addressed in our context, as in others, so as not to violate the constitutional rights of defendants.²⁸¹ Yet such difficulties have generally been found to justify neither the complete abolition of most forms of civil punishment, nor an automatic incorporation of any procedural safeguard employed in criminal law into the civil punitive mechanism.²⁸² In our case, while the punitive character of the legal response we propose is undeniable, we believe that the fact of its being administered in a civil action brought by a private party militates against applying most of the stringent procedural safeguards of the criminal process.

Still, we do believe that at least two procedural guarantees should be granted to the defendant in a wrongful abortion case once the possibility of a civil fine is introduced. First, the defendant must be explicitly warned, either by the plaintiff

279. For critical surveys of the development of American civil law to incorporate and enhance the scope of such sanctions, see Kenneth Mann, *Punitive Civil Sanctions: The Middleground Between Criminal and Civil Law*, 101 YALE L.J. 1795, 1844-61 (1992); and Franklin E. Zimring, *The Multiple Middlegrounds Between Civil and Criminal Law*, 101 YALE L.J. 1901 (1992).

280. For illuminating analyses of the distinction between civil and criminal sanctions and procedures see Susan R. Klein, *Redrawing the Criminal-Civil Boundary*, 2 BUFF. CRIM. L. REV. 679 (1999); and Mary M. Cheh, *Constitutional Limits on Using Civil Remedies To Achieve Criminal Law Objectives: Understanding and Transcending the Criminal-Civil Law Distinction*, 42 HASTINGS L.J. 1325 (1991).

281. In *Kennedy v. Mendoza-Martinez*, 372 U.S. 144 (1963), and *United States v. Ward*, 448 U.S. 242 (1980), the Supreme Court made it clear that the procedural safeguards of the criminal process may apply to an administrative action, if the purpose of such action is punitive. In *United States v. Halper*, 490 U.S. 435 (1989), the Court held that the Double Jeopardy Clause of the Fifth Amendment applies to a civil-punitive process instigated by the state and prevents the imposition of a civil penalty whenever a prior criminal fine has been imposed for the same act. However, to date, these precedents have not been interpreted to apply when the punitive sanction is imposed in a civil procedure instigated by a private party (e.g., an action for punitive damages). Assuming our conclusions in Sections III.C and III.D are valid, the risk of double jeopardy does not usually exist in a case of wrongful abortion since no other punitive process (criminal or administrative) is available against the negligent defendant.

282. The question of what constitutional safeguards should apply to a punitive-civil action has been largely debated in the context of punitive damages. For detailed accounts see Joseph J. Chambers, In Re Exxon Valdez: *Application of Due Process Constraints on Punitive Damages Awards*, 20 ALASKA L. REV. 195 (2003) (concluding that the Supreme Court's jurisprudence will require additional guidelines in order to settle various due process issues); and Malcolm E. Wheeler, *The Constitutional Case for Reforming Punitive Damages Procedures*, 69 VA. L. REV. 269 (1983).

in her statement of claim, or by the court itself (if the initiative to impose the fine comes from the court), of any intention to consider the imposition of a fine, and must accordingly be given a fair opportunity to convince the court that its imposition would be inappropriate or unnecessary in the circumstances of the case. Second, as mentioned above, the plaintiffs should be required to prove the defendant's negligence by a higher evidential standard than the ordinary "preponderance of evidence" standard. While there is no justification to require a standard of beyond reasonable doubt, the demand that the evidence be "clear and convincing" seems to us apposite, given the punitive nature of the fine and the detrimental effect of the judicial condemnation latent in its imposition.²⁸³

b. The Pragmatic Problem

A second and seemingly more disturbing difficulty concerns the fact that the imposition of the proposed fine wholly depends on the initiation of a civil action by the unborn child's parents against the negligent physician and its continuation to a successful end. This seems to undermine the public interest in responding to wrongful abortions in a significant number of cases. These are of two main types: (1) cases where the parents' financial, physical or mental condition cause them to avoid filing a suit in the first place and (2) cases where the parents have reached a settlement with the physician or the medical institution in which he or she is employed whereby they refrain from filing a lawsuit, or withdraw an existing one. In each of these cases, society's interest in vindicating the value of potential life seems to be left unanswered.

These concerns seem to us more apparent than real and, in any case, not insoluble. To begin with, we do not believe that the proposed method's dependence on the initiation and continuation of a civil suit by the parents would seriously reduce its deterrent effect. First, even if the proposed civil fine were not imposed in many cases, its very enactment and its implementation—though in a small number of cases—would send an important symbolic and deterrent message to professionals. It would significantly enhance their awareness of the importance of the social value at stake, and increase their level of care. Second, we believe that in most cases the parents' monetary and non-monetary incentives to claim damages would ensure a level of enforcement sufficient to preserve the required awareness and incentive.²⁸⁴

283. For a similar view expressed in another context, see Frank LaSalle, *The Civil False Claims Act: The Need for a Heightened Burden of Proof as a Prerequisite for Forfeiture*, 28 AKRON L. REV. 497 (1995) (recommending an evidentiary standard of "clear and convincing evidence" in claims for forfeitures and civil penalties under the FCA).

284. The parents' monetary incentive would derive from their willingness to win their share in the civil penalty and to receive compensation for their legal expenses (which should be available

Third, one must remember that even where a settlement is reached between the parties, society's need of deterrence is not necessarily left unanswered. Given the monetary reward awaiting the parents if they succeed in proving their case with clear evidence, it is only reasonable to assume that any settlement would reflect, at least partially, the extra-compensatory burden that would have been imposed on the physician in the absence of a settlement. Presumably, this extra-compensatory burden would be internalized by the negligent adviser, or by the medical institution that employed him or her, thereby influencing the future behavior of professionals, as well as officers at the managerial levels of medical institutions. Hence, as in other legal contexts (including plea bargaining), the prospect of settlement would not eliminate or significantly diminish the deterrent effect of the applicable sanction.²⁸⁵

In any case, various solutions may be suggested for this pragmatic problem. A first possible solution is that in addition to the parents' private claim, the very statute authorizing the imposition of the civil fine would recognize a parallel cause of action to the government, or any other nominated authority. Such a public authority would be entitled at any stage to file a separate suit (civil or administrative) in order to vindicate the same public value for whose sake the ability to make the adviser incur an extra burden is granted to the parents.²⁸⁶ This prerogative may be used where a wrongful abortion is detected and the parents decide not to claim damages or where they settle their claim against the negligent adviser. In the latter case, the public authority should take the settlement into account in deciding whether the public interest still justifies the instigation of a separate procedure.²⁸⁷

However, this simple solution clearly creates new problems. First, such a

only upon success in proving the defendant's negligence with clear and convincing evidence). The non-monetary incentive exists where the parents feel indignant at the negligence that caused the loss of their potential offspring. In some cases, this incentive may be even stronger than the financial incentive to bring suit.

285. Another possible way to alleviate the pragmatic concern discussed in this Subsection may be to prohibit an out-of-court settlement without the court's approval. *See, e.g.,* Sanford I. Weisburst, *Judicial Review of Settlements and Consent Decrees: An Economic Analysis*, 28 J. LEGAL STUD. 55 (1999). However, this solution could be applied only where a settlement was reached after the filing of the parents' suit.

286. Parallel causes of action are very often recognized in Federal statutes that provide for civil penalties. *See, e.g.,* 31 U.S.C. § 3730(a)-(b) (2000) (FCA); 18 U.S.C. § 1964(b)-(c) (2000) (RICO); 15 U.S.C. § 15(a)-(c) (2000) (Clayton Act).

287. Some settlements may reflect the parents' interest in extracting an extra-compensatory payment from the physician or the medical institution. When such is the case, the public interest in deterring wrongdoers is vindicated indirectly, although usually to a lesser extent, since the details of most settlement agreements do not come to the attention of the general public.

process may require establishing, funding, and administering a new governmental agency able to handle such a punitive claim professionally. Even if such a task were assigned to an existing department of the district attorney's office, it would increase the government's expenditure on enforcement. Contrary to the proposed private process, it would entail the instigation of a wholly separate civil or administrative action against the defendant for the same offense. The advantage of our proposal in terms of efficient administration would be immediately reduced.²⁸⁸

Second, the parallel process would be instigated and conducted by the state's prosecutorial bodies for the exclusive goal of inflicting punishment on the adviser, and as such would strikingly resemble a criminal process in which the problem of power imbalance between the parties is immanent. The quasi-criminal nature of the process may entail the use of more stringent procedural safeguards and would thereby complicate and prolong the proceedings, making them even more expensive.

We consequently believe that recognizing a parallel public cause of action for wrongful abortions is undesirable. The primary target of the legal response we have been seeking throughout this Article is, after all, an act of ordinary negligence, committed by a competent physician in an attempt to provide medical care in good faith. Moreover, the adviser's liability toward the parents already provides a certain incentive for prudent behavior. That is why we have contended that solving the current problem of under-deterrence does not necessarily entail a dramatic rise in the adviser's expected sanction. A moderate response might suffice to increase the potential wrongdoer's awareness of the risks of negligent inducement of abortions and motivate professional improvement. The possibility of a civil fine proposed herein enhances any deterrent effect created by tort law, making the marginal benefit of a parallel public cause of action rather trivial. The amount of public resources invested in an attempt to prevent wrongdoing and punish wrongdoers should be proportionate to the societal benefit derived from such efforts. It seems, therefore, that the creation and maintenance of a wholly new public enforcement system to handle the proposed mechanism's slight deficiencies would be unjustified from a social welfare standpoint.

We wish to conclude this Part by suggesting a more plausible means to back

288. Moreover, the information required in order to open a public investigation in a case of wrongful abortion would in many cases depend on the parents' readiness to file a complaint to the relevant public authority. At least in some cases in which the parents were reluctant to bring a private suit, they would undoubtedly be reluctant to forward a public complaint either. The societal loss from the lack of a public procedure would not always be recovered by the creation of a parallel cause of action.

up the incentives created by the civil fine proposed, namely a limited reform in disciplinary law. An expansion of the authority of medical boards to enable them, in cases of wrongful abortion, to impose monetary penalties for acts of ordinary negligence may reinforce the deterrent effect of the punitive-civil action proposed above, without adding a significant burden in terms of enforcement cost. As the reader may recall, in Section III.D we explained why the typical disciplinary causes of action, as well as the typical sanctions of revocation, suspension, and the like, are unavailable in wrongful abortion cases. Allowing a medical board to consider the imposition of a monetary fine as an additional means of enforcement might meaningfully add to the preservation of the potential wrongdoers' awareness of the need to avoid negligence in this field.

True, the prospect of such a parallel administrative form of enforcement raises similar problems to those we mentioned as justifying the rejection of the proposal to create a parallel public cause of action. However, a disciplinary proceeding differs from an ordinary public process at least in three important respects. First, in a disciplinary action the professional is indicted and judged not by a representative of the public, to whom he is a stranger, but by the very professional group to which he belongs. Not only does this diminish to some extent the power imbalance between the parties, it also enhances the possibility of a more sympathetic attitude to the defendant, especially in cases of ordinary one-time negligence, which is the focus of our discussion. This may justify relaxing some of the procedural safeguards that would need to be introduced into an ordinary public procedure and thus would make our proposal more attractive in terms of administrative efficiency. Second, the medical boards are professional bodies, which in their very nature are competent to consider and decide whether an act leading to wrongful abortion had been negligent. Third, a disciplinary process is an administrative procedure, hence far less costly than a judicial-legal process.

To conclude, with or without the aid of the disciplinary process, we believe that under the enforcement mechanism proposed in this Article the need to prevent wrongful abortions and to vindicate the public interest in preserving the potentiality of human life would be satisfied more effectively than it is today.

CONCLUSION

In this Article we have endeavored to rectify a disturbing anomaly in American law. On the one hand, the potentiality of human life embodied in the living cells of a fetus is a well-recognized social value. On the other hand, as our inquiry into the intricacies of existing law has demonstrated, the legal protection of this value, at least in cases of wrongful abortion, is to date relatively feeble, despite its supreme importance. We argued that the poor reaction to wrongful abortions may suffice to deter negligent advisers once in a while, but may not

render the required level of deterrence in most cases.

In searching for a balanced legal response to medical malpractice leading to unwarranted abortions, we started from the classical defender of public interests, the criminal law. Recognizing the shortcomings of this harsh mechanism in the context of an act of ordinary negligence such as the one discussed in this Article, we moved on to examine an apparently more moderate solution, that of expanding the civil liability of the negligent adviser. Struggling with the theoretical complexities of this proposal, we argued that such an extension of liability, though at first sight very problematic, is not theoretically impossible to justify. Nevertheless, given the drawbacks of the traditional remedial mechanism of civil law in terms of the ability to guarantee sufficient but not exaggerated levels of deterrence and retribution, we had to reject this proposal as well.

Finally, recognizing the vices and virtues of the traditional deterrence mechanisms of both the criminal and the civil law, we concluded the article by pointing out what seems to us the way out of this legal labyrinth. We proposed the enactment of a statutory provision as follows. A civil court adjudicating a wrongful abortion case, upon a finding by clear and convincing evidence of negligence on the part of the defendant that led to the loss of potential life, would be allowed to impose on the defendant an extra-compensatory civil fine. The court would do so if and only if it was convinced that such an additional sanction could improve the legal protection of potential life in terms of efficient deterrence and just desert. The amount of the fine would be determined with reference to all the circumstances of the case, and its monies would be divided between the parents of the unborn child and the state. In our view, this original solution, providing for a flexible and adjustable civil penalty upon a finding of a private wrong committed by the defendant against the parents, would best serve society's need in affording protection to the value of potential life. It would do so without imposing undue burdens or unjustified punishments on negligent physicians and without imposing a significant economic burden on the legal system.

As we mentioned at the very outset of this Article, we do not presume to claim with any scientific certainty that the other solutions examined above, or any other solution that has not been discussed, are impossible to defend. As the late Justice Frankfurter illuminatingly clarified many years ago:

How to effectuate policy—the adaptation of means to legitimately sought ends—is one of the most intractable of legislative problems. Whether proscribed conduct is to be deterred by *qui tam* action or triple damages or injunction, or by criminal prosecution, or merely by defense to actions in contract, or by some, or all, of these remedies in combination, is a matter within the legislature's range of choice. Judgment on the deterrent effect of the various weapons in the armory of the law can lay little claim to scientific basis. Such

judgment as yet is largely a prophecy based on meager and uninterpreted experience.²⁸⁹

Our modest hope is that this Article has succeeded in making the point that regardless of the ongoing disagreement about the exact meaning and weight of potential life, any community interested in the preservation of potential life should pay careful attention to the way its legal system reacts to wrongful abortions.

289. *Tigner v. Texas*, 310 U.S. 141, 148 (1940).

The FDA and the Tort System: Postmarketing Surveillance, Compensation, and the Role of Litigation

Catherine T. Struve, J.D.*

INTRODUCTION

Both the tort system and the FDA seek to protect consumers of medical products. The tort system provides compensation when a consumer is harmed by a defective product and sets incentives for companies to design safer products. The FDA imposes an elaborate system of prior restraint: Pharmaceuticals and some medical devices must undergo extensive testing and stringent risk/benefit analysis before the FDA will approve them for marketing.¹

Formerly, the FDA viewed its risk/benefit analysis as setting a floor but not a ceiling for product safety: FDA-approved products could be marketed, but the manufacturer might still incur liability if a court later decided that a product was defective or a warning was inadequate.² This view has changed in recent years,

* Assistant Professor, University of Pennsylvania Law School. I am grateful to Kristin Madison, Theodore Ruger, William Sage, and Kim Scheppele for their suggestions on the initial stages of this project, to participants in a University of Pennsylvania Law School ad hoc workshop and participants in faculty workshops at Rutgers School of Law-Camden and at Brooklyn Law School, and to Allison Arwady, Stephen Burbank, Rochelle Chodock, Frank Goodman, Geoffrey Hazard, Richard Nagareda, Kermit Roosevelt, Chris Sanchirico, Anthony Sebok, Jonathan Siegel, Allan Stein, Aaron Twerski, and an anonymous reviewer for extremely helpful comments on prior drafts. I thank Dylan Steinberg for superb research assistance and for being a thoughtful sounding board for many of the ideas discussed here, and Ronald Day and the staff of the Biddle Law Library for assistance in obtaining sources. Remaining errors, of course, are mine.

1. See *infra* text accompanying notes 25-43. But see source cited *infra* note 43 (observing that the FDA does not require comparisons between the product under review and alternative treatments).

2. See, e.g., Margaret Jane Porter, *The Lohr Decision: FDA Perspective and Position*, 52 FOOD & DRUG L.J. 7, 11 (1997) ("FDA product approval and state tort liability usually operate independently, each providing a significant, yet distinct, layer of consumer protection."). As Porter—then FDA's Chief Counsel—explained, "FDA regulation of a device cannot anticipate and

however, as policymakers have stressed the need to bring innovative medical treatments to market. Some now argue that the FDA review process should set both a floor and a ceiling: FDA approval of a new product indicates not only that the product can be marketed, but that it should be; FDA rejection of a proposed product warning means not only that the warning is unnecessary, but that it could be counterproductive.

FDA officials who hold this view consider the tort system dangerous. The threat of tort liability, they warn, deters pharmaceutical companies and device makers from developing much-needed new technologies.³ Even if those innovations are merely delayed rather than abandoned altogether, the cost is felt not merely in financial terms but also in the suffering of people whose illnesses could have been treated with the new drug or device.

These critics argue that the tort system—and juries in particular—should not be permitted to determine product safety. Lay juries, it is claimed, are incapable of understanding the complex scientific and statistical evidence relevant to product safety; they are eager to help injured plaintiffs—especially when the defendant has deep pockets—and they overlook the many consumers who might benefit from the product; they award excessive compensatory damages, especially for pain and suffering; and they often compound the problem by awarding staggering sums in punitive damages.⁴ With these concerns in mind, the FDA's then-Chief Counsel took the controversial step, in 2002 through 2004, of submitting amicus briefs in support of the defendants in several cases

protect against all safety risks to individual consumers.” *Id.*

3. See Amicus Brief at 26, *Horn v. Thoratec Corp.*, 376 F.3d 163 (3d Cir. 2004) (No. 02-4597) (arguing that tort awards “can harm the public health by retarding research and development and by encouraging ‘defensive labeling’ by manufacturers to avoid state liability, resulting in scientifically unsubstantiated warnings and underutilization of beneficial treatments”).

4. As the FDA argued last year (with respect to medical devices) in a submission to the United States Court of Appeals for the Third Circuit:

State actions are not characterized by centralized expert evaluation of device regulatory issues. Instead, they encourage, and in fact require, lay judges and juries to second-guess the balancing of benefits and risks of a specific device to their intended patient population—the central role of FDA—sometimes on behalf of a single individual or group of individuals. That individualized redetermination of the benefits and risks of a product can result in relief—including the threat of significant damage awards or penalties—that creates pressure on manufacturers to add warnings that FDA has neither approved, nor found to be scientifically required, or withdrawal of FDA-approved products from the market in conflict with the agency’s expert determination that such products are safe and effective.

Id. at 25-26.

concerning FDA-approved products.⁵ Detailed FDA scrutiny of a product, the briefs contended, should preempt litigation challenging the product's safety (unless the defendant has violated FDA requirements).

There are strong reasons to question the view that FDA approval should preempt products liability claims.⁶ To establish that preemption is warranted, proponents should be required to provide convincing evidence of serious flaws in the current system.⁷ In this regard, it should be noted that the opponents of the

5. See Amicus Brief for Defendant, *Horn* (No. 02-4597); *Horn*, 376 F.3d at 178 (quoting Statement of Interest of the United States at 7-9, *Murphree v. Pacesetter, Inc.* (Tenn. Cir. Ct. Dec. 12, 2003) (No. 005429-00-3)); Amicus Curiae Brief of the United States, in Support of Defendants/Respondents, *Dowhal v. Smithkline Beecham Consumer Healthcare*, 88 P.3d 1 (Cal. 2003) (No. S109306); Amicus Brief for the United States in Support of the Defendant-Appellee and Cross-Appellant, *Motus v. Pfizer, Inc.*, 358 F.3d 659 (9th Cir. 2002) (Nos. 02-55372, 02-55498).

Concerns over the FDA's stance arose partly from reports that the FDA's Chief Counsel solicited input from industry lawyers concerning cases in which the FDA could usefully intervene on behalf of defendants. In December 2003, then-FDA Chief Counsel Daniel Troy participated in a "roundtable" entitled "The Case for Preemption" at a continuing legal education conference designed for in-house and outside counsel for pharmaceutical and medical device companies. Program, 8th Annual Conference for In-House Counsel and Trial Attorneys: Drug and Medical Device Litigation (Dec. 14-16, 2003), <http://www.gibbonslaw.com/publications/uploadedfiles/602L04-NYC.pdf>.

Jessica Dart, who represents a number of products liability plaintiffs, attended the conference after learning that the roundtable agenda included two of her cases. Affidavit of Jessica R. Dart at ¶¶ 1, 3, *Dusek v. Pfizer, Inc.*, No. Civ. A. H-02-3559, 2004 WL 2191804 (S.D. Tex. Feb. 20, 2004). According to Dart, Troy "made it clear that he was interested in filing even more amicus briefs on behalf of pharmaceutical manufacturers and actually invited his defense counsel audience to approach him with requests . . ." *Id.* ¶ 5.

Daniel Troy resigned from his post as Chief Counsel in November 2004. See FDA, Statement of Dr. Lester M. Crawford, Acting Commissioner of Food and Drugs on the Resignation of Daniel E. Troy (Nov. 16, 2004), <http://www.fda.gov/bbs/topics/news/2004/NEW01135.html>. News reports, however, have suggested that the FDA's Acting Commissioner continues to support Troy's policy views. See, e.g., *FDA Chief Counsel Dan Troy Resigning, Masoudi Rumored as Replacement*, FDA WEEK, Nov. 19, 2004 (quoting an internal email from Lester Crawford that praised Troy for "put[ting] his personal reputation on the line defending the Agency's prerogatives from intrusion by courts applying state law in product liability actions").

6. For a summary of current law concerning preemption, see *infra* notes 105-110 and accompanying text.

7. The tort system is designed to promote product safety and compensate those injured by defective products—roles that were entirely compatible with the FDA's previous view that tort liability could supplement the FDA's efforts to ensure product safety. See *supra* note 2 and accompanying text. Moreover, products liability law lies within the area of consumer health and safety—an area traditionally within the regulatory powers of the states. Therefore, those who assert

tort system overstate the case:⁸ Empirical data indicate that juries do better than their critics assert at handling technical issues,⁹ that juries are not as eager as some think to award damages against business defendants,¹⁰ and that punitive damages are awarded rarely in products liability suits (and mainly in cases involving egregious misbehavior).¹¹

In addition to demonstrating a need for change, advocates of preemption should also be required to demonstrate that preemption is the best alternative to the status quo. It is true that the FDA possesses greater expertise concerning

that Congress should preempt state tort liability should bear the burden of showing that such preemption is necessary.

8. Theodore Eisenberg and James Henderson have argued that, in fact, data indicate a pro-defendant trend in recent decades. See Theodore Eisenberg & James A. Henderson, Jr., *Inside the Quiet Revolution in Products Liability*, 39 UCLA L. REV. 731, 741 (1992) (noting "a continuing decline in plaintiff success rates" over the period from 1979 to 1989); Theodore Eisenberg, *Judicial Decisionmaking in Federal Products Liability Cases, 1978-1997*, 49 DEPAUL L. REV. 323, 323-24 (1999) (noting low plaintiff win rates at trial and also observing that "[o]f those cases that survive early pretrial skirmishing, and end in pretrial judgment, an increasing percentage is resulting in pretrial judgment in favor of defendants" based on data extending "through fiscal 1997").

9. See, e.g., NEIL VIDMAR, *MEDICAL MALPRACTICE AND THE AMERICAN JURY: CONFRONTING THE MYTHS ABOUT JURY INCOMPETENCE, DEEP POCKETS, AND OUTRAGEOUS DAMAGE AWARDS* 265 (1995); Joe S. Cecil et al., *Citizen Comprehension of Difficult Issues: Lessons from Civil Jury Trials*, 40 AM. U. L. REV. 727, 729 (1991).

10. See VALERIE P. HANS, *BUSINESS ON TRIAL: THE CIVIL JURY AND CORPORATE RESPONSIBILITY* 23, 175-77 (2000). Other researchers have noted that plaintiffs' win rates are relatively low in products liability jury trials (compared to other types of cases). See Theodore Eisenberg et al., *Litigation Outcomes in State and Federal Courts: A Statistical Portrait*, 19 SEATTLE U. L. REV. 433, 437 (1996) (reporting with respect to products liability claims (other than asbestos claims) tried in 1991-1992 that "success rates are 40 percent in state court and 37 percent in federal court"). Of course, this finding does not prove that juries are particularly unsympathetic to products liability plaintiffs; the mix of cases selected for trial can differ across types of cases and can affect win rates. However, the finding does suggest that juries are not automatically receptive to plaintiffs' claims in products liability cases.

11. See, e.g., Michael Rustad, *In Defense of Punitive Damages in Products Liability: Testing Tort Anecdotes with Empirical Data*, 78 IOWA L. REV. 1, 23 (1992) (describing a study of products liability verdicts that indicated that "punitive damages were rarely awarded," which showed that "[t]he gap between what was awarded and collected was great," and that cases resulting in punitive awards involved "corporate misconduct and serious injuries"). Researchers recently summarized the empirical findings on punitive awards: "Contrary to popular belief, juries rarely award such damages, and award them especially rarely in products liability and medical malpractice cases. Rather, juries tend to award punitive damages in intentional misconduct cases. When juries do award punitive damages, they do so in ways that relate strongly to compensatory awards." Theodore Eisenberg et al., *Juries, Judges, and Punitive Damages: An Empirical Study*, 87 CORNELL L. REV. 743, 745 (2002) (footnotes omitted).

product safety than a civil jury. The FDA is correct to suggest that FDA regulation and the tort system should not operate entirely independently; the FDA's expertise gives its views on product safety considerable authority and those views should play a role in assessing product liability. But I will argue that even if systemic change is shown to be necessary, allowing FDA regulation to supplant the tort system is not the only, or the best, solution.

Permitting FDA approval to preclude the possibility of tort liability does more than ensure that product safety decisions are reserved to the FDA. Preemption of tort litigation removes the opportunity for litigation to aid the FDA in its goal of monitoring product safety.¹² Preemption also denies compensation to persons harmed by an FDA-approved product—even if they were harmed after a safety problem first surfaced but before the FDA took regulatory action to remove the product from the market or to require additional warnings.¹³

There exists a large body of literature concerning the appropriate scope of FDA regulatory preemption.¹⁴ I will argue, however, that this literature fails to contemplate the full range of possible options.¹⁵ Even if proponents of reform can ultimately carry their burden of showing the need for change, policymakers

12. See Robert L. Rabin, *Reassessing Regulatory Compliance*, 88 GEO. L.J. 2049, 2069 (2000) (“[I]f we are substantially dependent on the tort system to provide the educational function of revealing massive cover-ups of health information by industries like asbestos, or occasional efforts to conceal risk information from regulatory agencies like the FDA, then it is undeniably the case that tort law is serving a positive function of some consequence.”).

13. Plaintiffs could, under some circumstances, assert a claim against the United States based upon the FDA's failure to act concerning a product, but such claims would often fail due to the “discretionary function” exception in the Federal Tort Claims Act. See *Berkovitz v. United States*, 486 U.S. 531, 545 (1988) (stating that “application of the discretionary function exception” to a claim concerning agency determinations that a vaccine complied with federal standards “hinges on whether the agency officials making that determination permissibly exercise policy choice”).

14. See, e.g., Robert S. Adler & Richard A. Mann, *Preemption and Medical Devices: The Courts Run Amok*, 59 MO. L. REV. 895 (1994); Barbara L. Atwell, *Products Liability and Preemption: A Judicial Framework*, 39 BUFF. L. REV. 181 (1991); Richard C. Ausness, *Federal Preemption of State Products Liability Doctrines*, 44 S.C. L. REV. 187 (1993); Brian J. Donato & Mary Beth Neraas, *Federal Preemption of Product Liability Claims Involving Drugs and Medical Devices Regulated Under the Federal Food, Drug, and Cosmetic Act*, 48 FOOD & DRUG L.J. 305 (1993); David R. Geiger & Mark D. Rosen, *Rationalizing Product Liability for Prescription Drugs: Implied Preemption, Federal Common Law, and Other Paths to Uniform Pharmaceutical Safety Standards*, 45 DEPAUL L. REV. 395 (1996); Betsy J. Grey, *Make Congress Speak Clearly: Federal Preemption of State Tort Remedies*, 77 B.U. L. REV. 559 (1997); Lars Noah, *Amplification of Federal Preemption in Medical Device Cases*, 49 FOOD & DRUG L.J. 183 (1994).

15. Cf. Richard A. Nagareda, *In the Aftermath of the Mass Tort Class Action*, 85 GEO. L.J. 295, 298 (1996) (arguing that “the debate over procedural reform within the tort system ignores the relationship between that system and the modern administrative state”).

should keep in mind that preemption is not the only alternative to the status quo. In addition to considering whether there are ways to improve the performance of the current litigation system, policymakers should ask whether litigation could be restructured in a way that could improve the FDA's regulatory performance.

This Article considers whether Congress could create structural links between the litigation system and the FDA—either by providing for agency adjudication of products liability claims or by requiring federal courts to refer issues of product safety and causation to the agency for determination. After comparing four possible configurations, I conclude that policy considerations would weigh in favor of adjudication in federal court, with referral of technical questions to the FDA. In discussing this possibility, I draw upon insights provided by Richard Nagareda, who has argued that such a referral could be accomplished through the use of the primary jurisdiction doctrine.¹⁶ I conclude, however, that such a mechanism could well violate the Seventh Amendment if applied to private products liability claims.¹⁷ Accordingly, I describe an alternative scheme in which product safety claims for damages by the United States as *parens patriae*¹⁸ could be litigated in federal court by *qui tam* relators.¹⁹

The system would employ a somewhat novel process to adjudicate claims. After a period of discovery, a suit that survived summary judgment would proceed to a bench trial. Instead of ruling upon the issues of product safety and causation, however, the judge would refer those issues to an FDA advisory panel. The panel's determinations would be reviewed by the FDA, and the FDA's final determinations would be conclusive regarding product safety and causation.²⁰ If warranted, the court would then determine an aggregate amount of damages and would enter judgment. A fraction of the damages would be paid to the *qui tam* relator, and the bulk of the damages would finance a federal compensation fund

16. *See id.* at 352 (suggesting that Congress enact a framework within which courts would “apply the doctrine of primary jurisdiction to defer their disposition of individual claims pending agency action”); *see also* Richard A. Nagareda, *Turning from Tort to Administration*, 94 MICH. L. REV. 899, 978 (1996) (suggesting a regulatory scheme for facilitating claim resolution, within which “[a] mass tort centered upon a medical device like breast implants appropriately might come within the expertise of the FDA—the agency that originally licensed that product”).

17. *See infra* notes 238-265 and accompanying text.

18. This phrase, which translates “parent of the country,” denotes the state's ability to sue to protect its interest in the health and safety of its citizens. *See infra* notes 153-157 and accompanying text.

19. *Qui tam* relators are litigants who sue on behalf of the government (and who may earn a bounty for doing so successfully). *See infra* notes 158-170 and accompanying text.

20. The FDA's findings would be reviewed in the federal court proceeding for compliance with procedural requirements and to ensure that the findings were supported by some evidence. *See infra* notes 143-151 and accompanying text.

for those harmed by the product.²¹

Such a mechanism might improve the FDA's postmarketing surveillance of regulated products. The filing of such a suit could flag possible safety problems for the FDA. Discovery obtained in such a suit might uncover evidence that had not been reported to the FDA, or upon which the FDA had not yet focused. And the FDA's review of panel determinations would provide the agency with an opportunity to reassess its product safety determinations in light of the record developed in the litigation.

The scheme would also change the landscape of compensation. The amount of damages awarded could vary depending on the presence and degree of fault on the part of the company: A carelessly overlooked safety problem could trigger compensatory damages, while instances of willful deception might generate additional penalties. Even in the absence of fault, the scheme might require the company to provide some minimum level of damages to compensate for harms traceable to the product. Damages awarded in such a scheme would likely be smaller than some jury awards in similar cases tried in the tort system. On the other hand, damages could be recovered in cases where suits would currently be preempted, as well as in cases where preemption would not exist but where state law would not impose liability.

Companies could choose whether or not to opt in to the new system; if they declined to opt in, there would be no preemption of state-law tort claims. The opt-in feature would have an interesting policy implication.²² Though preemption advocates argue that the size and variability of jury awards deter companies from pursuing desirable projects, that assertion is hotly contested and difficult to evaluate. Companies themselves are better suited than legislators to determine the incentive effects of litigation, but their statements are likely to be self-interested. By permitting companies to choose between the tort system and the new alternative, the scheme would elicit a more accurate picture of companies' preferences.

My argument proceeds in four steps. I begin by summarizing, in Part I, the FDA's role in scrutinizing product safety. Part I explains why premarketing review predictably will fail to identify all potential safety issues and discusses flaws in the FDA's current postmarketing surveillance system. In Part II, I consider several possible structural changes that could link litigation more closely to the regulatory process. After weighing the relative merits of agency and court adjudication, and comparing the types of entities that might press a claim concerning product safety, I suggest that *qui tam* suits on behalf of the

21. See *infra* notes 168-169, 176-178 and accompanying text.

22. Companies' ability to choose whether or not to opt in to the system also addresses some possible constitutional concerns about the scheme. See *infra* note 268 and accompanying text.

United States as *parens patriae*, litigated in federal court, might be the best option for creating structural links. Part III considers the advantages of the system's opt-in feature and argues that the system could supplement and improve the FDA's postmarketing surveillance efforts. In conclusion, I consider potential disadvantages of the system—including the possibility that a pro-industry bias or a lack of resources might compromise the agency's role in the proposed scheme—and I assess the place of the hybrid scheme in the debate over preemption. While this Article does not establish that the hybrid scheme is superior to the status quo, it does support the contention that the hybrid scheme is superior to preemption. Advocates of preemption, then, not only must demonstrate that the current system is undesirable, but also should be required to show that preemption is preferable to a hybrid system.

I. THE ROLE OF POSTMARKETING SURVEILLANCE IN THE REGULATORY SYSTEM

An FDA task force recently summarized the agency's role in promoting consumer safety:

The Agency establishes and enforces product quality standards intended to prevent defective products from reaching the market. For products of acceptable quality, the central element of FDA's risk management is controlling product entry to the marketplace. The majority of FDA program resources are devoted to premarketing scientific risk identification and assessment and approval or nonapproval. Significant, but substantially fewer, resources are devoted to postmarketing surveillance and risk assessment activities.²³

In this Section I explain why postmarketing surveillance is critical to consumer safety, and I argue that despite the FDA's efforts to improve postmarketing surveillance, that aspect of its program still falls short.

A. Premarket Scrutiny

During premarket review, the FDA weighs a medical product's known risks and determines whether the product should be approved for marketing and, if so, whether warnings should be included in the labeling.²⁴ New drugs and medical devices undergo varying degrees of FDA scrutiny depending on their originality and other factors.

23. TASK FORCE ON RISK MGMT., FDA, MANAGING THE RISKS FROM MEDICAL PRODUCT USE—CREATING A RISK MANAGEMENT FRAMEWORK 29-30 (1999), <http://www.fda.gov/oc/tfrm/riskmanagement.pdf> [hereinafter RISK MANAGEMENT REPORT].

24. *Id.* at 30.

The standard drug approval proceeds in four steps:

Phase I seeks pharmacologic effects information and early evidence on effectiveness in several dozen healthy persons. Phase II measures several hundred closely monitored sick patients for the clinical effectiveness of the drug. Both prepare the product for its real test, the multiple Phase III effectiveness and safety tests which form the basis for risk assessments and label warnings, during which more than a thousand patients are likely to be exposed.

The final phase of the review process leads to formal acceptance of the proposed [new drug application].²⁵

The process is lengthy; approval can be somewhat speedier, however, for certain urgently needed drugs and for generic versions of drugs already on the market.²⁶ A “fast track” approval process is available for new drugs that “treat[] a serious or life-threatening condition and . . . demonstrate[] the potential to address unmet medical needs for such a condition.”²⁷ Manufacturers of generic drugs can take advantage of the abbreviated new drug application process,²⁸ often bypassing “the extensive clinical testing that a pioneer product would endure.”²⁹

Medical devices are grouped in three categories, in ascending order of riskiness. In Class I are seemingly innocuous devices such as dental floss;³⁰ these receive the least demanding regulatory oversight.³¹ Devices whose safety and

25. 1 JAMES T. O'REILLY, *FOOD AND DRUG ADMINISTRATION* § 13.11, at 13-66 to -67 (2d ed. 1995).

26. See James O'Reilly & Amy Dalal, *Off-Label or Out of Bounds? Prescriber and Marketer Liability for Unapproved Uses of FDA-Approved Drugs*, 12 *ANNALS HEALTH L.* 295, 304 (2003) (“The FDA approval process is complex and detailed, and its timing cannot keep up with the fast pace of medical discovery about pharmaceutical benefits. Even with the advent of the accelerated ‘fast-track’ approval process, the process for drug approval is one that is still lengthy and time-consuming.” (footnote omitted)).

27. 21 U.S.C. § 356(a)(1) (2000); see also 21 C.F.R. § 314.500 (2004). For drugs that receive fast track evaluation, the FDA may impose safety restrictions on the distribution or use of the drug, see *id.* § 314.520, and may require postapproval studies, see *id.* § 314.510, and “[p]ostapproval reporting of adverse events is much more closely monitored,” 1 O'REILLY, *supra* note 25, § 13.13, at 13-83.

28. See 21 C.F.R. § 314.92 (2004).

29. 1 O'REILLY, *supra* note 25, § 13.15, at 13.93.

30. See 21 C.F.R. § 872.6390 (2004).

31. See *id.* § 860.3(c)(1) (“Class I means the class of devices that are subject to only the general controls authorized by or under sections 501 (adulteration), 502 (misbranding), 510 (registration), 516 (banned devices), 518 (notification and other remedies), 519 (records and

effectiveness require greater scrutiny—such as contact lenses³²—are grouped in Class II and subjected to additional controls.³³ When more study is needed to determine the safety and effectiveness of a medically important or potentially dangerous device, it is assigned to Class III and the manufacturer must obtain premarket approval from the FDA.³⁴

A company seeking premarket approval of an innovative Class III device must test the product and must provide detailed data, including “full reports of all information, published or known to or which should reasonably be known to the applicant, concerning investigations which have been made to show whether or not such device is safe and effective.”³⁵ However, an applicant can use “premarket notification” to bypass the full premarket application process if it can show that its device is “substantially equivalent” to a device already on the market.³⁶ Though a premarket notification applicant will need to submit an analysis of existing data on the device, the FDA will not usually demand testing.³⁷ If the FDA accepts the notification, it will approve the product for marketing; if not, the applicant will have to proceed to the premarket application.

The FDA’s mission of protecting consumer safety dictates rigorous premarketing review, but its mandate to foster innovation creates a countervailing pressure. In 1997, finding that “prompt approval of safe and effective new drugs and other therapies is critical to the improvement of the public health,”³⁸ Congress directed the FDA to employ the “least burdensome”

reports), and 520 (general provisions) of the act.”).

32. See *id.* § 886.5916(b)(1) (stating that daily wear rigid gas permeable contact lenses are Class II devices); *id.* § 886.5925(b)(1) (stating that daily wear soft contact lenses are Class II devices).

33. See *id.* § 860.3(c)(2) (“A device is in class II if general controls alone are insufficient to provide reasonable assurance of its safety and effectiveness and there is sufficient information to establish special controls, including . . . performance standards, postmarket surveillance, patient registries, . . . guidance documents . . . , recommendations, and other appropriate actions . . .”).

34. The regulations explain:

A device is in class III if insufficient information exists to determine that general [or special] controls are sufficient to provide reasonable assurance of its safety and effectiveness . . . and if, in addition, the device is life-supporting or life-sustaining, or for a use which is of substantial importance in preventing impairment of human health, or if the device presents a potential unreasonable risk of illness or injury.

Id. § 860.3(c)(3).

35. 21 U.S.C. § 360e(c)(1)(A) (2000).

36. See *id.* § 360e(b)(1)(B).

37. See 1 O’REILLY, *supra* note 25, § 18.07, at 18-32.

38. Food and Drug Administration Modernization Act of 1997, Pub. L. No. 105-115, § 101(1), 111 Stat. 2296, 2298 (reprinted at 21 U.S.C. § 379g note (2000)).

methods of evaluating products in the premarket notification and premarket approval processes.³⁹ Pursuant to this mandate, the FDA has declared that premarket approval can sometimes be based on “well-designed bench and/or animal testing” rather than clinical tests.⁴⁰ Moreover, the FDA will consider the extent to which measures such as postmarketing trials can substitute for premarket scrutiny.⁴¹ Though clinical data are in any event not required for most premarket notifications, the FDA responded to the “least burdensome” directive by emphasizing that “substantial equivalence” determinations should also be streamlined.⁴²

Though the details of the approval process are complex, the bottom line is plain: When a company seeks FDA approval of an innovative drug or Class III device, the FDA will require the manufacturer to submit test data concerning safety and effectiveness. The FDA will weigh the product’s potential risks and benefits in determining whether to grant approval,⁴³ and consideration of the product’s risks can lead the agency to impose detailed requirements concerning product warnings. Premarketing scrutiny can provide a significant increase in product safety, but, as the next Section discusses, there is no way that it can discern all potential risks.⁴⁴

39. 21 U.S.C. § 360c(a)(3)(D)(ii) (2000) (premarket approval); *id.* § 360c(i)(1)(D) (premarket notification).

40. FDA, THE LEAST BURDENSOME PROVISIONS OF THE FDA MODERNIZATION ACT OF 1997: CONCEPT AND PRINCIPLES; FINAL GUIDANCE FOR FDA AND INDUSTRY 10 (2002), <http://www.fda.gov/cdrh/ode/guidance/1332.pdf>.

41. *See id.* at 4.

42. *See id.* at 9.

43. However, the FDA does not require comparative testing of the product’s performance relative to competitor brands (though marketing or practical considerations may dictate that the company perform such tests). *See* Editorial, *Comparing Prescription Drugs*, N.Y. TIMES, Aug. 27, 2003, at A20 (“[T]he drugs used in this country are seldom tested against one another in head-to-head combat. Instead, each is tested separately against a placebo and then, if shown to be safe and effective, is approved for marketing.”).

44. Even as to risks that could be discerned at the premarket review stage, some have argued that the FDA’s reliance on the regulated company to supply the necessary safety data can lead to problems. *See, e.g.,* Thomas O. McGarity, *Beyond Buckman: Wrongful Manipulation of the Regulatory Process in the Law of Torts*, 41 WASHBURN L.J. 549, 559 (2002) (“When the onus is on the regulatee to provide data establishing that its product is ‘safe and effective’ . . . , the temptation is strong for a company to discount data indicating that the product may not meet the statutory test.”).

B. The Need for Postmarketing Surveillance

Even if it is rigorously conducted,⁴⁵ a process that focuses on prior approval inevitably will fail to capture all relevant information.⁴⁶ Clinical trials normally will fail to reveal a number of types of problems:⁴⁷ those that occur relatively

45. There is some reason to question whether the current premarketing approval process is always sufficiently rigorous. Results of a 2002 survey of FDA scientists revealed that of the 360 recipients who responded to the question, "Have you ever been pressured to approve or recommend approval for an NDA despite reservations about the safety, efficacy, or quality of the drug?," eighteen percent responded "Yes." OFFICE OF INSPECTOR GEN., U.S. DEP'T HEALTH & HUMAN SERVS., HHS SURVEY, at question 25 (2002), http://www.peer.org/docs/fda/12_14_04_FDA_survey.pdf [hereinafter HHS SURVEY]. The results to some of the survey questions, which were made public pursuant to a request under the Freedom of Information Act (FOIA), are available at http://www.peer.org/docs/fda/12_14_04_FDA_survey.pdf; see also News Release, Pub. Employees for Env't'l Responsibility, FDA Scientists Issued Early Warnings on Drug Approvals (Dec. 16, 2004), http://www.peer.org/news/news_id.php?row_id=449 (stating that the survey results were obtained under FOIA). Health and Human Services researchers estimated that 846 Center for Drug Evaluation and Research reviewers were eligible to participate in the survey; 401 responses were received. See OFFICE OF INSPECTOR GEN., DEP'T HEALTH & HUMAN SERVS., FDA'S REVIEW PROCESS FOR NEW DRUG APPLICATIONS: A MANAGEMENT REVIEW 37 app. F (2003), <http://oig.hhs.gov/oei/reports/oei-01-01-00590.pdf>. The researchers noted "three main limitations" of their survey: The survey used web-based technology and technical difficulties may have caused some "non-responses"; though the survey was anonymous, "some respondents may not have participated out of concerns for their anonymity"; and "although survey access was limited to CDER employees, the potential exists that some individuals not in our intended population completed the survey." *Id.*

46. See Michael D. Green, *Statutory Compliance and Tort Liability: Examining the Strongest Case*, 30 U. MICH. J.L. REFORM 461, 496 (1997) (discussing "the inability of the [investigational new drug] testing process to identify all of the risks associated with use of a drug"); William M. Sage, Note, *Drug Product Liability and Health Care Delivery Systems*, 40 STAN. L. REV. 989, 990 (1988) (noting with respect to drugs that "extensive use in humans is the only way to measure safety or efficacy").

47. This is true even if the trial is designed and executed evenhandedly and in good faith. Some observers, however, assert that "bias is now rampant in drug trials." MARCIA ANGELL, *THE TRUTH ABOUT THE DRUG COMPANIES: HOW THEY DECEIVE US AND WHAT TO DO ABOUT IT* 106 (2004). Angell notes a number of strategies that can skew study results; one is "to enroll only young subjects in trials Because young people experience fewer side effects, drugs will look safer in these trials than they would in practice." *Id.* at 107-08. Results of a 2002 survey of FDA scientists reveal doubts about the sufficiency of the information provided in new drug applications. Three hundred and sixty-one survey recipients responded to the question "[H]ow often do NDAs, including amendments submitted during the PDUFA time clock, contain enough data to adequately assess the SAFETY of a drug?" While fifty-six percent responded "Most of the time," thirty-two percent responded "Some of the time" (the other options were "All of the time" (two percent),

rarely,⁴⁸ those involving relatively subtle increases in the risk of already common problems, those that disproportionately affect a population subset not represented in the trial,⁴⁹ and those with a long latency period.⁵⁰ Thus, “premarketing studies cannot guarantee product safety.”⁵¹

Though some problems may surface almost immediately after marketing commences,⁵² others may take years to appear.⁵³ Once problems do manifest themselves, however, it is critical for the FDA to recognize and respond to those problems promptly, so as to minimize the danger to consumers.⁵⁴ An official from the General Accounting Office has suggested that there is growing cause for concern:

“Rarely” (nine percent), and “Never” (two percent)). Eighty-seven percent (out of 354 respondents) stated that “additional SAFETY data [would] improve CDER’s ability to adequately assess the safety of a drug” at least “[t]o some extent.” HHS SURVEY, *supra* note 45, at questions 12, 13.

48. See Funmilayo O. Ajayi et al., *Adverse Drug Reactions: A Review of Relevant Factors*, 40 J. CLINICAL PHARMACOLOGY 1093, 1094 (2000) (“Most clinically relevant [adverse drug reactions] occur at a rate of 1 in 10,000 or less.”).

49. See Ajayi et al., *supra* note 48, at 1094 (“[A] major shortcoming of clinical trials can be the failure to account for variability among patients in terms of age, gender, genetic background, coadministered drugs, the coexistence of other diseases, and their concurrent effects on drug metabolism and/or excretion.”); Am. Med. Ass’n, *Reporting Adverse Drug and Medical Device Events: Report of the AMA’s Council on Ethical and Judicial Affairs*, 49 FOOD & DRUG L.J. 359, 359-60 (1994) [hereinafter *AMA Report*] (noting that “the patient population used in clinical trials does not usually include vulnerable populations such as the elderly, the young, women, those with complicated disease, or those taking other medications”).

50. See, e.g., Timothy Brewer & Graham A. Colditz, *Postmarketing Surveillance and Adverse Drug Reactions: Current Perspectives and Future Needs*, 281 JAMA 824, 824 (1999) (noting that “[p]remarketing trials . . . lack the follow-up necessary to detect [adverse drug reactions] widely separated in time from the original use of the drug or delayed consequences associated with long-term drug administration”).

51. *AMA Report*, *supra* note 49, at 359-60.

52. See *Examining the Incidence of Medical Errors, Focusing on Understanding Adverse Drug Events: Hearing Before the Senate Comm. on Health, Educ., Labor, & Pensions*, 106th Cong. 6 (2000) [hereinafter *Senate Medical Errors Hearing*] (statement of Dr. Janet Woodcock, Director, Ctr. for Drug Evaluation and Research, FDA) (“New types of risks, and rare risks, may well be uncovered in the first year a drug is on the market.”).

53. See Ajayi et al., *supra* note 48, at 1099 (suggesting that “approximately 2 to 3 years of postmarketing experience is required to fully understand the safety profile of a new drug”).

54. Cf. Alastair J. J. Wood, *The Safety of New Medicines: The Importance of Asking the Right Questions*, 281 JAMA 1753, 1753 (1999) (discussing five drugs withdrawn from the market, and stating that “a staggering 19.8 million patients . . . were estimated to have been exposed to these 5 drugs before their removal”); Ajayi et al., *supra* note 48, at 1094 (“The [adverse drug reactions] undetected prior to approval of a drug product may pose serious health threats once released into the general population . . .”).

[T]he pressures on the U.S. system of pharmaceutical risk management are increasing. Prescription drug use in the U.S. continues to increase; . . . roughly 10 prescriptions were filled for every American in 1998. Further, direct-to-consumer advertising and other marketing techniques can greatly accelerate the rate at which a new drug is prescribed to large numbers of patients.⁵⁵

Other recent changes also serve to raise the stakes: It is now more likely that a drug will enter the U.S. market before it has developed a track record abroad and that it will do so on the basis of a less searching "fast track" review by the FDA.⁵⁶ Postmarketing surveillance, then, must play an increasingly vital role in ensuring consumer safety;⁵⁷ but as the next Section discusses, the available resources fall short.⁵⁸

C. Deficiencies in the FDA's Postmarketing Surveillance

The goals of the FDA's postmarketing surveillance are "to detect adverse events not previously observed, improve understanding of the potential severity of previously unanticipated risks, detect events resulting from drug interactions or drug effects in particular populations, and assess the potential for causal relationships."⁵⁹ The FDA employs several different methods, including reporting systems, medical databases, and studies and registries focused on

55. *Senate Medical Errors Hearing*, *supra* note 52 (letter from Janet Heinrich, Associate Director, Health Fin. & Pub. Health Issues, U.S. Gen. Accounting Office, to the Honorable James M. Jeffords, Chairman, Sen. Comm. on Health, Educ., Labor, & Pensions).

56. As Marcia Angell has explained:

[U]ntil a decade ago, drugs were usually first approved in Europe But now, most drugs are approved first in the United States. Furthermore, an increasing number of them are given accelerated review by the FDA, which means they come to market on the basis of less evidence. Thus, a drug may come into widespread use with very little research to back it up, and no experience in another country.

ANGELL, *supra* note 47, at 162.

57. See Ajayi et al., *supra* note 48, at 1097 ("Although fraught with certain limitations such as underreporting, the use of postmarketing surveillance is still very critical in collecting data on drug safety because the true adverse reaction profile of a drug is often not revealed until it has been widely used."); Sage, *supra* note 46, at 1015 ("At the time a drug is approved, many adverse effects are undiscoverable. Though the first such ADRs to arise are unpreventable, effective postmarketing surveillance can greatly reduce the total damage.").

58. See *Senate Medical Errors Hearing*, *supra* note 52 (statement of Sen. Edward M. Kennedy) ("Approximately 48 percent of prescription drugs on the market today have become available only since 1990. FDA needs additional resources to identify adverse reactions . . ."); Green, *supra* note 46, at 495-96 (noting with respect to "the post-approval period" concerning new drugs that "the FDA has inadequate resources to enforce regulatory compliance").

59. RISK MANAGEMENT REPORT, *supra* note 23, at 52.

specific issues.⁶⁰ Although the FDA publicly takes a generally positive view of its own efforts,⁶¹ there are reasons to question the agency's effectiveness. Indeed, a 2002 internal survey of reviewers in the FDA's Center for Drug Evaluation and Research (CDER) found that some two-thirds of respondents were either "[n]ot at all confident" or only "[s]omewhat confident" that the CDER "adequately monitors the safety of prescription drugs once they are on the market."⁶² These concerns are well-founded: The FDA receives large amounts of data both from regulated companies and from healthcare providers, but those data will sometimes be incomplete or lack sufficient detail. Further, the FDA does not have the capability—or, some charge, the motivation—to analyze thoroughly and act swiftly upon all the information that it does receive.

Federal law imposes significant reporting duties on manufacturers of medical devices,⁶³ as well as on certain healthcare facilities where those devices are used ("user facilities").⁶⁴ Manufacturers must report deaths, serious injuries, and device malfunctions, as well as baseline data, to the FDA within set time periods. User facilities must also report deaths to the FDA, and must report serious injuries to the manufacturer.⁶⁵ Pharmaceutical companies have similar reporting duties with respect to adverse drug events.⁶⁶

60. *Id.* at 54. An FDA task force recently listed these methods:

spontaneous reporting systems to rapidly identify potential new problems; large healthcare databases with product use linked to subsequent diagnoses, hospitalizations, and other adverse events; cohort and case-control studies conducted as needed to investigate a specific safety issue in depth; and registries initiated when potential risks (particularly those apparent only with long-term follow-up) are sufficient to warrant identification and active follow-up of individuals exposed to a product.

61. *See id.* at 51 ("The Task Force believes that FDA's postmarketing surveillance and risk assessment programs are, for the most part, accomplishing the purposes for which they were designed.").

62. HHS SURVEY, *supra* note 45, at question 45. Twenty-eight percent of respondents were "[m]ostly confident" and six percent of respondents were "[c]ompletely confident." *Id.*

63. *See* 21 U.S.C. § 360i(a) (2000) (imposing reporting requirements on manufacturers and importers of medical devices); 21 C.F.R. § 803.10(c) (2004) (summarizing device manufacturers' reporting requirements).

64. *See* 21 U.S.C. § 360i(b) (2000) (imposing reporting requirements on "device user facilities").

65. *See* 21 C.F.R. § 803.10 (2004) (summarizing user facilities' reporting duties).

66. *Id.* § 314.80 (specifying reporting requirements regarding adverse drug experiences). As Barbara Noah has explained:

Within fifteen days, manufacturers must submit reports of all adverse drug experiences that are both "serious" and "unexpected" and they must "promptly investigate" all such adverse experiences. By contrast, manufacturers need only submit periodic reports for non-serious or expected adverse events. The periodic reports must contain summaries of

Potential tort liability (where it exists) and the possibility of FDA penalties give companies incentives to monitor and report adverse events.⁶⁷ But there are countervailing pressures as well: As three government researchers recently noted, “[t]here are strong disincentives for companies . . . to identify safety problems with licensed drugs quickly and efficiently [S]eeking out and sharing bad news about a product are unlikely to increase business.”⁶⁸ Commenting on “episodes of falsification and concealment of research by manufacturers,” William Sage has observed that “[s]ince a manufacturer may have invested several million dollars in a drug before a single adverse reaction is reported, this misbehavior is predictable albeit unforgiveable.”⁶⁹ In a reflection of these pressures, there are indications that Merck was aware of potential problems with Vioxx long before it withdrew the drug from the market in September 2004, and that the company may have attempted to retard the spread of information concerning such safety concerns.⁷⁰

all fifteen-day reports, along with reports of other adverse experiences, and explanations of any action that the manufacturer has taken in response to reported information.

Barbara A. Noah, *Adverse Drug Reactions: Harnessing Experiential Data To Promote Patient Welfare*, 49 CATH. U. L. REV. 449, 470-71 (2000) (citing 21 C.F.R. § 314-80 (1999)). Noah notes that manufacturers of new drugs incur additional reporting obligations:

The regulations also require that holders of an approved [new drug application] submit quarterly adverse drug experience reports for the first three years of marketing and annual reports afterwards Finally, additional regulations for new drugs require that manufacturers submit a brief summary of new information accumulated during the preceding year that “might affect the safety, effectiveness, or labeling of the drug product” along with a description of the manufacturer’s intended response to this information.

Id. at 471 (citing 21 C.F.R. §§ 314-80 to -81 (1999)).

67. See Thomas Scarlett, *The Relationship Among Adverse Drug Reaction Reporting, Drug Labeling, Product Liability, and Federal Preemption*, 46 FOOD DRUG COSM. L.J. 31, 35 (1991) (noting that “there are severe regulatory and other penalties” for violating FDA’s reporting rules, and that “product liability pressure . . . pushes in the direction of reporting everything that could conceivably be reported as an [adverse drug reaction] and making sure it shows up in the labeling”).

68. Marie R. Griffin et al., *Commentary: Postmarketing Surveillance for Drug Safety: Surely We Can Do Better*, 75 CLINICAL PHARMACOLOGY & THERAPEUTICS 491, 492 (2004). The authors are investigators at the Centers for Education and Research on Therapeutics. See *id.* at 494.

69. Sage, *supra* note 46, at 1019-20; cf. Green, *supra* note 46, at 488 (noting that “the pharmaceutical industry’s history is littered with instances of deliberate or negligent withholding of information from the FDA in the new drug approval process”).

70. See, e.g., Gardiner Harris, *F.D.A. Failing in Drug Safety, Official Asserts*, N.Y. TIMES, Nov. 19, 2004, at A1 (noting the existence of documents in which “Merck executives and scientists discussed the possible link between Vioxx and heart damage years before the company publicly

The FDA also relies upon health professionals to identify potential problems. To this end, it created the MedWatch program, which solicits reports from health professionals regarding deaths or serious injuries associated with drugs, medical devices or other regulated products;⁷¹ reports received through this program are evaluated and entered into databases.⁷² The system, however, is plagued by underreporting.⁷³ For one thing, doctors may notice unexpected harms, but they are less likely to discern an increase in the probability of familiar harms.⁷⁴ For another, doctors may be unwilling to report events that might get them into trouble.⁷⁵ (As some medical devices are marketed for use outside of medical settings, the likelihood of spontaneous reporting decreases still further.⁷⁶)

admitted that the drug could cause harm”); Anna Wilde Mathews & Barbara Martinez, *Warning Signs: E-mails Suggest Merck Knew Vioxx’s Dangers at Early Stage*, WALL ST. J., Nov. 1, 2004, at A1 (stating that “internal Merck e-mails and marketing materials as well as interviews with outside scientists show that the company fought forcefully for years to keep safety concerns from destroying the drug’s commercial prospects”); Barry Meier, *Questions Are Seen on Merck’s Stance on Pain Drug’s Use*, N.Y. TIMES, Nov. 24, 2004, at A1 (stating that Merck was aware, “as far back as 2001,” that Vioxx might not provide gastrointestinal benefits for older users who were also taking aspirin regularly, and that Merck “never followed up with a plan in 2001 to run a definitive test about the drug’s advantages, if any, to aspirin users”).

71. See Brewer & Colditz, *supra* note 50, at 825.

72. See RISK MANAGEMENT REPORT, *supra* note 23, at 54-55 (discussing Adverse Event Reporting System database for drugs and biological products); *id.* at 58 (discussing the Manufacturer and User Device Experience database for medical devices).

73. See U.S. GEN. ACCOUNTING OFFICE, ADVERSE EVENTS; SURVEILLANCE SYSTEMS FOR ADVERSE EVENTS AND MEDICAL ERRORS: STATEMENT OF JANET HEINRICH 3 (2000), <http://www.gao.gov/archive/2000/he00061t.pdf> (noting that the “FDA believes that its . . . Adverse Event Reporting System . . . receives reports for only about 1 to 10 percent of all [adverse drug events]”); Brewer & Colditz, *supra* note 50, at 825 (“[S]erious adverse events that may represent [adverse drug reactions] are underreported by physicians to either manufacturers or the FDA.”).

74. See Brewer & Colditz, *supra* note 50, at 825 (“Unusual . . . events that occur during initial or long-term drug use are more likely to be detected by case reports than increases in common events or events that occur remotely in time from the medication use.”); Griffin et al., *supra* note 68, at 492 (noting that “voluntary reports are less likely to be helpful in determining whether a drug causes or increases the severity of a condition that is relatively common in the background population”).

75. See Roxana Mehran et al., *Post-Market Approval Surveillance: A Call for a More Integrated and Comprehensive Approach*, 109 CIRCULATION 3073, 3074 (2004) (“[P]hysicians encountering adverse events while performing off-label procedures may be reluctant to call undue attention to themselves for using a device in an unapproved manner.”); Sage, *supra* note 46, at 1022 (“Fear of malpractice liability discourages reporting.”).

76. See Barry Meier, *Flawed Device Places F.D.A. Under Scrutiny*, N.Y. TIMES, Dec. 15, 2004, at A1 (noting that “[t]he growing use of [defibrillators] in settings like offices, schools and homes puts them outside the [FDA]’s problem-reporting system”).

Overreporting is an issue as well: The MedWatch system generates some 22,000 reports each year,⁷⁷ and of these a substantial number may not involve a causal link between the product and the injury. The quality of the reports can limit their usefulness: "Much of the data FDA receives do not allow a complete understanding of the problems associated with an adverse event or allow the Agency to be proactive in protecting the public."⁷⁸

More generally, the FDA's reporting programs generate a deluge of information. Annually, the agency has received more than 200,000 adverse event reports regarding drugs or biologic products, and more than 80,000 adverse event reports concerning devices.⁷⁹ It is thus unsurprising that the agency describes its analysis of this flood of data as "triage,"⁸⁰ and that the agency laments the difficulty of its task: "Like the proverbial search for a needle in a haystack, the number and variety of products and the lack of reliable usage information, make it difficult to distinguish variability and *noise* from a real concern. . . . More work in this area is needed."⁸¹ But though more work is needed, the resources

77. See Michelle Meadows, *MedWatch: Managing Risks at the FDA*, FDA CONSUMER MAG., Sept.-Oct. 2003, http://www.fda.gov/fdac/features/2003/503_risk.html.

78. RISK MANAGEMENT REPORT, *supra* note 23, at 63-64.

79. See *id.* at 54 ("In FY 1998, more than 230,000 reports of suspected adverse events were received by [the Adverse Event Reporting System]."); *id.* at 58 ("The Agency receives approximately 80,000 to 85,000 device-related adverse event reports every year."). The numbers appear to be increasing. See David W. Feigal, et al., *Ensuring Safe and Effective Medical Devices*, 348 NEW ENG. J. MED. 191, 191 (2003) ("The FDA received more than 120,000 [device-related] reports in 2002.").

80. RISK MANAGEMENT REPORT, *supra* note 23, at 58 ("When received, [medical device] reports are first triaged by medical professionals."). The agency's "triage" efforts include some measures designed to make the flow of information more manageable by cutting its volume. The FDA permits "summary reporting" of events concerning some medical devices with "well-documented adverse event histories." *Id.* at 58-59. Statutory changes in 1997, see Food and Drug Administration Modernization Act of 1997, 21 U.S.C. § 360i(b)(5) (2000), "direct[] the FDA to move away from universal, mandatory adverse event reporting by user facilities to a system based on reporting by a representative sample of facilities," RISK MANAGEMENT REPORT, *supra* note 23, at 53. In addition, the FDA is improving its electronic data systems and is seeking ways to use technology to look for emerging safety issues. See *id.* at 3 (noting that the "FDA has initiated several changes in the adverse event reporting system, such as consolidating reporting system components and using electronic reporting"); see also *FDA To Use Data Mining To Monitor Adverse Events*, 22 BIOTECHNOLOGY L. REP. 481, 481 (2003) (reporting that the FDA "has signed a Cooperative Research and Development Agreement . . . with Lincoln Technologies, Inc. . . . to use safety data as an early indicator of populations at particular risk of adverse effects and of drug interactions," and stating that "[t]he data mining will be applied to information the FDA collects from postmarket reports").

81. RISK MANAGEMENT REPORT, *supra* note 23, at 67-68.

necessary to perform the task are sorely lacking.⁸² Observers assert that the FDA's funding arrangements have led it to privilege new drug approval while starving the CDER's postmarketing surveillance arm.⁸³

Other critics suggest that the FDA suffers not only from a lack of resources but also from a lack of will to pursue safety issues aggressively. David Graham, the Associate Director for Science and Medicine in the FDA's Office of Drug Safety, has charged that the CDER resists airing safety concerns about approved drugs, both because the officials who approved the drug wish not to be proven wrong and because upper-level managers in the Office of Drug Safety tend to support the positions taken by those officials.⁸⁴

In summary, though the FDA has made efforts to improve its postmarketing surveillance,⁸⁵ more should be done. The problem of insufficient resources

82. See Green, *supra* note 46, at 499 ("If the FDA had adequate resources to monitor manufacturer post-approval reporting behavior, detect violations, impose adequate sanctions, and thereby provide an appropriate deterrent, we could be more sanguine about the efficacy of the [adverse reaction reporting] process. But, once again, there is the problem of inadequate regulatory resources.").

83. See Gardiner Harris, *At F.D.A., Strong Drug Ties and Less Monitoring*, N.Y. TIMES, Dec. 6, 2004, at A2 (noting that the FDA no longer has the resources to fund independent studies of emerging safety issues and that "[i]n the past 11 years, spending on [new drug] reviews has increased to more than four-fifths of the agency's drug center budget from about half"). Citing figures from 1997, Barbara Noah has observed that "the FDA only devotes the equivalent of fifty-five full-time employees to post-approval surveillance, as compared with over 1700 full-time equivalents engaged in pre-market review of new drug applications." Noah, *supra* note 66, at 452.

84. See *Merck and Vioxx: Putting Safety First?: Hearing Before the House Comm. on Finance*, 108th Cong. (2004) [hereinafter *Vioxx Hearings*] (statement of David J. Graham, Assoc. Dir. for Sci., Office of Drug Safety, FDA).

85. Efforts continue to be made to strengthen the FDA's postmarketing oversight. For example, the FDA can premise its approval of a product on the company's commitment to perform postmarketing studies. Concerns were raised in the mid-1990s about the FDA's capacity to supervise such studies. REPORT TO CONGRESS: REPORTS ON POSTMARKETING STUDIES [FDAMA 130] 5-6 (2001), <http://www.fda.gov/cber/fdama/pstmrktfdama130.pdf>. In response, Congress in 1997 expanded the FDA's authority to follow up on drug and biologic postmarketing studies. See 21 U.S.C. § 356b (2000); Griffin et al., *supra* note 68, at 492. Some have suggested that those studies have not yet fulfilled their potential, *see id.* at 492 ("As of February 2002, only 37% of the 2400 postmarketing commitments for new drugs had been completed and many had never been started. Despite changes in FDA procedures, potential concerns or 'signals' generated before licensing can still remain unexplored for years after marketing."), and an FDA official recently remarked that the FDA "ha[s] very little authority to make sure those postmarketing commitments are carried out," Denise Grady, *A Medical Journal Calls for a New Watchdog on Drugs*, N.Y. TIMES, Nov. 23, 2004, at A1 (quoting Sandra Kweder, Deputy Director of the FDA's Office of New Drugs). *But cf.* FDA Report on the Performance of Drug and Biologics Firms in Conducting

persists, as does the concern that the FDA may be loath to move swiftly to address emerging safety issues.⁸⁶ Commentators have suggested a number of measures that might help: For example, Congress could create a new regulatory body—independent of the FDA’s medical product approval arm—that would be devoted to postmarketing surveillance.⁸⁷ The Institute of Medicine, which has been asked to review the FDA’s postmarketing surveillance system,⁸⁸ may suggest other measures. But even if changes are made, it is likely that litigation will continue to play an important role in identifying and substantiating problems. In the next Part, I consider whether the litigation and regulatory processes might be restructured so as to improve the FDA’s postmarketing surveillance.

Postmarketing Commitment Studies; Availability, 69 Fed. Reg. 12,162, 12,163 & tbl.1 (Mar. 15, 2004) (asserting that though 1338 drug postmarketing commitments remain open, and studies have not yet begun with respect to 864 of those commitments, only twenty-one are “delayed” in the sense that “[t]he study is behind the original schedule”).

86. As a group of doctors recently noted with respect to device regulation, “[w]hereas large resources have been devoted to . . . early development and clinical evaluation . . . , few resources have been focused on post-market surveillance” Mehran et al., *supra* note 75, at 3073.

87. See *Vioxx Hearings*, *supra* note 84 (statement of Bruce M. Psaty, M.D., Ph.D., Professor, Medicine, Epidemiology & Health Servs.); see also Editorial, *Looking for Adverse Drug Effects*, N.Y. TIMES, Nov. 27, 2004, at A14 (“Critics have proposed a wide range of reforms—a more active search for adverse consequences, increasing the power of the safety office within the F.D.A., ending the agency’s reliance on user fees from the industry and establishing a wholly independent drug safety board . . .”). Requiring advance registration of all drug trials would reduce a company’s ability to suppress adverse information through confidentiality agreements with researchers. See Barry Meier, *Contracts Keep Drug Research Out of Reach*, N.Y. TIMES, Nov. 29, 2004, at A1. There have been many additional proposals. See, e.g., RISK MANAGEMENT REPORT, *supra* note 23, at 14-15 (listing options); U.S. GEN. ACCOUNTING OFFICE, ADVERSE DRUG EVENTS: THE MAGNITUDE OF HEALTH RISK IS UNCERTAIN BECAUSE OF LIMITED INCIDENCE DATA 18 (2000) (noting proposal to “establish[] a network of health care facilities to serve as ‘sentinel sites’ for closely monitoring the experiences of the first patients to take a new drug”); Brewer & Colditz, *supra* note 50, at 827-28 (suggesting that measures such as meta-analysis of existing studies, and analysis of information in large databases, may help to identify problem drugs); Mehran et al., *supra* note 75, at 3076 (suggesting that claims databases and device registries may provide safety information); Sage, *supra* note 46, at 992 (proposing that knowledge of drug risks “can be improved by (1) medical structures such as HMOs, which can gather information about delayed or low probability adverse drug reactions, and (2) intelligently selected legal rules governing physician competence and manufacturers’ profit incentives”).

88. See Harris, *supra* note 83.

II. LINKING THE REGULATORY AND LITIGATION SYSTEMS

In theory, the regulatory and litigation systems could operate entirely independently: Compliance with regulations would be irrelevant in litigation, and litigation outcomes would not directly affect agency regulation.⁸⁹ Few, however, would advocate total independence. It seems clear that the FDA's expert assessments of product safety should not be irrelevant in litigation arising from alleged safety defects. Rather, the dispute is over what the *effects* of the FDA safety determinations should be.

As noted above,⁹⁰ some argue that the FDA's expert balancing of product risks and benefits leaves no room for disagreement within the tort system. In this view, there is no reason for judges or juries to second-guess the FDA's judgments, and, indeed, second-guessing is likely to produce undesirable results because of the limited capabilities and circumscribed perspective of a civil jury.

Others, however, point out that the FDA cannot discern and address all product safety issues ahead of time, and that the agency may not act quickly enough to address those issues when they arise after a product enters the market. Even if agency capture does not inhibit the FDA's investigation of a safety problem, other limits on the FDA's postmarketing surveillance capacity may produce a similar effect. Scholars have also noted a substantial body of data that suggests juries do better at assessing technical and scientific questions than their critics assert.⁹¹

Courts considering the effects of FDA determinations have struggled to balance these competing considerations and have developed a number of doctrinal methods for doing so. FDA determinations can help a plaintiff establish a claim, but they may also help a defendant avoid liability. And in recent years, some—including, recently, the FDA itself—have asserted that certain types of FDA determinations ought to preclude litigation altogether.

Though there is no private right of action for violation of requirements imposed under the Food, Drug, and Cosmetic Act (FDCA),⁹² such a violation is hardly irrelevant in cases asserting products liability under state law. A violation of FDA-imposed requirements can be the basis for a finding of negligence *per*

89. Obviously, litigation itself has regulatory effects. *See, e.g.,* W. Kip Viscusi et al., *Detering Inefficient Pharmaceutical Litigation: An Economic Rationale for the FDA Regulatory Compliance Defense*, 24 SETON HALL L. REV. 1437, 1448 (1994) ("The common law regulates behavior through the imposition of damage awards against tortfeasors."). My point here, however, is to consider the extent to which litigation outcomes might operate independently from agency decisions.

90. *See supra* text accompanying notes 3-4.

91. *See supra* note 9.

92. 21 U.S.C. §§ 301-399 (2000); *see* *Ellis v. C.R. Bard, Inc.*, 311 F.3d 1272, 1284 n.10 (11th Cir. 2002) ("[N]o private right of action exists for a violation of the FDCA.").

se.⁹³ Even if the violation does not establish negligence *per se*, it can be considered to be evidence of negligence.⁹⁴

Conversely, some have argued that compliance with FDA requirements should establish a defense to negligence claims.⁹⁵ Under a regulatory compliance defense, “[m]anufacturers of drugs and extensively regulated devices would be shielded from liability by compliance with FDA regulations, including conformance with agreed-upon testing protocols and timely submission and complete, accurate description of all required information.”⁹⁶ Proponents assert that this system “would strengthen current incentives to comply with FDA regulations, while attenuating current incentives to exceed FDA safety standards.”⁹⁷ Acting upon such principles, some states have barred punitive damages where a defendant has met FDA requirements.⁹⁸

Similarly, some states have essentially rejected the notion that an FDA-approved drug can suffer from a design defect. An influential comment in the Restatement (Second) of Torts set the terms of the debate by asserting that many drugs are “unavoidably unsafe” and that the manufacturers of such products should not incur liability in the absence of manufacturing defects or inadequate warnings.⁹⁹ Most jurisdictions purport to follow this rule, but they disagree on its

93. See RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 4 note (1998) (“The overwhelming majority of American courts hold that violations of product safety regulations cause products to be defective as a matter of law in cases involving both design and failure to warn.”); *Lukaszewicz v. Ortho Pharm. Corp.*, 510 F. Supp. 961, 965 (E.D. Wis. 1981) (holding that a violation of FDA requirements concerning contraceptive labeling would establish negligence *per se* under Wisconsin law), *amended by* 532 F. Supp. 211 (E.D. Wis. 1981). *But see, e.g., King v. Danek Med. Inc.*, 37 S.W.3d 429, 460 (Tenn. Ct. App. 2000) (rejecting negligence *per se* claim that was based on defendant’s failure to obtain premarket approval or clearance for medical device).

94. See, e.g., *MacDonald v. Ortho Pharm. Corp.*, 475 N.E.2d 65, 71 (Mass. 1985) (stating in dictum that “violation of FDA requirements is evidence, but not conclusive evidence, of negligence”).

95. See, e.g., *Viscusi et al., supra* note 89, at 1478-80 (arguing that in the absence of fraud, compliance with FDA requirements should preclude tort liability).

96. STEVEN GARBER, *PRODUCT LIABILITY AND THE ECONOMICS OF PHARMACEUTICALS AND MEDICAL DEVICES*, at xxxii (1993). The defense could take several different forms. For example, compliance could provide a rebuttable presumption that liability should not attach. See, e.g., N.J. STAT. ANN. § 2A:58C-4 (West 2005) (“If the warning or instruction given in connection with a drug or device or food or food additive has been approved or prescribed by the [FDA] . . . a rebuttable presumption shall arise that the warning or instruction is adequate.”).

97. GARBER, *supra* note 96, at xxxii.

98. See *Viscusi et al., supra* note 89, at 1476 n.140 (citing statutes).

99. As the comment explained:

There are some products which, in the present state of human knowledge, are quite incapable of being made safe for their intended and ordinary use. These are especially

scope. Many courts engage in a case-by-case risk/benefit analysis to determine whether a particular drug or device is “unavoidably unsafe.”¹⁰⁰ Some other courts, however, have concluded that all prescription drugs should be viewed as “unavoidably unsafe,” such that the manufacturer should not be liable on a design defect theory.¹⁰¹ Though a blanket application of the rule seems less persuasive with regard to medical devices than with regard to pharmaceuticals,¹⁰² some courts have found whole categories of medical devices to be “unavoidably unsafe” as well.¹⁰³ A strong undercurrent in the case law broadly applying the

common in the field of drugs Such a product, properly prepared, and accompanied by proper directions and warning, is not defective, nor is it *unreasonably* dangerous.

RESTATEMENT (SECOND) OF TORTS, § 402A cmt. k (1965). The *Restatement (Third) of Torts: Products Liability* proposes a different test:

A prescription drug or medical device is not reasonably safe due to defective design if the foreseeable risks of harm posed by the drug or medical device are sufficiently great in relation to its foreseeable therapeutic benefits that reasonable health-care providers, knowing of such foreseeable risks and therapeutic benefits, would not prescribe the drug or medical device for any class of patients.

RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY, § 6(c) (1997). This standard has been criticized by both courts, *see, e.g., Freeman v. Hoffman-La Roche, Inc.*, 618 N.W.2d 827, 839-40 (Neb. 2000) (reviewing objections to Section 6(c) and rejecting it because “recovery [under this standard] would be nearly impossible”), and commentators, *see, e.g., George W. Conk, Is There a Design Defect in the Restatement (Third) of Torts: Products Liability?*, 109 YALE L.J. 1087, 1089 (2000) (arguing that Section 6(c)’s “declaration that manufacturers of medical products need not make a safer product if the existing product does more good than harm reverses thirty-five years of safety-advancing products-liability law”). *But see James A. Henderson & Aaron D. Twerski, Drug Designs Are Different*, 111 YALE L.J. 151 (2001) (responding to Conk’s critique).

100. *See, e.g., Freeman*, 618 N.W.2d at 840 (holding in a prescription drug case that comment k will provide an affirmative defense “when it is shown that (1) the product is properly manufactured and contains adequate warnings, (2) its benefits justify its risks, and (3) the product was at the time of manufacture and distribution incapable of being made more safe”); *Tansy v. Dacomed Corp.*, 890 P.2d 881, 886 (Okla. 1994) (applying similar test in medical device case).

101. *See, e.g., Grundberg v. Upjohn Co.*, 813 P.2d 89, 99 (Utah 1991).

102. Alternative designs of prescription drugs may often be impossible to find or create. *But see* 5 LOUIS R. FRUMER & MELVIN I. FRIEDMAN, *PRODUCTS LIABILITY* § 50.03A[3], at 50-29 (2004) (noting that birth control pills “can be designed in many different ways”). However, it seems likely that alternative designs of many medical devices could be pursued. *See* GARBER, *supra* note 96, at xxviii (noting that medical devices “can often be made safer at low or moderate costs”). Take for example the variety of possible designs for intrauterine devices (IUDs). *See* RONALD J. BACIGAL, *THE LIMITS OF LITIGATION: THE DALKON SHIELD CONTROVERSY* 10 (1990) (describing the choice between monofilament and multifilament tail strings for IUDs and explaining that the Dalkon Shield’s multifilament tail strings “wicked” bacteria into the uterus).

103. *See, e.g., Hufft v. Horowitz*, 5 Cal. Rptr. 2d 377, 384 (Cal. Ct. App. 1992) (holding that “all implanted medical devices” should be viewed as unavoidably unsafe). *But see* GARBER, *supra*

“unavoidably unsafe” notion is that the FDA’s approval of a medical product evidences an authoritative judgment that the product’s benefits outweigh its risks.¹⁰⁴

Even in the absence of structural connections between the litigation and regulatory systems, then, strong substantive connections exist. Violation of FDA requirements can help establish liability, while compliance can sometimes help defend against a claim or mitigate its damages. Some advocates of “tort reform,” however, contend that the regulatory-adjudicative relationship must be structured more formally through the mechanism of preemption. Under the current system, when FDA regulation preempts state tort claims, the regulatory system displaces the litigation system. Because no federal cause of action currently exists, preempting state tort claims eliminates the potential for lawsuits concerning product safety.

Questions of preemption currently turn upon both the nature of the claim and the degree of prior FDA scrutiny of the product. Claims seeking damages from a company that violated FDA requirements are not preempted.¹⁰⁵ Nor are claims challenging the safety of a medical device approved under the relatively streamlined “substantial equivalence” process.¹⁰⁶ Claims asserting that the defendant perpetrated a fraud on the FDA, however, are preempted.¹⁰⁷ And while there is a circuit split on the question of preemption for claims with respect to medical devices that have survived the more rigorous “premarket approval” process, the emerging majority view is that such claims are impermissible.¹⁰⁸ In

note 96, at 39 (“It is unclear whether comment k has been as widely applied to devices as to drugs and biologicals.”).

104. For example, the *Grundberg* court explained its holding as follows:

In light of the strong public interest in the availability and affordability of prescription medications, the extensive regulatory system of the FDA, and the avenues of recovery still available to plaintiffs by claiming inadequate warning, mismanufacture, improper marketing, or misrepresenting information to the FDA, we conclude that a broad grant of immunity from strict liability claims based on design defects should be extended to FDA-approved prescription drugs in Utah.

Grundberg, 813 P.2d at 99.

105. See *Medtronic, Inc. v. Lohr*, 518 U.S. 470, 495 (1996) (holding that “[n]othing . . . denies Florida the right to provide a traditional damages remedy for violations of common-law duties when those duties parallel federal requirements”).

106. See *id.* at 494.

107. See *Buckman Co. v. Plaintiffs’ Legal Comm.*, 531 U.S. 341, 348 (2001) (holding that “state-law fraud-on-the-FDA claims conflict with, and are therefore impliedly pre-empted by, federal law”). But see McGarity, *supra* note 44, at 572 (arguing that *Buckman*’s holding should be narrowly construed).

108. Compare *Horn v. Thoratec Corp.*, 376 F.3d 163 (3d Cir. 2004) (finding preemption where

the area of prescription drugs, although most lower courts have rejected the preemption defense in failure-to-warn cases, at least one has disagreed.¹⁰⁹

The FDA, in its notorious 2002-2004 court filings, took up the defendants' side of the argument in both of the latter disputes.¹¹⁰ Moreover, there are indications that the Bush Administration intends to expand the reach of preemption in other ways. A bill introduced in the Senate during the 108th Congress would have immunized manufacturers from punitive damages in connection with medical products unless the plaintiff shows by clear and convincing evidence that the manufacturer violated a specific requirement imposed under the FDCA.¹¹¹ Despite recent events concerning Vioxx and other controversial FDA-approved drugs, it appears likely that the Administration will continue to press for passage of this measure.¹¹² Thus, it continues to be important to assess the arguments of those who support preemption of claims for medical products liability.

As can be seen from this summary, each proposal to take tort claims away from civil juries rests upon the assertion that jury determinations of product safety are at best duplicative—because the FDA exists to make just such safety assessments—and at worst harmful because unwarranted jury awards can deter companies from developing and marketing useful products. But, as I have noted, those positions have been subjected to powerful critiques. In addition to

medical device went through premarket approval process); *Medtronic, Inc.*, 254 F.3d 573 (5th Cir. 2001) (same); *Brooks v. Howmedica, Inc.*, 273 F.3d 785 (8th Cir. 2001) (same); *Kemp v. Medtronic, Inc.*, 231 F.3d 216, 227 (6th Cir. 2000) (same); *Mitchell v. Collagen Corp.*, 126 F.3d 902 (7th Cir. 1997) (same); *Papike v. Tambrands, Inc.*, 107 F.3d 737 (9th Cir. 1997) (same), *with Goodlin v. Medtronic, Inc.*, 167 F.3d 1367 (11th Cir. 1999) (finding no preemption).

109. *Compare* *Motus v. Pfizer, Inc.*, 127 F. Supp. 2d 1085, 1092 (C.D. Cal. 2000) (rejecting preemption defense in prescription drug failure-to-warn case and noting that “most courts have found that FDA regulations as to design and warning standards are minimum standards which do not preempt state law defective design and failure to warn claims”), *with* *Ehls v. Shire Richwood, Inc.*, 233 F. Supp. 2d 1189, 1198 (D.N.D. 2002) (holding failure-to-warn claim preempted because “[t]he FDA dictates the contents of the label for Adderall® and defendants were prohibited from changing it without prior approval from the FDA, except in limited circumstances for a limited period of time”), *aff’d on other grounds*, 367 F.3d 1013 (8th Cir. 2004).

110. *See, e.g., Horn*, 376 F.3d at 177-79 (describing FDA’s arguments in support of preemption, based on FDA premarket approval, in a medical device case); Brief of Amicus Curiae United States at 2, *Motus v. Pfizer, Inc.*, 358 F.3d 659 (9th Cir. 2002) (No. 02-55372) (advocating preemption, in a prescription drug case, because “[t]o require a warning of a supposed danger that FDA concludes has no actual scientific basis, no matter the warning’s language, would be to require a statement that would be false or misleading, and thus contrary to federal law”).

111. Patients First Act of 2003, S. 11, 108th Cong. § 7(c)(1).

112. *See* Bob Herbert, *A Gift for Drug Makers*, N.Y. TIMES, Jan. 14, 2005, at A23 (criticizing the Administration’s position).

challenging the notion of jury incompetence, commentators have argued persuasively that some amount of redundancy is desirable: The tort system should remain free to redetermine product safety in the light of information developed during litigation, because the FDA may not always uncover relevant safety information and may not act quickly enough upon the information that it does receive.

Proposals for FDA regulation to displace tort litigation—either through preemption or through a regulatory compliance defense—cannot fully meet this objection. A carefully designed regulatory compliance defense might attempt to improve postmarketing surveillance, for instance by precluding liability if and only if the defendant had complied with regulatory requirements, including disclosure requirements.¹¹³ Thus, proponents have urged that the defense should be available only where there was full disclosure.¹¹⁴ But even with this caveat, the effectiveness of such a system would require that the FDA act quickly and effectively to address all indications of emerging safety problems. Especially in the light of recent questions concerning the FDA's performance, this assumption seems unduly optimistic. A regulatory compliance defense would remove a company's incentive to work proactively to address emerging safety issues; to avoid liability, the company would simply have to disclose any relevant information to the FDA.¹¹⁵ And such disclosures might well not facilitate the FDA's task: A system in which disclosure provided immunity would encourage companies to inundate the FDA with information.

So long as the regulatory and litigation systems remain structurally separate, the policy debate may have reached an impasse: In order to privilege FDA

113. Michael Green has pointed out that incorporating such nuances into the regulatory compliance defense will render that defense complex and costly to litigate. See Green, *supra* note 46, at 507-08.

114. See 2 AM. LAW INST., REPORTERS' STUDY, ENTERPRISE RESPONSIBILITY FOR PERSONAL INJURY 97 (1991) [hereinafter ALI REPORTERS' STUDY] (arguing that a regulatory compliance defense should apply only if the defendant "publicly disclosed to the relevant regulatory agency any material information . . . of which it has reason to be aware . . . concerning the risks posed by the defendant's activities and/or the means of controlling them," and stating that the requirement should "extend to information indicating that agency standards or tests may be inadequate or inappropriate").

115. As Michael Green has noted:

With a regulatory compliance defense available, manufacturers would no longer have an incentive to seek labeling changes that would disclose additional risks discovered in the post-marketing period. The impetus for such changes would be left to the FDA The specter of inadequate resources available to the FDA makes this role reversal of significant concern.

Green, *supra* note 46, at 502.

decisionmaking, one must lose the added benefits that the tort system could provide. In particular, eliminating litigation would deprive regulators of a potentially useful source of information on product safety¹¹⁶ and would repose in the FDA and the regulated companies a level of trust that seems unwarranted in the light of recent events. In this Part, I explore some structural options that could offer a way around the dilemma. Each of the options I consider would preserve some opportunity for persons injured by medical products to obtain compensation, and some of those options would also preserve a role for the private plaintiffs' bar in bringing safety problems to light. As I will argue, a system that preserves those compensatory and monitoring functions is preferable to preemption, which would sacrifice both.

I will compare four ways in which Congress could link litigation to regulation. In each option described here, Congress would preempt state tort claims and substitute a federal cause of action. On the assumption that one goal would be to submit safety and causation questions to the FDA (or other expert agency) for resolution, each of the options described here would incorporate agency determinations of liability. A basic question in that regard is whether, in light of the fact that key liability questions would be determined by the FDA, the rest of the proceeding should unfold within an agency setting, or whether the suit should be litigated in federal court with a mechanism for referring specific questions to the FDA. I first consider two options for situating the adjudication within the agency itself; I then outline two possible frameworks for litigation in federal court. Finally, I compare the four options in the light of a number of constitutional and policy considerations.

Before embarking on this comparison, I should note that my discussion assumes that the tort system should seek to apply the same substantive standard, and roughly the same evidentiary requirements, that the FDA employs in making its safety determinations. This assumption is, of course, debatable; it is not uniformly reflected in current state tort law, and it need not guide the choice of substantive and evidentiary standards under a new federal cause of action either. However, much of the debate over the interaction between FDA regulation and tort liability presumes that the standards should be the same and focuses on asserted flaws in one or the other institution's application of those standards. My project is not to defend a particular choice of substantive liability rules, but rather to examine whether structural changes could improve the application of the chosen standard. On that premise, I will proceed to consider possible alternatives.

116. *Cf. id.* at 482 ("Sometimes it is the tort system that uncovers instances of noncompliance with FDA regulatory standards, rather than the FDA itself.").

A. Agency Adjudication

One possible approach would be to situate the adjudication of product safety claims within the agency itself. Such adjudication could proceed on the government's initiative; in addition, Congress could authorize private persons to bring claims.

1. Agency Enforcement

In an agency enforcement model, Congress would preempt private state tort claims and replace them with a claim by the government for penalties. The germ of such a penalty system already exists within the framework of the FDCA.

Though the FDA's principal enforcement options include injunctions, civil seizures, and criminal penalties,¹¹⁷ it also has authority to seek civil penalties¹¹⁸ for violations of certain laws governing prescription drugs¹¹⁹ and medical devices.¹²⁰ Civil penalty proceedings begin with a complaint by the relevant center within the FDA.¹²¹ The respondent can request a hearing, at which it can be represented by counsel.¹²² The presiding officer at the hearing has the power to subpoena witnesses and evidence.¹²³ Discovery is more circumscribed than in civil court proceedings: Though parties can obtain discovery of documents if they establish that the documents are "relevant to the issues before the presiding officer,"¹²⁴ to obtain permission to take depositions, they must show that the

117. See Megan Hanley Baer & Whitney Moore, *Federal Food and Drug Act Violations*, 40 AM. CRIM. L. REV. 613, 631-33 (2003) (discussing criminal penalties for violations of the FDCA).

118. In addition, at least one court has held that the FDA can bring a claim for restitution arising from violations of the FDCA. See *United States v. Universal Mgmt. Servs., Inc.*, 191 F.3d 750, 762 (6th Cir. 1999) (holding that "nothing in the FDCA precludes a court sitting in equity from ordering restitution in appropriate cases"). This view, however, is not universally shared. See *id.* at 761 (noting "a number of district court cases that determine that recalls and disgorgement are unavailable under the FDCA"); see also Jeffrey N. Gibbs & John R. Fleder, *Can FDA Seek Restitution or Disgorgement?*, 58 FOOD & DRUG L.J. 129, 147 (2003) (criticizing *Universal Management Services*).

119. See 21 U.S.C. § 333(b)(2)-(3) (2000).

120. See *id.* § 333(f)(1)(A).

121. See 21 C.F.R. § 17.5 (2004).

122. See *id.* §§ 17.9, 17.15.

123. See *id.* §§ 17.19, 17.27.

124. *Id.* § 17.23. This standard may roughly correspond to the current presumptive standard in federal civil practice. See FED. R. CIV. P. 26(b)(1) (setting general rule that "[p]arties may obtain discovery regarding any matter, not privileged, that is relevant to the claim or defense of any party"). However, in federal litigation a party may be able to obtain a court order authorizing broader discovery. See *id.* (providing that "[f]or good cause, the court may order discovery of any

information sought is not available in some other way and that “relevant and probative evidence may otherwise not be preserved for presentation by a witness at the hearing.”¹²⁵ Direct testimony at the hearing is given in writing, but cross-examination occurs through live testimony.¹²⁶ The evidentiary rules are more relaxed than those applied in federal court, though the hearing officer may draw upon the Federal Rules of Evidence for guidance.¹²⁷ Liability and size of penalty must be proven by a preponderance of the evidence.¹²⁸ Either side may appeal the hearing officer’s decision within the FDA; findings of fact are reviewed for “substantial evidence” and conclusions of law are reviewed *de novo*.¹²⁹ A dissatisfied respondent may then seek judicial review of the final agency decision.¹³⁰

Though the current system provides a starting point for an agency enforcement model, it would require some adjustment. Penalties available under current law are directed toward deterrence but not compensation. Neither the maximum allowable penalties¹³¹ nor the factors to be considered¹³² relate to the extent of harm caused by a violation. The money recovered goes into the general treasury,¹³³ not toward compensation of injured persons. If an agency enforcement proceeding were to substitute for private civil actions, the amount of the penalties could be keyed to the level of damages incurred by consumers, and the proceeds could be earmarked for distribution to injured persons. The other

matter relevant to the subject matter involved in the action”).

125. 21 C.F.R. § 17.23 (2004).

126. *See id.* § 17.37.

127. *See id.* § 17.39.

128. *See id.* § 17.33. The decisionmaker must consider aggravating and mitigating circumstances and articulate the reasons for the chosen penalty. *See id.* § 17.34.

129. *See id.* § 17.47.

130. *See id.* § 17.51. Because the FDA’s penalty procedure includes a public hearing, *see id.* § 17.33(d) (“The hearing shall be open to the public unless otherwise ordered by the presiding officer . . .”), and results in the development of an administrative record, it appears likely that the FDA’s decision would be reviewed in federal court using the “substantial evidence” standard set forth in 5 U.S.C. § 706(2)(E) (2000). *See Citizens to Preserve Overton Park, Inc. v. Volpe*, 401 U.S. 402, 414 (1971) (“Review under the substantial-evidence test is authorized only when the agency action is taken pursuant to a rulemaking provision of the Administrative Procedure Act itself . . . or when the agency action is based on a public adjudicatory hearing.”).

131. *See* 21 U.S.C. § 333(b)(2)-(3), (f)(1)(A) (2000) (setting maximum penalties); 21 C.F.R. § 17.2 (2004) (adjusting maximum penalties for inflation).

132. *See, e.g.,* 21 U.S.C. § 333(f)(3)(B) (2000) (with respect to penalties for medical device violations, factors include “the nature, circumstances, extent, and gravity of the violation or violations and, with respect to the violator, ability to pay, effect on ability to continue to do business, any history of prior such violations, [and] the degree of culpability”).

133. *See* 21 C.F.R. § 17.54 (2004).

major change would be that the agency's enforcement resources would have to be increased, as would its staffing for internal hearings.

2. *Private Enforcement*

Alternatively, Congress could preempt state tort claims, and substitute a system of private products liability claims that would be adjudicated within the agency. That adjudication could employ procedures similar to those discussed in Subsection II.A.1. above. Some differences, however, would arise from the presence of private plaintiffs in the suit. For example, it would be necessary to provide procedures to govern the joinder of multiple plaintiffs, either as named parties or as members of a plaintiff class. In non-class actions where liability was proven, damages would be determined on an individual basis; in class actions, a finding of liability might be followed by a determination of aggregate damages and the adoption of a set of guidelines for distributing those damages to class members. Although a private enforcement system of this type would remove the need for additional government enforcement resources, it would still entail a substantial increase in the number of agency personnel staffing the hearing process.

B. Hybrid Adjudication

Thus far, the discussion has assumed that the desirability of obtaining FDA resolution of safety and causation issues would dictate that the litigation should occur within the administrative system. An alternative, however, would be to permit claims to proceed in federal court, but refer certain issues to the FDA for resolution.¹³⁴ In a 1996 article, Richard Nagareda made a similar proposal for the treatment of mass torts.¹³⁵ He suggested that Congress enact a scheme under which the doctrine of "primary jurisdiction" would come into play when a particular mass tort resulted in federal litigation that merited consolidation by the

134. Existing rules provide somewhat analogous mechanisms. A court can refer a matter to the FDA for administrative determination, and if the Commissioner accepts the referral, the FDA can employ a range of procedures to determine the referred matter. *See* 21 C.F.R. § 10.60 (2004). Thus, for example, courts have referred to the FDA the question of whether a product falls within the definition of a "new drug" under 21 U.S.C. § 321(p) (2000). *See Weinberger v. Bentex Pharm., Inc.*, 412 U.S. 645, 653 (1973) ("[T]he District Court's referral of the 'new drug' and the 'grandfather' issues to FDA was appropriate, as these are the kinds of issues peculiarly suited to initial determination by the FDA.").

135. *See Nagareda, supra* note 15, at 353; *see also id.* at 359 (stating that "the FDA would seem to be a strong candidate for" inclusion in his proposal).

Panel on Multidistrict Litigation.¹³⁶

The primary jurisdiction doctrine requires a court to stay (or dismiss) an action so as to defer to agency determination of an issue (or a claim) in appropriate cases.¹³⁷ The rationales for deference to the agency may include a need for uniform agency determination of an issue,¹³⁸ as well as a recognition of superior agency expertise, particularly with respect to specialized facts within the agency's field of experience.¹³⁹ Primary jurisdiction has loomed large in certain areas of federal regulation; for example, the doctrine has played a prominent role in coordinating the reach of federal antitrust lawsuits with the authority of other federal regulatory schemes. The doctrine has not, however, yet been employed as a way to link the mass tort and regulatory systems. Nagareda proposed that the doctrine could be used as a way, in effect, to refer to an agency (such as the FDA) a question (such as general causation) that would benefit from the agency's decisionmaking.¹⁴⁰ The proposal I describe here is similar to Nagareda's in that it contemplates that litigation would be commenced in court but that the court would (at an appropriate juncture) refer issues of product safety and causation to the FDA for determination.

Discovery, in this system, would be supervised by the federal court. After discovery, the defendant could obtain summary judgment¹⁴¹ unless the plaintiff

136. *See id.* at 361 ("The primary jurisdiction doctrine would apply only if [the] litigation . . . progresses to the point that similar claims inundate the federal system, so many as to warrant consolidation by the MDL Panel and to trigger the opportunity to petition the relevant regulatory agency.").

137. *See, e.g., Reiter v. Cooper*, 507 U.S. 258, 268 (1993) (stating that where a claim "contain[s] some issue within the special competence of an administrative agency," the doctrine of primary jurisdiction "requires the court to enable a 'referral' to the agency, staying further proceedings so as to give the parties reasonable opportunity to seek an administrative ruling").

138. *See, e.g., Tex. & Pac. Ry. v. Abilene Cotton Oil Co.*, 204 U.S. 426, 440-41 (1907) (stressing need for uniform determinations concerning railroad rates). *But see Great N. Ry. v. Merchs. Elevator Co.*, 259 U.S. 285, 290-91, 294 (1922) (holding that prior resort to agency is not necessary in order to obtain uniform determination of a pure question of law, because review by the Supreme Court can ensure uniformity with respect to such questions).

139. *See, e.g., Far E. Conference v. United States*, 342 U.S. 570, 574-75 (1952) ("Uniformity and consistency in the regulation of business entrusted to a particular agency are secured . . . by preliminary resort for ascertaining and interpreting the circumstances underlying legal issues to agencies that are better equipped than courts by specialization, by insight gained through experience, and by more flexible procedure.").

140. *See Nagareda, supra* note 15, at 361.

141. As in current federal civil litigation, a defendant moving for summary judgment would "bear[] the initial responsibility of informing the district court of the basis for its motion, and identifying those portions of 'the pleadings, depositions, answers to interrogatories, and admissions on file, together with the affidavits, if any,' which it believes demonstrate the absence of a genuine

pointed either to evidence that the defendant violated FDA requirements or to new information concerning product harmfulness. To survive summary judgment on the latter ground, the plaintiff would have to show the existence of information, material to the product's safety, that the FDA did not consider when it initially approved the product. That showing, or a showing of facts from which a reasonable decisionmaker could infer that the defendant violated FDA requirements, would entitle the plaintiff to a bench trial.¹⁴²

The trial would be segmented, because the court would refer questions of product safety and causation to the FDA.¹⁴³ The FDA would initially submit the questions to an advisory committee for nonbinding determination. The advisory committee could resemble those currently employed by the FDA to assist it with new product reviews and other matters.¹⁴⁴ Advisory committees can enhance the accuracy of the FDA's decisionmaking and improve its credibility; they can also

issue of material fact." *Celotex Corp. v. Catrett*, 477 U.S. 317, 323 (1986).

142. Procedure under the proposed system would differ from ordinary summary judgment procedure for two reasons: first, because the system does not contemplate a jury trial, and second, because the system divides liability questions between the district judge and the FDA.

As to questions relegated to the district judge, a summary judgment motion might sometimes provide an occasion for the judge to resolve the questions without taking live testimony:

When evidentiary issues are in dispute, when the credibility of witnesses may be in issue, when conflicting evidence must be weighed, a full trial is clearly necessary regardless of whether it is a bench or jury trial But when the question for decision concerns drawing inferences from undisputed evidence, or interpreting and evaluating evidence to derive legal conclusions, a trial may not add to the judge's ability to decide.

WILLIAM W. SCHWARZER ET AL., *THE ANALYSIS AND DECISION OF SUMMARY JUDGMENT MOTIONS* 39 (1991).

By contrast, as to questions entrusted to the FDA, it might be appropriate for the judge to play even less of a role in screening cases through summary judgment than the judge would ordinarily play in a case where the right to a jury is asserted. In the context of jury trials, the judge plays the role of gatekeeper by determining the admissibility of expert testimony. Judges may be well suited, in comparison to juries, to serve such a function. However, as to questions that the proposed system would relegate to the FDA, little purpose would be served by requiring the judge rigorously to screen expert evidence for admissibility prior to sending the issues to the FDA: The advisory committee and agency officials are better equipped to assess such evidence.

143. Some safety and causation issues might be suitable for determination by the district court (assisted where necessary by a special master). For example, safety determinations could be straightforward in cases involving violations of existing FDA requirements. Also, one of the causation issues in failure-to-warn cases is whether the physician would have prescribed the product even if the appropriate warning had been given; that issue might not require resolution by the FDA. *See infra* note 332.

144. Such committees are governed by the Federal Advisory Committee Act, 5 U.S.C. app. 2 §§ 1-16 (2000), as well as by FDA regulations, *see* 21 C.F.R. §§ 14.1-.174 (2004).

provide an opportunity for stakeholder and public input concerning important decisions. Employing an advisory panel to assist the FDA in liability determinations could carry similar benefits.

Panel members would be selected through a public nomination process,¹⁴⁵ and would include researchers with relevant scientific and medical expertise. FDA committees can also include members selected to represent consumer, patient, and industry interests. The liability panel could include such stakeholder representatives, but it would be necessary to screen carefully for conflicts of interest and to protect against an appearance of bias. Conflicts screening would also be key as to medical experts.¹⁴⁶ Although members of FDA advisory committees are subject to federal disclosure and conflicts provisions that ban participation by those with financial interests in the outcome,¹⁴⁷ the conflict can be waived if “the official responsible for the employee’s appointment . . . certifies in writing that the need for the individual’s services outweighs the potential for a conflict of interest created by the financial interest involved.”¹⁴⁸ Critics charge that conflicts are routinely waived, even in instances where waiver is unwarranted.¹⁴⁹ Because the proposed system would place significant reliance

145. *Cf.* 21 C.F.R. § 14.82 (2004) (providing nomination process for voting members of standing advisory committees); *id.* § 14.84 (providing nomination process for nonvoting members of standing technical advisory committees).

146. *Cf.* Gardiner Harris & Alex Berenson, *10 Voters on Panel Backing Pain Pills Had Industry Ties*, N.Y. TIMES, Feb. 25, 2005, at A1 (“Ten of the 32 government drug advisers who last week endorsed continued marketing of . . . Celebrex, Bextra and Vioxx have consulted in recent years for the drugs’ makers, according to disclosures in medical journals and other public records.”). Partly as a result of the Bayh-Dole Act, which fostered ties between industry and academia, *see* ANGELL, *supra* note 47, at 8 (“The Reagan years and Bayh-Dole . . . transformed the ethos of medical schools and teaching hospitals One of the results has been a growing pro-industry bias in medical research . . .”), a large proportion of biomedical researchers derive material benefits from their relations with industry. A recent analysis found multiple studies documenting ties between researchers and industry: “Studies suggest that 23% to 28% of academic investigators in biomedical research receive research funding from industry. A 1998 survey found that 43% of investigators also receive research-related gifts Approximately one third of investigators at academic institutions have personal financial ties with industry sponsors.” Justin E. Bekelman et al., *Scope and Impact of Financial Conflicts of Interest in Biomedical Research*, 289 JAMA 454, 456 (2003).

147. *See* 18 U.S.C. § 208(a) (2000).

148. *Id.* § 208(b)(3); *see also* FDA, Policies and Procedures for Handling Conflicts of Interest with FDA Advisory Committee Members, Consultants, and Experts, at <http://www.fda.gov/oc/advisory/conflictsofinterest/policies.html> (last visited Sept. 19, 2004).

149. *See* ANGELL, *supra* note 47, at 210 (stating that the FDA “regularly waives [the conflicts rules] on the unlikely grounds that someone’s advice is indispensable”). Angell cites a *USA Today* study that “examined FDA hearing records in 2000 and found that ‘at 92 percent of the meetings at

on the panel process, measures should be taken to ensure that conflicts are identified, and waivers should be granted only rarely and only upon a rigorous showing of necessity.

Panel meetings would presumptively be open to the public,¹⁵⁰ and they could include an opportunity for public comment.¹⁵¹ The panel would consider evidence submitted by the parties, and could request additional information that it considered necessary. The panel's determinations would be reviewed by the FDA, which would render the final determination concerning the safety and causation questions. Regarding safety, the FDA would determine whether the product is too dangerous to remain on the market, and whether (if the product is worth keeping on the market) it should be subjected to restrictions such as additional safety warnings. The FDA could address questions of causation by listing the factors and analysis that would determine whether a particular person's injury was caused by the defect in question.

The FDA's determination would be sent to the district court. The judge would review the FDA's safety and causation findings to ensure that they were supported by some evidence and that the agency had complied with the procedural requirements described above. If warranted, the district court would then apply the FDA's causation guidance, assess damages, and enter judgment.

1. Private Claims

Congress might attempt to use the hybrid system to adjudicate private claims. To accomplish this, Congress would preempt state tort claims and substitute a federal products liability claim that could be brought in federal district court. The claim would be adjudicated using the procedures described above.

Though this option has the advantage of being relatively uncomplicated, it would be vulnerable to constitutional challenge (as I explain in Subsection II.C.1.c below). Thus, it is worthwhile to consider whether a constitutionally permissible alternative exists.

least one member had a financial conflict of interest,' and 'at 55 percent of meetings, half or more of the FDA advisers had conflicts of interest.'" *Id.* (quoting Dennis Cauchon, *FDA Advisors Tied to Industry*, USA TODAY, Sept. 25, 2000, at A1).

150. *Cf.* Federal Advisory Committee Act, 5 U.S.C. app. 2 § 10 (2000) (setting presumption of openness); 21 C.F.R. § 14.20 (2004) (providing for public notice of advisory committee meetings); *id.* § 14.22 (providing for public access to the open portion of advisory committee meetings). Portions of the meetings could be closed as necessary. *Cf. id.* § 14.25 (providing for closed presentation of data and closed deliberations under certain circumstances).

151. *Cf. id.* § 14.29 (providing opportunity for public comment at committee meetings).

2. *Qui Tam* Claims

Qui tam claims on behalf of the government could provide another way to harness litigation as a supplement to the regulatory process. Under such a system, state tort claims would be preempted but the United States would possess a *parens patriae* claim for harms to consumers. A *qui tam* mechanism would permit such claims to be initiated and litigated by a private person, subject to federal supervision and review.¹⁵² A portion of the defendant's damages payment would provide a bounty for the *qui tam* relator and the rest would fund an administrative compensation scheme for victims.

Congress can authorize the United States to sue as *parens patriae* to recover damages for injuries arising from a company's violations of federal law.¹⁵³ *Parens patriae* suits are an appropriate way for a government to protect the health and welfare of its citizens,¹⁵⁴ and suits concerning the safety of FDA-

152. Readers familiar with Jonathan Siegel's discussions of *qui tam* suits and state sovereign immunity will see that I am borrowing from him the notion of *qui tam* suits in *parens patriae*. See Jonathan R. Siegel, *Congress's Power To Authorize Suits Against States*, 68 GEO. WASH. L. REV. 44 (1999) [hereinafter Siegel, *Suits Against States*]; Jonathan R. Siegel, *The Hidden Source of Congress's Power To Abrogate State Sovereign Immunity*, 73 TEX. L. REV. 539 (1995).

The *qui tam* device came into use in English law long before the founding of the United States. See *Vt. Agency of Natural Res. v. United States ex rel. Stevens*, 529 U.S. 765, 774-75 (2000). The first Congress under the Constitution adopted the device, and since then various American statutes have employed it; the most prominent current example is the False Claims Act. See *id.* at 768 & n.1, 776.

Some have proposed the extension of the *qui tam* mechanism to provide for enforcement of federal regulatory requirements in areas such as medical product regulation or environmental law. See, e.g., Pamela H. Bucy, *Private Justice and the Constitution*, 69 TENN. L. REV. 939, 940 (2002) (arguing that "the [False Claims Act's] private justice model should be expanded to two areas: protection of financial markets and protection of the environment"); McGarity, *supra* note 44, at 580 (suggesting that a "statute, modeled on the [False Claims Act], creating a federal private right of action for damages caused by wrongful manipulation of a licensing regime administered by a federal agency" could help to address "situations in which companies make false claims to a federal agency about the safety and efficacy of their regulated products"). However, such discussions have not proposed a hybrid adjudicatory scheme such as the one outlined here.

153. See Siegel, *Suits Against States*, *supra* note 152, at 69. As Siegel notes, "[a] statute authorizing the federal government to espouse private claims . . . may give the government the right to collect any sums that the defendant would have had to pay in a suit brought by the injured private party." *Id.* (noting as an example that the Fair Labor Standards Act "empowers the Secretary of Labor to bring suit against any employer who has violated the Act and to distribute any sums recovered to affected employees").

154. "[A] State has a quasi-sovereign interest in the health and well-being—both physical and economic—of its residents in general." *Alfred L. Snapp & Son, Inc. v. Puerto Rico ex rel. Barez*,

regulated products clearly implicate the federal government's interest in consumer welfare.¹⁵⁵

The *parens patriae* model is often thought to be particularly appropriate for harms that affect a substantial portion of the population. This is likely to be true of many medical products liability claims. Especially in light of the speed with which new medical products spread through the market, a safety problem with such a product is likely to create a large number of claimants. In particular, claims concerning pharmaceuticals will ordinarily involve large numbers of potential claimants, because harms to only a handful of people will not be amenable to proof. (A manufacturing defect might cause isolated injuries; but manufacturing defects are unusual in the field of pharmaceuticals, if not in the area of devices.)¹⁵⁶

Moreover, though most *parens patriae* actions allege harm to large numbers of citizens, arguably the real touchstone should be, not the number of persons already harmed, but the degree of government interest in regulating the challenged conduct. As the Court explained with respect to a *parens patriae* action by Puerto Rico, one factor "in determining whether an alleged injury to the health and welfare of its citizens suffices to give the State standing to sue as *parens patriae* is whether the injury is one that the State, if it could, would likely attempt to address through its sovereign lawmaking powers."¹⁵⁷ The FDA has a clear interest in addressing product safety problems before those problems harm large numbers of people. Allowing *parens patriae* claims of the type posited here would further that mission, even if the group of people who have so far suffered harm is a small one.

Parens patriae actions, then, could usefully enforce medical product safety standards and obtain damages for harm to consumers. However, the United States' litigation resources are limited, and as discussed above, the government will not always discern safety problems quickly. To address these issues,

458 U.S. 592, 607 (1982); see also Larry W. Yackle, *A Worthy Champion for Fourteenth Amendment Rights: The United States in Parens Patriae*, 92 NW. U. L. REV. 111, 142 (1997) (discussing *Snapp*).

155. Cf. Nagareda, *supra* note 15, at 327-28 (noting that mass torts are not "purely private disputes" and that they "more closely resemble the issues of broad public concern that constitute the daily business of the administrative state").

156. See RISK MANAGEMENT REPORT, *supra* note 23, at 8 ("Injury from product defects is unusual in the United States because of the great attention paid to product quality control and quality assurance during manufacturing.").

157. *Alfred L. Snapp & Son*, 458 U.S. at 607 (noting that "[a]lthough more must be alleged than injury to an identifiable group of individual residents, the indirect effects of the injury must be considered as well in determining whether the State has alleged injury to a sufficiently substantial segment of its population").

Congress could authorize private persons to bring the *parens patriae* suit on behalf of the United States.

This proposal is modeled loosely on the *qui tam* provisions in the False Claims Act,¹⁵⁸ although the specific features of the *qui tam* suit proposed here would be set by the statute authorizing the new system. The *qui tam* mechanism permits a private person (the “relator”) to sue to recover damages for harm to the United States; in return, a successful *qui tam* relator receives a cut of the damages recovered. The primary justifications for the use of the *qui tam* mechanism under the False Claims Act apply in the present context as well. In at least some instances, the best evidence concerning an unsafe product will be known only to company insiders; but those insiders often will not come forward without a monetary incentive.¹⁵⁹ As noted, the federal government’s limited resources prevent it from pursuing all potentially valid claims.¹⁶⁰ Moreover, in some instances an agency might fail to pursue a claim because of undue influence from the regulated industry.¹⁶¹ Allowing private litigants to press claims on behalf of the government could address these concerns.

Congress could authorize the assertion of *qui tam* claims when a medical product had harmed consumers. An injured consumer could bring the suit; so could a person—such as a company insider—who possesses significant nonpublic information that supports the claim.¹⁶² At the outset of the suit, the government would have an opportunity to review the relevant information and

158. That Act imposes treble damages and civil penalties upon persons who submit false monetary claims to the federal government. See 31 U.S.C. § 3729(a) (2000). Suits under the Act can be brought either by the Attorney General or by a private person “in the name of the Government.” *Id.* § 3730(a)-(b). The Act’s growing use against asserted health care fraud has engendered controversy. See, e.g., Joan H. Krause, “Promises To Keep”: *Health Care Providers and the Civil False Claims Act*, 23 CARDOZO L. REV. 1363 (2002).

159. Cf. Evan Caminker, *The Constitutionality of Qui Tam Actions*, 99 YALE L.J. 341, 350 (1989) (noting, with respect to the False Claims Act, that “detecting fraud against the Federal treasury often is extremely difficult for the government without the aid of ‘informers,’” in part “because often the only persons who know about frauds are associated with the perpetrators . . . and are therefore reluctant to notify the authorities”).

160. Cf. *id.* at 350-51 (“[G]iven the ‘harsh reality of today’s funding limitations of . . . the budgets of the government’s prosecuting agencies,’ public officials often cannot commit the time and resources necessary for the successful prosecution of fraud even when they have already somehow managed to detect it.”).

161. Cf. *id.* at 351 (noting with respect to the False Claims Act that “[g]overnment agencies may be sufficiently dependent upon (or co-opted by) specific players in the military-industrial complex that the desire to prosecute wrongdoers diligently is compromised”).

162. There would be no Article III standing bar to such a claim. See *infra* notes 269-270 and accompanying text.

decide whether to pursue the action on its own behalf, either through litigation or an administrative proceeding.¹⁶³ As discussed below, this early review not only would provide the government with early input on the suit, it also would alert the FDA to the possible existence of a safety problem. Even if the government decided not to press the civil claim, this early warning could spur other investigative action within the FDA.

If the government did not take over the case, it would still retain some supervisory control. It could require the relator to provide it with copies of pleadings and with relevant information gained through discovery.¹⁶⁴ If it changed its decision later in the litigation, it could seek to intervene at that point.¹⁶⁵ The government could obtain dismissal of the suit over the *qui tam* relator's objection by establishing good cause for the dismissal (as in the case of a demonstrably frivolous claim).

This distribution of power over the prosecution of the suit would balance two competing concerns. On one hand, the value of the *qui tam* system comes from the opportunity for a private party to press a *qui tam* claim despite government inaction. Such inaction may sometimes arise from an agency's unwillingness to press a claim that would reveal evidence of prior agency errors or that would disadvantage an influential company. In the light of these concerns, it would be desirable to place some constraints on the government's ability to secure dismissal of the suit over the relator's objection. On the other hand, it is possible that if the government's ability to obtain dismissal of the suit were too constrained, courts might find that the mechanism offended separation of powers principles.¹⁶⁶ If a requirement that the government show good cause for the

163. The False Claims Act requires the relator to provide the United States with "[a] copy of the complaint and written disclosure of substantially all material evidence and information the person possesses." 31 U.S.C. § 3730(b)(2) (2000). The United States then has at least sixty days to decide whether to take over the action. *See id.* § 3730(b)(2)-(3). The government can press the claim either in the civil suit, *see id.* § 3730(b)(4)(A), or in an administrative proceeding, *see id.* § 3730(c)(5).

164. *Cf. id.* § 3730(c)(3) ("If the Government so requests, it shall be served with copies of all pleadings filed in the action and shall be supplied with copies of all deposition transcripts (at the Government's expense).").

165. *Cf. id.* ("When a person proceeds with the action, the court, without limiting the status and rights of the person initiating the action, may nevertheless permit the Government to intervene at a later date upon a showing of good cause.").

166. *See infra* notes 275-279 and accompanying text. Constraints on the government's ability to control the suit might raise separation of powers questions, but should not raise any other constitutional problems. As discussed below, the standing of a *qui tam* relator to press the claim is supported by the notion that the government has assigned a part of its injury to the *qui tam* relator. *See infra* notes 269-72 and accompanying text. The strength of that rationale would not vary depending on the degree of government control of the suit, because the degree of government

dismissal were deemed to impinge improperly on the executive branch's authority, the standard could be changed to permit dismissal at the government's behest for any rational governmental reason.¹⁶⁷

A successful relator would be paid a share of any damages recovery or settlement (but the relator would receive nothing if the defendant prevailed). The relator's share would vary depending on the degree of the relator's participation and the extent to which information provided by the relator played a role in the recovery.¹⁶⁸ The statute could also require a losing defendant to pay the relator's reasonable expenses, including attorney's fees.¹⁶⁹ The terms of judgments and

control would not alter the existence of a cognizable injury. Indeed, if anything, a *qui tam* relator's eagerness to press a claim (despite government skepticism) indicates that the relator is the sort of litigant who will litigate the claim zealously, which would subserve the presentation of the merits. Nor would it be persuasive to suggest that the degree of government control over the suit should affect the question of whether there is a Seventh Amendment right to a jury trial on the claim; the absence of government control over the suit would not create a right to a jury trial where none would otherwise exist.

167. The latter standard would parallel court interpretations of the False Claims Act, under which the government can "cause the action to be dismissed for any rational governmental reason, notwithstanding the *qui tam* plaintiff's desire that it continue." *United States ex rel. Stevens v. Vt. Agency of Natural Res.*, 162 F.3d 195, 202-03 (2d Cir. 1998), *rev'd on other grounds*, 529 U.S. 765 (2000).

168. In False Claims Act cases taken over by the government, the relator can only receive up to ten percent of the proceeds if the action was "based primarily" on information that was in the public record, but otherwise receives from fifteen to twenty-five percent, "depending upon the extent to which the person substantially contributed to the prosecution of the action." 31 U.S.C. § 3730(d)(1) (2000). In cases that the government decides not to take over, the relator receives from twenty-five to thirty percent of the proceeds. *See id.* § 3730(d)(2).

Some provision would need to be made for cases in which the relator had worked for the defendant. *Cf. id.* § 3730(d)(3) ("[I]f [the relator] planned and initiated the violation . . . the court may, to the extent the court considers appropriate, reduce the [relator's] share of the proceeds . . . , taking into account the role of that person in advancing the case to litigation and any relevant circumstances pertaining to the violation"). On the one hand, current and former employees may have key information concerning product safety, and the scheme should provide an incentive to bring that information forward. *Cf., e.g., United States ex rel. Franklin v. Parke-Davis*, No. Civ. A.96-11651-PBS, 2003 WL 22048255, at *1 (D. Mass. Aug. 22, 2003) (False Claims Act case in which relator alleged that his former employer "promoted the drug Neurontin for uses not approved by the Food and Drug Administration"). On the other, it would be unseemly to award a substantial portion of the damages to a person who had been responsible for the safety problem in the first place.

If the relator were one of those injured by the product, she would also receive a share of the damages distributed to injured claimants. *See infra* notes 176-178 and accompanying text.

169. *Cf.* 31 U.S.C. § 3730(d)(1)-(2) (2000). The False Claims Act provides that if the government does not take over the case and if the defendant wins, the court "may" require the

settlements in *qui tam* actions would be a matter of public record.¹⁷⁰

Under the system sketched here, there might sometimes be competition among would-be relators and their counsel. The False Claims Act's *qui tam* system accords relator status to the first person or persons to file a particular *qui tam* claim, and excludes later *qui tam* suits concerning the same facts.¹⁷¹ The first-to-file rule may make sense in the area of false claims, where the value added by the relator may lie primarily in the initial disclosure of the fraud.¹⁷² But when the *qui tam* suit will settle the question of a product's safety—and the discovery and trial process may call for significant expertise—some safeguards should be imposed to ensure that the lawyers litigating the claim are experienced and competent.

Thus, a modified first-to-file rule could be employed: If more than one *qui tam* suit concerning the same facts is filed within a short time period, the actions could be consolidated in one district court, and the court could select an appropriate relator (or set of joint relators) and suitable counsel.¹⁷³ In evaluating

relator to pay the defendant's reasonable attorney fees, if the action was "clearly frivolous, clearly vexatious, or brought primarily for purposes of harassment." *Id.* § 3730(d)(4).

170. As discussed below, defendants may seek confidentiality when settling private lawsuits. *See infra* notes 351-353 and accompanying text. Secret settlements of *parens patriae* claims, however, would be inappropriate: In suits on behalf of the government, there is a legitimate public interest in the terms of a settlement. Accordingly, the opt-in system would require that the terms of settlements be public. For an explanation of the opt-in mechanism, see Section III.A below.

Such a requirement would not necessarily present a significant downside for defendants. The main reason why secret settlements appeal to defendants is that secrecy deprives other potential plaintiffs of useful information. Within the opt-in system, however, the likelihood of follow-on claims would be significantly reduced, because the claims of all existing claimants would be resolved in the settlement. Although publication of the terms of a settlement might generate adverse publicity and affect sales, it would not have a direct impact on the liability of a company within the opt-in system.

171. *See* 31 U.S.C. § 3730(b)(5) (2000) ("When a person brings an action under this subsection, no person other than the Government may intervene or bring a related action based on the facts underlying the pending action."); *see also* JOHN E. CLARK, ETHICS ISSUES IN *QUI TAM* LITIGATION: SOME THOUGHTS FROM THE PERSPECTIVE OF A RELATOR'S COUNSEL, (A.B.A. Ctr. for Continuing Legal Educ., Nat'l Inst., N02CFCB ABA-LGLED I-1, 2001) ("It is not unusual for two or more individuals who share knowledge about a prospective defendant's actions—typically because they are co-workers—to join forces and seek to pursue a *qui tam* action jointly.").

172. *See United States ex rel. LaCorte v. Smithkline Beecham Clinical Labs., Inc.*, 149 F.3d 227, 234 (3d Cir. 1998) ("[C]laimants alleging the same material facts as prior relators should not share in a *qui tam* award, because their allegations are unlikely to increase the total recovery.").

173. In some instances, would-be relators might include both an insider with information about the product and one or more persons injured by the product. The court would then face the task of selecting relators from among those persons. According relator status to a company whistleblower

proposed counsel for the relator, the court should consider, among other things, any work the lawyer has done to develop the potential claim, the lawyer's experience in relevant areas of product liability litigation, and the resources the lawyer can bring to the representation.¹⁷⁴ The selection of a relator and of the relator's attorney could be decoupled if necessary: For example, an industry insider who brought significant nonpublic information to the table, but who was represented by inexperienced counsel, could be directed to seek more experienced representation in order to be allowed to proceed as the *qui tam* relator.¹⁷⁵

If, upon referral, the FDA found a safety problem, the district court would

would serve the goal of encouraging those with relevant information to come forward; on the other hand, according relator status to one or more injured claimants would help to ensure that the claimants' perspective is presented in the litigation. In at least some cases, the optimal choice would be to appoint multiple persons to act jointly as relators.

174. When appointing counsel in a federal class action:

[T]he court . . . must consider:

- the work counsel has done in identifying or investigating potential claims in the action,
- counsel's experience in handling class actions, other complex litigation, and claims of the type asserted in the action,
- counsel's knowledge of the applicable law, and
- the resources counsel will commit to representing the class.

FED. R. CIV. P. 23(g)(1)(C). The court may also consider other relevant factors. *See id.*

175. The Private Securities Litigation Reform Act of 1995 (PSLRA) provides a precedent for such "decoupling." In federal securities fraud class actions, the PSLRA directs the court to "appoint as lead plaintiff the member or members of the purported plaintiff class that the court determines to be most capable of adequately representing the interests of class members (. . . the 'most adequate plaintiff')." 15 U.S.C. § 78u-4(a)(3)(B)(i) (2000). In turn, "[t]he most adequate plaintiff shall, subject to the approval of the court, select and retain counsel to represent the class." *Id.* § 78u-4(a)(3)(B)(v). This provision, by "permit[ting] the plaintiff to choose counsel rather than have counsel choose the plaintiff," S. REP. NO. 104-98, at 11 (1995), *reprinted in* 1995 U.S.C.C.A.N. 679, was designed to lessen the influence wielded by plaintiffs' class action lawyers. And while the provision puts the initial choice of counsel in the hands of the "most adequate plaintiff"—an entity that will frequently turn out to be a large institutional investor—the statute preserves authority in the court "to approve or disapprove the lead plaintiff's choice of counsel when necessary to protect the interests of the plaintiff class." *Id.* at 12. The PSLRA has generated debate over the extent to which the court should override the lead plaintiff's preference concerning counsel. *See, e.g.*, Third Circuit Task Force Report on the Selection of Class Counsel, 208 F.R.D. 340, 345 (2002) ("The Act raises a number of questions, including the degree to which a court should defer to the lead plaintiff's choice of counsel and whether a court-sponsored auction is permissible in securities class actions."). But it seems clear that in at least some cases the PSLRA will decouple the choice of plaintiff from the choice of counsel.

proceed to apply the FDA's guidance on causation in order to determine or estimate the number of persons injured by the product. The amount of compensatory damages would depend on the number, type, and severity of injuries. Often, damages determinations could be made on the basis of individualized evidence; however, in the case of a product that harmed huge numbers of people, the court might use statistical methods to set damages amounts.¹⁷⁶ In determining appropriate damages, the court would also take into account factors relating to the defendant's culpability, including an assessment of the time when the safety issue first became known or knowable, and whether the company was proactive in discerning and addressing the issue.¹⁷⁷

After determining damages, the court would enter judgment. The proceeds of the judgment would go into a compensation fund, which would be distributed by a special master to claimants based upon their exposure and injury.¹⁷⁸

Because the *parens patriae* suit would assert the government's interest in obtaining redress on behalf of all those currently injured by a product, the judgment would determine the question of the company's liability with respect to current injuries. The judgment's finality, however, would be subject to two major limitations.

One limitation concerns "exposure-only" claimants—those persons who have used a product, but who have not yet shown signs of injury. The court's assessment of damages might include a component designed to cover the cost of compensation for claimants whose latent injuries only manifest themselves after judgment. However, if it turned out that the class of persons with latent injuries was larger than the court had anticipated, the government should be able to reopen the judgment to seek additional compensatory relief.

The other limitation concerns cases in which the FDA determines either that the product is safe or that causation is absent. Such a determination will result in

176. A number of commentators have discussed the possibility of determining damages on an aggregate basis in mass tort cases. See, e.g., Michael J. Saks & Peter David Blanck, *Justice Improved: The Unrecognized Benefits of Aggregation and Sampling in the Trial of Mass Torts*, 44 STAN. L. REV. 815 (1992); Richard A. Nagareda, *Turning from Tort to Administration*, 94 MICH. L. REV. 899, 917 (1996).

177. Subsection III.A.2 discusses in more detail the factors relevant to the damages determination. See *infra* text accompanying notes 323-333.

178. The details of the fund's administration would depend on a number of factors. In cases where the district court's damages calculations were based on individualized assessments of damages, the fund administrator would apply those individualized assessments in distributing payments to claimants. In cases where aggregate damages calculations were employed by the district court, the fund administrator would need to require some showing from each claimant concerning exposure to the product and degree of injury; the administrator could then employ a schedule or matrix to set the award for each claimant.

a judgment in favor of the defendant, and ordinarily that should be the end of the matter. However, in some instances, advances in science and research may uncover evidence concerning safety and causation that was unavailable at the time of the initial *qui tam* action, and such evidence may provide the quantum of proof that was lacking during the first proceeding. In those instances, the government (or, in appropriate circumstances, a *qui tam* relator) should be able to seek to reopen the judgment and relitigate the question of safety and causation. However, the standard for reopening the judgment would have to be fairly demanding—both for practical reasons and because of constitutional concerns.¹⁷⁹

C. Comparing the Options

In this Section, I will compare the options described above along various dimensions. Constitutional constraints impose some limits on the range of structural options among which policymakers may choose. Thus, I first consider whether each of the options detailed above is constitutionally permissible. It seems likely that the system of hybrid adjudication for private tort claims would face a Seventh Amendment barrier, but that the other three options could comport with the Constitution. I next compare the remaining three options—government enforcement within an agency setting, private intra-agency enforcement, and *qui tam* claims—by considering their likely effects on the cost and speed of litigation, the skill and zeal with which claims would be presented, and the expertise and neutrality of the decisionmaker. I argue that *qui tam* enforcement may be the most desirable, because it harnesses the skills of the private bar, and it provides the protections of an independent, generalist judicial decisionmaker.

179. David Shapiro has noted:

[T]he need to recognize the finality of judgments—their immunity from reopening or nullification at the hands of the executive or legislature (as well as the oft-repeated canon that the courts do not sit to render “advisory opinions”) is fundamental to the status of the federal courts under Article III of the Constitution

DAVID L. SHAPIRO, *CIVIL PROCEDURE: PRECLUSION IN CIVIL ACTIONS* 14 (2001). The contours of this constraint are uncertain because “[t]he Supreme Court has seldom had to consider how much res judicata effect is necessary.” RICHARD H. FALLON, JR. ET AL., *HART AND WECHSLER’S THE FEDERAL COURTS AND THE FEDERAL SYSTEM* 105 (5th ed. 2003). It seems clear, however, that some latitude to reopen judgments is permissible. In civil actions, a federal court has discretion to grant relief from a judgment on the ground of “newly discovered evidence which by due diligence could not have been discovered in time to move for a new trial,” if the relief is sought within a year after entry of judgment. FED. R. CIV. P. 60(b)(2). Although the system described in the text would be more lenient than Rule 60(b)(2) in at least some respects—for example, it would not include the one-year time limitation—it could presumably be designed so that the judgment in the initial suit would have enough finality to comport with Article III.

1. Constitutional Constraints

To the modern eye, regulation and litigation overlap. It is thus tempting to consider the various configurations—administrative adjudication, court adjudication, hybrid adjudication—from a purely functional perspective. The choice of structure, however, can have constitutional implications as well. In this Subsection, I review constitutional issues posed by each of the four schemes sketched above.¹⁸⁰ The internal-agency-enforcement model is clearly constitutional. The private-enforcement/agency-adjudication model may be constitutional as well. *Qui tam* claims in the hybrid system also pass constitutional scrutiny. Private claims in the hybrid system are questionable, however, because plaintiffs would likely have a Seventh Amendment right to a jury, and the referral of safety and causation issues to the FDA would likely violate that right.

It is useful, at the outset, to review the concerns that underlie the requirements set by Article III and the Seventh Amendment. Article III serves structural values: The requirement that many types of disputes be adjudicated by life-tenured, salary-protected judges maintains the function of the Article III courts and prevents the other two branches from aggrandizing themselves at the expense of the judiciary.¹⁸¹ In addition, Article III protects the litigant's right to a fair, independent tribunal.¹⁸² The Seventh Amendment protects individual rights by ensuring that disputes within its scope can be heard by juries, which can provide an independent check on government decisionmaking. The Court has

180. The schemes discussed here would operate prospectively: Companies could choose to opt in when submitting new products for FDA review, *see infra* Section III.A. Thus, the proposals considered in this Subsection would not apply to products already on the market, and thus would not affect any vested legal rights. The preemption of potential state tort claims therefore would not violate due process. *See* Olivia A. Radin, Note, *Rights as Property*, 104 COLUM. L. REV. 1315, 1328 (2004) ("The Supreme Court [has] found that laws that affect future actions do not implicate a property interest.").

181. As the Court has explained:

Article III, § 1, safeguards the role of the Judicial Branch in our tripartite system by barring congressional attempts 'to transfer jurisdiction [to non-Article III tribunals] for the purpose of emasculating' constitutional courts, *National Ins. Co. v. Tidewater Co.*, 337 U.S. 582, 644 . . . (1949) (Vinson, C.J., dissenting), and thereby preventing 'the encroachment or aggrandizement of one branch at the expense of the other.' *Buckley v. Valeo*, 424 U.S. 1, 122 . . . (1976) (per curiam).

Commodity Futures Trading Comm'n v. Schor, 478 U.S. 833, 850 (1986).

182. *See id.* at 848 ("Article III, § 1's guarantee of an independent and impartial adjudication by the federal judiciary of matters within the judicial power of the United States . . . serves to protect primarily personal, rather than structural, interests.").

been protective of Seventh Amendment rights within the court system,¹⁸³ but has permitted Congress some latitude to render that Amendment inapplicable by assigning disputes to agencies instead of courts.¹⁸⁴

a) Internal Agency Enforcement

The internal-agency-enforcement scheme is standard fare in the administrative state. "Congress has often created new statutory obligations, provided for civil penalties for their violation, and committed exclusively to an administrative agency the function of deciding whether a violation has in fact occurred."¹⁸⁵ Such an arrangement comports with Article III because government enforcement of civil penalties for violation of an administrative scheme falls within the traditional core of "public rights" cases that can be committed to non-Article III tribunals¹⁸⁶ (or, by extension, to administrative agencies with limited review in Article III courts¹⁸⁷). And though a civil penalty defendant would have a right to a jury if the action took place in an Article III court,¹⁸⁸ no such right attaches when the penalty proceeding unfolds within an administrative agency.¹⁸⁹

183. See *infra* note 240.

184. See *infra* note 239.

185. *Atlas Roofing Co. v. Occupational Safety & Health Review Comm'n*, 430 U.S. 442, 450 (1977).

186. See *N. Pipeline Constr. Co. v. Marathon Pipe Line Co.*, 458 U.S. 50, 64-69 (1982) (plurality opinion) (describing the requirement of Article III adjudication, and listing exceptions concerning territorial courts, courts-martial, and cases involving "public rights"). Although the Court has since held that the "public rights" category includes some disputes to which the government is not a party, see *Thomas v. Union Carbide Agric. Prods. Co.*, 473 U.S. 568, 586 (1985) ("Insofar as appellees interpret [prior cases] as establishing that the right to an Article III forum is absolute unless the Federal Government is a party of record, we cannot agree."), cases brought by or against the government continue to fall within the core of the "public rights" doctrine.

187. See *N. Pipeline*, 458 U.S. at 67 n.18 (plurality opinion) ("Congress' power to create legislative courts to adjudicate public rights carries with it the lesser power to create administrative agencies for the same purpose, and to provide for review of those agency decisions in Art. III courts.").

188. See *Tull v. United States*, 481 U.S. 412, 420 (1987) (stating that an action for civil penalties under Clean Water Act "is clearly analogous to the 18th-century action in debt, and federal courts have rightly assumed that the Seventh Amendment required a jury trial").

189. As the *Atlas Roofing* Court explained:

[W]hen Congress creates new statutory 'public rights,' it may assign their adjudication to an administrative agency with which a jury trial would be incompatible, without violating the Seventh Amendment's injunction that jury trial is to be 'preserved' in 'suits at common law.' . . . This is the case even if the Seventh Amendment would have

b) Private Intra-Agency Enforcement

The constitutionality of the private intra-agency enforcement proceeding would depend on whether the rationales described above, with respect to government enforcement, could extend to private claims in the context of the FDA regulatory scheme. Taken together, two cases—*Thomas v. Union Carbide Agricultural Products Co.*¹⁹⁰ and *NLRB v. Jones & Laughlin Steel Corp.*¹⁹¹—suggest that the scheme could be permissible. But because both of these cases are distinguishable from the FDA products liability proposal, the proposal's constitutionality is not entirely free from doubt.

In *Thomas* the Court held that the “public rights” doctrine extends to some disputes between private parties. *Thomas* concerned a pesticide maker's right to compensation when its data were used to facilitate regulatory approval (under the Federal Insecticide, Fungicide, and Rodenticide Act (“FIFRA”)) of a competitor's similar pesticide.¹⁹² Disagreements between pesticide makers over the appropriate amount of compensation were sent to binding arbitration, with very limited federal court review.¹⁹³ In rejecting a participant's Article III challenge to the scheme, the Court held that the pesticide maker's right to compensation was “not a purely ‘private’ right,” because it had “many of the characteristics of a ‘public’ right.”¹⁹⁴ In particular, the use of the pesticide data played “an integral part” in “a complex regulatory scheme” to protect public health.¹⁹⁵

Private intra-agency enforcement of claims for violation of FDA requirements would arguably fall within the *Thomas* Court's statement that “Congress, acting for a valid legislative purpose pursuant to its constitutional powers under Article I, may create a seemingly ‘private’ right that is so closely integrated into a public regulatory scheme as to be a matter appropriate for

required a jury where the adjudication of those rights is assigned instead to a federal court of law instead of an administrative agency.

Atlas Roofing, 430 U.S. at 455. For a critique of this distinction, see Ellen E. Sward, *Legislative Courts, Article III, and the Seventh Amendment*, 77 N.C. L. REV. 1037, 1140 (1999) (“To allow Congress to avoid the Seventh Amendment by assigning adjudication of certain matters otherwise meeting the constitutional test for the Seventh Amendment to non-Article III courts is tantamount to informally amending the Constitution to limit the reach of the Seventh Amendment.”).

190. 473 U.S. 568 (1985).

191. 301 U.S. 1 (1937).

192. See *Thomas*, 473 U.S. at 571-73.

193. See *id.* at 573-74 (“The arbitrator's decision is subject to judicial review only for ‘fraud, misrepresentation, or other misconduct.’”).

194. *Id.* at 589.

195. *Id.*

agency resolution with limited involvement by the Article III judiciary.”¹⁹⁶ As in *Thomas*, the private claims adjudication would play an integral role in the FDA’s regulatory scheme: Private claims would help to enforce FDA requirements, would provide information to assist the FDA in its regulatory role, and would result in compensation for those injured by safety problems.

However, some key differences could limit the application of *Thomas*’ holding to the scheme described here. The *Thomas* Court emphasized that those involved in FIFRA’s compensation scheme were “voluntary participants in the program.”¹⁹⁷ As explained below,¹⁹⁸ potential defendants in the FDA enforcement scheme could validly consent to the system in advance. However, the claimants in the private enforcement scheme would be people injured by medical products; those claimants would not have meaningfully consented to the use of the non-Article III system.¹⁹⁹

In *Thomas*, the Court also found it significant that the rights at issue were federal rights that did not “depend on or replace a right to . . . compensation under state law.”²⁰⁰ In the FDA enforcement scheme, although the claim would arise under federal law, it would substantially resemble (and supplant) state tort claims for products liability. To the extent that federal rights that displace similar state common law causes of action fall closer to the core of Article III concerns,²⁰¹ this difference might cut against extending *Thomas* to the FDA

196. *Id.* at 593-94.

197. *Id.* at 589; *see also id.* at 592 (“[U]nder FIFRA, the only potential object of judicial enforcement power is the follow-on registrant who explicitly consents to have his rights determined by arbitration.”).

198. *See infra* Section III.A (discussing opt-in mechanism); *see also infra* note 268.

199. Perhaps it could be argued that by using a medical product (labeled with an announcement of the administrative compensation scheme) one consents in advance to the use of the non-court, non-jury proceeding. However, there are serious questions as to the practicality and fairness of such a position. *Cf. Sage*, *supra* note 46, at 990 (noting, with respect to treatment decisions involving drugs, that “[m]any consumers are under physical and emotional burdens that may preclude true freedom of choice”).

200. *Thomas*, 473 U.S. at 584.

201. As the *Thomas* Court explained:

In assessing the degree of judicial involvement required by Article III in this case, we note that the statute considered in *Crowell* [v. Benson, 285 U.S. 22 (1932)] is different from FIFRA in significant respects. Most importantly, the statute in *Crowell* displaced a traditional cause of action and affected a pre-existing relationship based on a common-law contract for hire. Thus it clearly fell within the range of matters reserved to Article III courts under the holding of *Northern Pipeline*. *See* 458 U.S., at 70-71 n.25 (plurality opinion) (noting that matters subject to a “suit at common law or in equity or admiralty” are at “protected core” of Article III judicial powers); *id.*, at 90 (opinion concurring in judgment) (noting that state law contract actions are “the stuff of the traditional actions

enforcement scheme.

This is not to say that claims replacing state-law causes of action can never be assigned to agencies for adjudication. *CFTC v. Schor*²⁰² demonstrates that even state-law claims can be heard by federal agencies in some circumstances. *Schor* addressed a scheme whereby customers injured by a commodity broker's violation of federal law could seek reparations in a proceeding before the Commodity Futures Trading Commission ("CFTC").²⁰³ CFTC regulations permitted a broker to assert factually related counterclaims in the reparations proceeding.²⁰⁴ Although such counterclaims were state law claims "of the kind assumed to be at the 'core' of matters normally reserved to Article III courts,"²⁰⁵ the Court upheld the scheme.

The *Schor* Court relied upon two factors that could apply with equal strength to the FDA scheme. First, the Court emphasized that the CFTC's jurisdiction over state-law counterclaims was necessary to the success of the regulatory scheme.²⁰⁶ If, as I have argued, private enforcement is a necessary supplement to FDA regulation, then the claim of regulatory necessity could similarly support the permissibility of the FDA adjudicatory scheme. Second, the Court noted that the CFTC's jurisdiction was limited to a particular field—claims concerning violations of federal commodities laws, plus factually related claims—rather than extending to all sorts of state-law claims.²⁰⁷ Likewise, the FDA enforcement scheme would only concern claims regarding injury from certain FDA-regulated products.

However, *Schor*, even more than *Thomas*, turned upon the notion of consent.²⁰⁸ *Schor* stands for the proposition that the assignment of a private-rights

at common law tried by the courts at Westminster in 1789").

Thomas, 473 U.S. at 587.

202. 478 U.S. 833 (1986).

203. *See Schor*, 478 U.S. at 836.

204. *See id.* at 837.

205. *Id.* at 853.

206. *See id.* at 856 ("It was only to ensure the effectiveness of [the reparations] scheme that Congress authorized the CFTC to assert jurisdiction over common law counterclaims. Indeed . . . absent the CFTC's exercise of that authority, the purposes of the reparations procedure would have been confounded.").

207. *See id.* at 852-53 ("The CFTC . . . deals only with a 'particularized area of law,' . . . whereas the jurisdiction of the bankruptcy courts found unconstitutional in *Northern Pipeline* extended to broadly 'all civil proceedings arising under title 11 or arising in or related to cases under title 11.'" (citation omitted)).

208. As the Court emphasized:

Schor indisputably waived any right he may have possessed to the full trial of Conti's counterclaim before an Article III court. Schor expressly demanded that Conti proceed

dispute to a non-Article III tribunal need not offend *structural* Article III concerns, so long as *Schor*'s balancing test is met.²⁰⁹ But since the litigant in *Schor* had consented to submit the claim to the CFTC, the holding in *Schor* did not extend to cases in which no such waiver had occurred. In a subsequent case, *Granfinanciera, S.A. v. Nordberg*,²¹⁰ the Court held that in the absence of litigant consent, a private-rights claim that would carry a jury right if litigated in federal court is not assignable to a non-Article III tribunal for juryless adjudication.²¹¹ The *Granfinanciera* Court explicitly equated the scope of the Seventh Amendment constraint with that of the Article III constraint.²¹² Thus, though *Schor* indicates that private-rights disputes may be assigned to non-Article III tribunals when the litigants consent, *Granfinanciera* indicates that absent litigant consent, a case must fall within the public-rights category (or another traditional exception)²¹³ in order to be validly assigned to a non-Article III tribunal.²¹⁴

In that respect, *Jones & Laughlin Steel* may provide more support for the private intra-agency enforcement scheme, because it did not involve litigant consent. Assuming that *Jones & Laughlin Steel*'s holding concerning the appropriateness of agency adjudication is still good law—a fair assumption,

on its counterclaim in the reparations proceeding rather than before the District Court, and was content to have the entire dispute settled in the forum he had selected until the ALJ ruled against him on all counts; it was only after the ALJ rendered a decision to which he objected that Schor raised any challenge to the CFTC's consideration of Conti's counterclaim.

Id. at 849 (citation omitted).

209. *See id.* at 851 (explaining that factors to be weighed include the degree to which the “essential attributes of judicial power” are reserved to Article III courts, and, conversely, the extent to which the non-Article III forum exercises the range of jurisdiction and powers normally vested only in Article III courts”; “the origins and importance of the right to be adjudicated”; and Congress’s reasons for “depart[ing] from the requirements of Article III”).

210. 492 U.S. 33 (1989).

211. *See id.* at 53-54.

212. *See id.*

213. In addition to the public rights doctrine, traditional exceptions that justify assignment of a matter to a non-Article III tribunal include matters assigned to territorial courts and to courts martial. *See* *N. Pipeline Constr. Co. v. Marathon Pipe Line Co.*, 458 U.S. 50, 64-66 (1982).

214. *Granfinanciera* does tacitly suggest that a private rights claim that would carry a jury right if litigated in federal court could be sent to a non-Article III tribunal if that tribunal employed a jury and acted merely as an adjunct to an Article III court. The existence of such a possible exception explains why the Court in *Granfinanciera*, having determined that the claim at issue was a private rights claim, nonetheless left open the question whether bankruptcy judges (who are not Article III judges) could conduct a jury trial on the claim, subject to district court oversight. *See Granfinanciera*, 492 U.S. at 64. But the juryless adjudication of private claims within an agency setting obviously would not fit within that possible exception.

given that the Court has made no suggestion to the contrary—the case must now be read to rest upon the conclusion that the claim at issue fell within the public rights doctrine. Thus, the private intra-agency enforcement scheme could be validated as a public-rights scheme if it were considered sufficiently similar to the scheme at issue in *Jones & Laughlin Steel*.

In that case, a union instituted a proceeding against an employer before the NLRB seeking both injunctive remedies and back pay under federal law.²¹⁵ Among other objections, the employer asserted that it had a Seventh Amendment right to a jury trial on the back pay issue.²¹⁶ Though the Court rejected this contention partly because it viewed the back pay award as merely incidental to the injunctive relief,²¹⁷ the Court also suggested that juryless adjudication within the NLRB was appropriate because the claims at issue were created by Congress: “The instant case is not a suit at common law or in the nature of such a suit. The proceeding is one unknown to the common law. It is a statutory proceeding.”²¹⁸ Since consent was not a basis for the holding in *Jones & Laughlin Steel*, that case may support the constitutionality of the scheme outlined here. On the other hand, the case is not directly on point because it is somewhat difficult to argue that a claim for products liability is “unknown to the common law.”

In addition, *Thomas* and *Jones & Laughlin Steel* may also be distinguishable from the FDA scheme outlined here in that both cases involved decisionmakers relatively insulated from the executive branch: *Jones & Laughlin Steel* involved the NLRB, an independent agency, and in *Thomas*, the arbitrators were appointed by a separate, independent federal agency.²¹⁹ Although the FDA Commissioner must be confirmed by the Senate, he or she “serves at the pleasure of the . . . Secretary [of Health and Human Services] and, therefore, the President.”²²⁰ Thus,

215. See *NLRB v. Jones & Laughlin Steel Corp.*, 301 U.S. 1, 22 (1937). Because the union initiated the NLRB proceeding, the case can be viewed as one involving a dispute between private parties. It appears, however, that the actual disputants during the adjudication within the NLRB were the employer and the Board—not the union. See *id.* at 24-25 (describing proceedings before the NLRB).

216. See *id.* at 48.

217. See *id.* (holding that the Seventh Amendment “has no application to cases where recovery of money damages is an incident to equitable relief”).

218. *Id.*

219. See *Thomas v. Union Carbide Agric. Prods. Co.*, 473 U.S. 568, 574 n.1 (1985) (quoting statute that provided for appointment of arbitrator by Federal Mediation and Conciliation Service). *Schor* is similarly distinguishable: The Court noted that the CFTC, an independent agency, was viewed as “relatively immune from political pressures.” *Commodity Futures Trading Comm’n v. Schor*, 478 U.S. 833, 855 (1986).

220. John W. Lundquist & Sandra L. Conroy, *Defending Against Food and Drug Prosecutions*, THE CHAMPION, July 1997, at 20, 20.

the FDA lacks some of the attributes of an independent agency.²²¹

In sum, it seems possible that Congress could assign private claims under the FDCA to agency adjudication, but the answer is not entirely clear, because the boundaries of the “public rights” doctrine are incompletely defined.

c) Private Enforcement in a Hybrid System

The immediately preceding analysis demonstrates that if a private claim under the FDCA is deemed to fall within the “public rights” doctrine, the claim can be adjudicated within the FDA without offending either Article III or the Seventh Amendment. But if such a claim is brought, instead, in federal court, and if the right and remedy involved are legal in nature, then a Seventh Amendment jury right attaches and this right would prevent referral of safety and causation issues to the FDA for binding determination.

It should be noted that Article III itself would not pose a barrier to the private/hybrid scheme. Even if private claims under the FDCA did not fall within the “public rights” doctrine, Article III would pose no bar to the adjudication of those claims in federal court with referral of the safety and causation issues to the FDA. From the perspective of Article III analysis, the private/hybrid scheme conforms well to the “adjunct” model exemplified by *Crowell v. Benson*.²²²

In *Crowell*, the Court considered a workers’ compensation scheme devised by Congress as a substitute for traditional federal negligence law in admiralty jurisdiction.²²³ Under the scheme, claims for compensation were heard and determined by deputy commissioners within the United States Employees’ Compensation Commission.²²⁴ Enforcement of any resulting compensation order was to be sought in federal court.²²⁵ Though the court would review the administrator’s legal determinations de novo,²²⁶ the administrator’s factual determinations generally were reviewed only under a deferential “supported by

221. See Paul R. Verkuil, *The Purposes and Limits of Independent Agencies*, 1988 DUKE L.J. 257, 259 (noting that agency independence typically involves “three statutory arrangements: the bipartisan appointment requirement; the fixed term requirement; and the requirement that removal be limited to express causes”).

222. 285 U.S. 22 (1932).

223. See *id.* at 36-38 (explaining that federal Longshoremen’s and Harbor Workers Compensation Act “deals exclusively with compensation in respect of disability or death resulting ‘from an injury occurring upon the navigable waters of the United States’ if recovery ‘through workmen’s compensation proceedings may not validly be provided by State law’”).

224. See *id.* at 42-43 (describing hearing procedure).

225. See *id.* at 44.

226. See *id.* at 45 (“Rulings of the deputy commissioner upon questions of law are without finality.”).

evidence" standard.²²⁷ However, as the Court interpreted the statutory framework,²²⁸ it provided for independent judicial determination of facts relevant to the commission's jurisdiction²²⁹ or to constitutional rights.²³⁰

At the time, the Court viewed the "public rights" doctrine as extending only to cases in which the government was a party.²³¹ But though the dispute in *Crowell* was thus classified as a "private rights" case,²³² the Court upheld the statutory delegation of fact-finding to the agency: Even in private rights cases, the Court held, "there is no requirement that, in order to maintain the essential attributes of the judicial power, all determinations of fact in constitutional courts shall be made by judges."²³³ It sufficed, in *Crowell*, that independent federal court review was available with respect to jurisdictional and constitutional facts.²³⁴

The private/hybrid scheme fits comfortably within *Crowell*'s holding with respect to the requirements of Article III. As in *Crowell*, a federal statutory claim would replace a judicially developed cause of action. The field covered by the statute would be limited to medical products regulated by the FDA.²³⁵ The referral of technical questions to the FDA would "furnish a prompt, continuous, expert, and inexpensive method for dealing with a class of questions of fact which are peculiarly suited to examination and determination by an administrative agency specially assigned to that task."²³⁶ And the FDA's decisions would be subject to federal court review for compliance with the statutory scheme, though the FDA's judgments on safety and causation would

227. *Id.* at 46.

228. The Court adopted the interpretation noted in the text in order to avoid the constitutional issues that would have arisen had it found that the statute required judicial deference to the commissioner with respect to jurisdictional and constitutional facts. *See id.* at 62 ("When the validity of an act of the Congress is drawn in question, and even if a serious doubt of constitutionality is raised, . . . this Court will first ascertain whether a construction of the statute is fairly possible by which the question may be avoided.").

229. *See id.* at 63.

230. *See id.* at 60.

231. *See id.* at 50 ("[T]he distinction is at once apparent between cases of private right and those which arise between the government and persons subject to its authority in connection with the performance of the constitutional functions of the executive or legislative departments.").

232. *See id.* at 51 ("The present case does not fall within the categories just described, but is one of private right, that is, of the liability of one individual to another under the law as defined.").

233. *Id.*

234. *See id.* at 62.

235. *Cf. id.* at 54 ("The statute has a limited application, being confined to the relation of master and servant . . .").

236. *Id.* at 46.

receive deference.

Crowell, however, does not settle the Seventh Amendment question: Because *Crowell* concerned a statutory replacement for a claim in admiralty—not a claim at common law—the Seventh Amendment was not at issue in the case.²³⁷ The Seventh Amendment jury trial requirement applies to a claim under a federal statute if the right and remedy are legal in nature.²³⁸ Admittedly, if such a claim falls within the “public rights” doctrine it can be relegated to administrative adjudication and the Seventh Amendment jury right will not apply in the administrative proceeding.²³⁹ But if, instead, such a claim is brought in federal court, the Seventh Amendment requires a jury.²⁴⁰

Under this analysis, situating the federal products liability claim in federal court would trigger a Seventh Amendment right to jury trial. The claim would be analogous to a state-law tort claim for products liability and the remedies sought—compensatory and punitive damages—would fall within the category of legal relief.²⁴¹ Accordingly, the binding determination of safety and causation issues by the FDA would be impermissible, because it would violate the right to a jury trial.²⁴² Of course, the jury trial right is waivable; but though the defendant could validly waive the right in advance, the claimant would not have done so.²⁴³

237. *See id.* at 45 (“As the claims which are subject to the provisions of the Act are governed by [federal] maritime law . . . and are within the admiralty jurisdiction, the objection raised by the respondent’s pleading as to the right to a trial by jury under the Seventh Amendment is unavailing.”).

238. *See Curtis v. Loether*, 415 U.S. 189, 194 (1974) (“The Seventh Amendment does apply to actions enforcing statutory rights, and requires a jury trial upon demand, if the statute creates legal rights and remedies, enforceable in an action for damages in the ordinary courts of law.”).

239. *See Granfinanciera, S.A. v. Nordberg*, 492 U.S. 33, 42 n.4 (1989) (“If a claim that is legal in nature asserts a ‘public right,’ . . . then the Seventh Amendment does not entitle the parties to a jury trial if Congress assigns its adjudication to an administrative agency or specialized court of equity.”).

240. As the Court has explained, when a federal scheme contemplates “enforcement of statutory rights in an ordinary civil action in the district courts, where there is obviously no functional justification for denying the jury trial right, a jury trial must be available if the action involves rights and remedies of the sort typically enforced in an action at law.” *Curtis*, 415 U.S. at 195.

241. It would be possible to design a claim that sought solely equitable relief. For example, the statute could provide for restitution of money the company had derived from sales of a defective product. However, such a remedy would often not meet the goal of compensation, because the harm done by a defective product may exceed the profits a company derived from it. And such a remedy would not serve the purposes furthered by punitive damages, either.

242. *See Ex parte Peterson*, 253 U.S. 300, 314 (1920) (“A compulsory reference with power to determine issues is impossible in the federal courts because of the Seventh Amendment . . .”).

243. The waiver analysis here parallels that discussed in the previous Section with respect to individual Article III rights. *See supra* text accompanying notes 198-199.

Conceptualizing the referral scheme as an application of the primary jurisdiction doctrine would not remove the Seventh Amendment difficulty. The Supreme Court has not explicitly addressed Seventh Amendment constraints on the application of primary jurisdiction. The Court's silence is perhaps unsurprising, because many of the cases that have presented issues of primary jurisdiction involved no Seventh Amendment right—for example, because the lawsuit in question was brought in state rather than federal court,²⁴⁴ or sought equitable rather than legal relief.²⁴⁵ In other cases, a litigant who might have had a Seventh Amendment right failed to argue that claim to the Supreme Court as a bar to the application of the primary jurisdiction doctrine.²⁴⁶

The two Supreme Court primary jurisdiction cases that most directly presented a Seventh Amendment issue are inapposite to the question considered in this Article. In *Keogh v. Chicago & N.W. Ry. Co.*,²⁴⁷ the plaintiff argued to the Court that he had a constitutional right to a jury trial on his claims for antitrust damages.²⁴⁸ However, the court below had dismissed Keogh's antitrust claims because by the time those claims reached trial, the Interstate Commerce Commission had approved the rates challenged by the plaintiff.²⁴⁹ Given this procedural history, the Supreme Court's holding that Keogh's action was

244. See, e.g., *San Diego Bldg. Trades Council v. Garmon*, 359 U.S. 236, 239 (1959); *Tex. & Pac. Ry. v. Abilene Cotton Oil Co.*, 204 U.S. 426, 426 (1907).

245. See, e.g., *Bd. of R.R. Comm'rs v. Great N. Ry.*, 281 U.S. 412, 416 (1930) (suit to enjoin enforcement of state agency's rate-setting order); *U.S. Navigation Co. v. Cunard S.S. Co.*, 284 U.S. 474, 478 (1932) (suit to enjoin alleged antitrust violation); *Far E. Conference v. United States*, 342 U.S. 570, 571, 573 (1952) (suit by the United States to enjoin alleged antitrust violations).

246. See, e.g., *Ricci v. Chi. Mercantile Exch.*, 409 U.S. 289 (1973) (claims for damages under Commodities Exchange Act and Sherman Act; in Supreme Court briefs, petitioner did not assert a right to jury trial as a basis for reversal); *Chi. Mercantile Exch. v. Deaktor*, 414 U.S. 113, 113-14 (1973) (per curiam) (deciding case for claims for damages under Commodities Exchange Act and Sherman Act without a merits briefing; certiorari briefs did not mention right to jury trial); *Andrews v. Louisville & Nashville R.R.*, 406 U.S. 320, 320-21, 324-25 (1972) (damages suit that had been removed from state to federal court; majority refused to address "[t]he constitutional issue discussed in the dissent" because the issue was not included in the petition for certiorari); *id.* at 331 (Douglas, J., dissenting) ("Under the First Amendment . . . [the plaintiff] is petitioning the Government 'for a redress of grievances' in the traditional manner of suitors at common law; and by the Seventh Amendment is entitled to a jury trial."); *Port of Boston Marine Terminal Ass'n v. Rederiaktiebolaget Transatlantic*, 400 U.S. 62, 64-65 (1970) (action for damages and declaratory relief that had been removed from state to federal court; Supreme Court briefs did not mention a jury trial right).

247. 260 U.S. 156 (1922).

248. See Brief and Argument for Plaintiff in Error at 13, 15, *Keogh v. Chi. & N.W. Ry.*, 260 U.S. 156 (1922) (No. 823).

249. See *Keogh*, 260 U.S. 156.

properly dismissed²⁵⁰ does not provide support for the proposition that a claim carrying a right to a jury can be stayed pending referral of a jury question to an agency for decision; rather, *Keogh* can be seen as applying the principle (later explicitly adopted by the Court²⁵¹) that a litigant can be precluded from relitigating an issue determined in a prior proceeding even if the prior proceeding was one in which there was no jury trial.

In *Carnation Co. v. Pacific Westbound Conference*,²⁵² the Court was asked to decide whether the Federal Maritime Commission's jurisdiction over shipping rates precluded a plaintiff from suing for antitrust damages arising from shipping conferences' implementation of rate agreements.²⁵³ The plaintiff and defendants focused their Supreme Court briefs on the question of exclusivity: Did the Shipping Act (administered by the Federal Maritime Commission (FMC)) provide the sole avenue for challenging rate agreements, or could a plaintiff also bring an antitrust claim in federal court? The plaintiff argued that if the Shipping Act were construed to exclude the antitrust remedy, that construction would "improperly . . . deprive [the plaintiff] of a right to trial by jury."²⁵⁴

The FMC argued that the antitrust action should be stayed, rather than dismissed, so that the FMC could determine whether the rate agreements violated the Shipping Act.²⁵⁵ The defendants disagreed, insisting that dismissal, rather than a stay, was the appropriate disposition.²⁵⁶ The plaintiff, as well, continued to focus on the question of dismissal, and continued to argue that if the Shipping Act provided the exclusive remedy, that would violate the Seventh Amendment.²⁵⁷ Though the plaintiff also contended that the legality of the rates

250. *See id.* at 163.

251. *See Parklane Hosiery Co. v. Shore*, 439 U.S. 322, 337 (1979).

252. 383 U.S. 213 (1966).

253. *See id.* at 215.

254. Petition for a Writ of Certiorari at 4, *Carnation Co. v. Pac. Westbound Conference*, 383 U.S. 213 (1966) (No. 20).

255. *See* Memorandum for the Federal Maritime Commission at 5-6, *Carnation Co.* (No. 657); Brief for the United States and the Federal Maritime Commission at 13, *Carnation Co.* (No. 20).

256. *See* Supplemental Brief in Opposition for Respondents, Far East Conference, and Members and Certain Former Members Thereof Named as Defendants at 4, *Carnation Co.* (No. 657); Supplemental Brief in Opposition for Respondent Pacific Westbound Conference at 6-7, *Carnation Co.* (No. 20); Brief for Respondent Pacific Westbound Conference at 10-11, *Carnation Co.* (No. 20).

257. *See* Petitioner's Brief at 7, 56-59, *Carnation Co.* (No. 20). Responding to this argument, one of the defendants asserted that limiting the plaintiff to the Shipping Act's administrative remedy would not violate the Seventh Amendment because the plaintiff sought "damages resulting from a statutory violation unknown at common law"—i.e., the plaintiff was asserting a public

need not first be determined by the FMC, it did not stress the Seventh Amendment in connection with this facet of its argument.²⁵⁸

Meanwhile, the FMC concluded that its approval of an earlier agreement did not encompass the rate agreements challenged in the plaintiff's antitrust suit, and that the latter agreements violated the Shipping Act.²⁵⁹ In its reply brief, the plaintiff asserted that the FMC took the view that "any administrative questions presented that were for determination by the Commission . . . have been determined, and in such way that petitioner is entitled to pursue its [antitrust] litigation" (unless, as the defendants contended, the Shipping Act provided the only possible remedy).²⁶⁰ Unsurprisingly, the plaintiff's reply brief raised no Seventh Amendment challenge to the application of the FMC's findings in its antitrust suit.²⁶¹ Thus, when the Supreme Court held in *Carnation* that the plaintiff's antitrust suit should be stayed "pending the final outcome of the Shipping Act proceedings" (because the FMC's decision had been appealed),²⁶² there was no reason for the Court to consider whether the Seventh Amendment posed any barrier to such a stay, and the Court did not mention the question.

The Supreme Court, then, has not established whether the primary jurisdiction doctrine can be applied to require referral of factual issues to an agency for binding determination when those issues arise in a federal lawsuit on claims for which there is a Seventh Amendment right to a jury trial. The few commentators to discuss the question have noted the existence of doubt.²⁶³ There

rights claim, not a private rights claim. See Brief for Respondent Pacific Westbound Conference at 56 n.51, *Carnation Co.* (No. 20) (citing *NLRB v. Jones & Laughlin Steel Corp.*, 301 U.S. 1, 22 (1937)). Another set of defendants similarly disputed the contention that excluding antitrust suits would violate the Seventh Amendment. See Brief for the Respondents Far East Conference at 48-49, *Carnation Co.* (No. 20).

258. See Petitioner's Brief at 24-25, *Carnation Co.* (No. 20). The plaintiff did state—in response to the contention that the outcomes of jury trials on antitrust claims might vary from case to case—that "[i]f the results turn out to be not entirely consistent, but still sustainable, that must be laid to the workings of the Seventh Amendment." *Id.* at 79. This brief mention, however, did not present a clear argument that a stay, as opposed to dismissal, of the antitrust claims would violate the Seventh Amendment.

259. See *Carnation Co.*, 383 U.S. at 223 n.6.

260. Petitioner's Reply Brief at 9, *Carnation Co.* (No. 20).

261. See *id.*

262. *Carnation Co.*, 383 U.S. at 223-24.

263. See, e.g., Robert B. von Mehren, *The Antitrust Laws and Regulated Industries: The Doctrine of Primary Jurisdiction*, 67 HARV. L. REV. 929, 963 (1954) (noting that "at least in jury cases, there appears to be an insurmountable obstacle—the Seventh Amendment—to making an agency's findings of fact conclusive"); 5 JACOB A. STEIN ET AL., ADMINISTRATIVE LAW § 47.03[2] (2004) ("It has been argued that a court which refers to an agency questions of fact, as opposed to questions of law, should not be bound by the agency's decision because of the Seventh

seems to be no persuasive reason to distinguish such cases from any other instance in which a federal court contemplates referring a fact issue to a non-jury decisionmaker for binding determination; and the Court has made clear that the Seventh Amendment bars such referrals.²⁶⁴ Accordingly, it appears likely that private enforcement within the hybrid system could founder upon the Seventh Amendment difficulty.²⁶⁵

Amendment's guarantee of a right to trial by jury.") (citing von Mehren).

264. *See supra* note 242. It should be noted that other possible applications of the primary jurisdiction doctrine (or similar schemes) can be permissible, even as to plaintiffs who would have a right to a jury trial if they were permitted to sue in federal court. Thus, for example, Congress can preempt a common law claim for damages (and leave persons injured in the future without a remedy) without violating the Seventh Amendment. *See* Nagareda, *supra* note 15, at 354 n.288. Also, Congress can provide that a "public rights" claim for damages falls within the exclusive jurisdiction of an agency, such that the claim must be dismissed if it is brought in federal court. Such an arrangement would relegate the claim to decision by the agency without a jury, but, as discussed above, this would not violate the Seventh Amendment. *See supra* note 239 and accompanying text. However, if Congress instead attempts to provide that the "public rights" damages claim can be brought in federal court, but that certain fact issues must be referred to the agency for binding decision, the plaintiff would have a Seventh Amendment right to a jury trial and the referral would violate that right. *See supra* note 240 and accompanying text.

265. To avoid impairing the jury trial right, Congress could provide for a jury trial and direct that the FDA render an advisory opinion that could persuade, rather than bind, the jury. Such a system could assist the jury in determining difficult issues, but because the jury would retain the ability to reject the panel's findings, the system would not ensure uniformity, and it might not gain the confidence of potential defendants.

Another possible argument is that there should be a "complexity exception" to the Seventh Amendment: Some commentators contend that as to highly complex issues requiring technical or scientific expertise, there should be no jury trial right. *See, e.g.,* Graham C. Lilly, *The Decline of the American Jury*, 72 U. COLO. L. REV. 53, 80 (2001) (arguing that a complexity exception "seems especially appropriate . . . when a forthcoming trial is likely to be protracted and involve difficult technical or scientific issues"). But though at least one appellate court has endorsed such an exception, *see In re Japanese Elec. Prods. Antitrust Litig.*, 631 F.2d 1069, 1086 (3d Cir. 1980) (balancing due process and Seventh Amendment rights and finding "the most reasonable accommodation . . . to be a denial of jury trial when a jury will not be able to perform its task of rational decisionmaking with a reasonable understanding of the evidence and the relevant legal standards"), other courts have rejected it, *see, e.g., In re U.S. Fin. Sec. Litig.*, 609 F.2d 411, 431 (9th Cir. 1979) ("Not only do we refuse to read a complexity exception into the Seventh Amendment, but we also express grave reservations about whether a meaningful test could be developed were we to find such an exception."), and the Supreme Court has not yet resolved the dispute. Admittedly, an issue can be given to the judge rather than the jury—despite the fact that the issue arises in a case involving a jury right—if the issue is not one that was historically reserved for the jury and if the relative capabilities of juries and judges tilt the policy analysis in favor of judicial determination. *See Markman v. Westview Instruments, Inc.*, 517 U.S. 370, 372 (1996)

d) Qui Tam Claims in a Hybrid System

Maintenance of a *parens patriae* suit by a *qui tam* relator, by contrast, would both comport with Article III and avoid the Seventh Amendment problem.

A *qui tam* suit within the hybrid system would comply with Article III under *Crowell's* "adjunct" test, for the same reasons discussed in the preceding Section.²⁶⁶ Indeed, the Article III argument in favor of the hybrid system would be even stronger in the *qui tam* context, because the suit, brought on behalf of the United States, would all the more clearly fall within the "public rights" doctrine.

A defendant ordinarily would have a Seventh Amendment jury right regarding *qui tam* claims brought within the hybrid system; the fact that the claim was brought in the government's name would not change the analysis.²⁶⁷ However, as discussed in Section III.A, the defendant would waive that right in advance, when it opted into the federal products liability system.²⁶⁸ (It would, in any event, be an unusual products liability defendant that complained of being deprived of a jury trial.) And because the *qui tam* relator would be pressing a claim on behalf of the government, Congress could, in the statutory scheme, waive any right to a jury trial on the relator's behalf.

Admittedly, use of the *qui tam* mechanism would introduce some additional constitutional issues. But though the *qui tam* mechanism has been challenged on both Article II and Article III grounds, the more persuasive view holds that it is constitutional. The Supreme Court rejected the Article III challenge in *Vermont Agency of Natural Resources v. United States ex rel. Stevens*,²⁶⁹ when it held that "a *qui tam* relator under the [False Claims Act] has Article III standing."²⁷⁰

(holding that "the construction of a patent, including terms of art within its claim, is exclusively within the province of the court"). It is far from clear, however, that issues of safety and causation in a products liability case would meet this test.

266. See *supra* notes 222-236 and accompanying text.

267. See *Tull v. United States*, 481 U.S. 412, 420 (1987).

268. The proposal outlined here could hardly be viewed as impermissibly coercive. After all, the traditional baseline presumption is that states can regulate products that affect health and safety, and that such regulation can be accomplished through the tort system. Accordingly, a system that permits companies to opt in to an alternative system (or to opt out, and be subject to state-law tort claims) benefits the company by expanding its choices. Cf. Seth F. Kreimer, *Allocational Sanctions: The Problem of Negative Rights in a Positive State*, 132 U. PA. L. REV. 1293, 1300-01 (1984) (arguing that in assessing the permissibility of governmental allocations of benefits, courts should distinguish threats—i.e., "allocations that make a citizen worse off than she otherwise would be because of her exercise of a constitutional right"—from offers—i.e., allocations that "merely expand her range of options, leaving the citizen better off").

269. 529 U.S. 765 (2000).

270. *Id.* at 778. The Court cited "the doctrine that the assignee of a claim has standing to assert

However, the *Stevens* majority declined to address “whether *qui tam* suits violate Article II, in particular the Appointments Clause of § 2 and the ‘take Care’ Clause of § 3.”²⁷¹

The Appointments Clause vests the President with power to appoint federal officers with the advice and consent of the Senate, but provides that Congress may vest the appointment of “inferior” officers “in the President alone, in the Courts of Law, or in the Heads of Departments.”²⁷² Some have argued that the False Claims Act’s *qui tam* provision violates this clause by permitting *qui tam* relators to function as federal officers without an appropriate appointment.²⁷³ *Qui tam* relators, however, should not be viewed as “officers,” because they have no established position, they draw no salary, and they serve their function on an ad hoc basis.²⁷⁴

The Take Care Clause provides that the President shall “take Care that the Laws be faithfully executed.”²⁷⁵ Though it is not entirely clear whether this clause vests power in the President or instead imposes a duty,²⁷⁶ under either interpretation the clause should pose no problem for *qui tam* provisions. It seems clear that in cases where the government intervenes in a *qui tam* suit, the *qui tam* mechanism does not impair the government’s law-enforcement functions.²⁷⁷ And because the executive branch retains significant control over a *qui tam* suit even in cases where the government chooses not to intervene,²⁷⁸ the more persuasive

the injury in fact suffered by the assignor,” and reasoned that the False Claims Act “can reasonably be regarded as effecting a partial assignment of the Government’s damages claim.” *Id.* at 773.

271. *Id.* at 778 n.8.

272. U.S. CONST. art. II, § 2, cl.2.

273. See Richard A. Bales, *A Constitutional Defense of Qui Tam*, 2001 WIS. L. REV. 381, 430-33 (reviewing Appointments Clause arguments concerning *qui tam* mechanism).

274. See *United States ex rel. Stone v. Rockwell Int’l Corp.*, Nos. 99-1351, 99-1352, 99-1353, 2004 WL 433235, at *17-*18 (10th Cir. Mar. 5, 2004) (finding *qui tam* relators do not fall within Appointments Clause because they are not federal officers); *Riley v. St. Luke’s Episcopal Hosp.*, 252 F.3d 749, 757-58 (5th Cir. 2001) (en banc) (same); *United States ex rel. Taxpayers Against Fraud v. Gen. Elec. Co.*, 41 F.3d 1032, 1041 (6th Cir. 1994) (same); *United States ex rel. Kelly v. Boeing Co.*, 9 F.3d 743, 759 (9th Cir. 1993) (same); Bales, *supra* note 273, at 430-33 (reviewing reasons why *qui tam* relators should not fall within scope of Appointments Clause).

275. U.S. CONST. art. II, § 3.

276. See, e.g., Caminker, *supra* note 159, at 356 (“The Supreme Court has suggested occasionally that the ‘take Care’ clause vests the President with prosecutorial discretion over Federal law enforcement, but this clause is better viewed as a mandate to follow the will of Congress than as a grant of exogenously defined power.”).

277. See *Stone*, 2004 WL 433235, at *19 (rejecting the “contention that the presence of a *qui tam* relator in the litigation so hindered the Government’s prosecutorial discretion as to deprive the Government of its ability to perform its constitutionally assigned responsibilities”).

278. See *supra* notes 164-167 and accompanying text.

view holds that those cases do not violate the Take Care Clause either.²⁷⁹

Constitutional considerations, then, would likely eliminate one of the four options: The Seventh Amendment would probably bar the implementation of the private claim/hybrid adjudication model.²⁸⁰ In the Subsection that follows, I will compare the merits, from a policy standpoint, of the three remaining options.

2. Policy Considerations

The *qui tam*/hybrid scheme appears to be the most attractive of the remaining possibilities for linking litigation with the regulatory system. Considerations of cost and speed are inconclusive; however, the quality of decisionmaking and the quality of advocacy in the *qui tam* system would be better, on balance, than those in systems that relied upon agency adjudication of either government or private claims.

279. See *Riley*, 252 F.3d at 757 (“Any intrusion by the *qui tam* relator in the Executive’s Article II power is comparatively modest, especially given the control mechanisms inherent in the FCA to mitigate such an intrusion and the civil context in which *qui tam* suits are pursued.”); *Taxpayers Against Fraud*, 41 F.3d at 1041 (explaining in dictum that *qui tam* suits in which the government does not intervene do not contravene separation of powers principles because the government retains means of controlling *qui tam* suits even if it chooses not to intervene); *Kelly*, 9 F.3d at 755 (“[T]he Executive Branch exercises at least an equivalent amount of control over *qui tam* relators as it does over independent counsels. Thus, the FCA gives the Attorney General sufficient means of controlling or supervising relators to satisfy separation of powers concerns.”); see also Bales, *supra* note 273, at 435 (“Comparing the *qui tam* provisions of the FCA to the independent counsel provisions upheld in *Morrison v. Olson*, 487 U.S. 654 (1988)] presents the strongest argument as to why judicial branch involvement in *qui tam* actions does not violate separation of powers principles.”); Bucy, *supra* note 152, at 956 (suggesting that “because an effective private justice model brings an invaluable and otherwise unobtainable resource to public regulatory efforts, namely inside information, the executive branch is unable to ‘take Care’ that laws are faithfully executed without such a model”).

Objections to *qui tam* suits have also been based on separation of powers more generally; these arguments, too, should be rejected. See, e.g., *Kelly*, 9 F.3d at 755-56 (rejecting the contention that “the *qui tam* provisions disrupt the proper balance of power between the three branches by permitting the Judicial Branch to encroach on executive authority”); Bales, *supra* note 273, at 435 (arguing that “because *qui tam* disperses power among the citizens rather than concentrating it in the hands of a single political branch, the principles underlying the separation of powers doctrine are not threatened as they are when, for example, Congress seeks to retain the power constitutionally apportioned to another branch”).

280. This assumption is based on the likelihood that the proposed scheme would not fall within any possible “complexity” exception to the Seventh Amendment. See *supra* note 265.

a) *Cost and Speed*

The agency-adjudication options might produce some cost savings relative to the hybrid-adjudication options, but the additional cost of hybrid adjudication should be weighed against its benefits.²⁸¹

Agency adjudication might cut the costs of litigation if it provided a narrower range of discovery than is customary in civil litigation. Though some agencies provide for a range of discovery similar to that available in federal court,²⁸² the rules of other agencies “may severely restrict access to discovery.”²⁸³ The FDA’s rules for civil penalty proceedings, for example, permit depositions only upon a showing of necessity and then only for the purpose of preserving testimony.²⁸⁴

Such restrictions on discovery, however, would reduce the effectiveness of litigation as a way to uncover safety-related information. Depositions, for instance, can be a key tool to uncover internal policies and deliberations within a company.²⁸⁵ The savings achieved by restricting discovery to narrower limits than those imposed in federal court would therefore come at a significant cost.

In the hybrid system, some additional delay might result from the referral process, but that delay need not be excessive. The referral would occur at a point in the process when discovery was complete, and summary judgment motion practice would have served to narrow and focus the issues prior to trial. Because the proceedings in the hybrid system would not involve a jury, the referral process would not cause undue disruption; proceedings can more readily be segmented in bench trials than in jury trials.²⁸⁶

281. Cf. Kevin M. Clermont & Theodore Eisenberg, *Litigation Realities*, 88 CORNELL L. REV. 119, 130 (2002) (“Delay is an unavoidable feature of life, and it is not an evil in itself. The only evil is excessive delay, where excessive means that the costs of delay outweigh its benefits.”).

282. See 4 STEIN ET AL., *supra* note 263, § 23.01[2], at 23-15 (“Agencies such as the Federal Trade Commission, the Federal Maritime Commission, and the Federal Communications Commission . . . have closely modeled their discovery rules on the Federal Rules [of Civil Procedure].”).

283. *Id.* § 23.01, at 23-28 (discussing the NLRB). “For example, N.L.R.B. rules do not permit depositions except for the purpose of preserving testimony, and then only when in the discretion of the regional director or administrative law judge good cause has been shown.” *Id.* at 23-28.

284. See *supra* text accompanying note 125.

285. See *infra* Subsection III.B.2.

286. Another factor that bears upon litigation costs concerns the possibility of multiple suits. Both the agency-enforcement and *qui tam* options would structure the dispute as a single proceeding; by contrast, private claims in the agency setting could proceed singly as well as in a class format. Because they would require the resolution of all claims in a single proceeding, the agency-enforcement and *qui tam* systems could, overall, prove more efficient—though they also

b) Decisionmaking: Bias and Expertise

A consideration that supports situating the proceeding in federal court is that the court could provide better decisionmaking than the agency with respect to discovery and damages. The agency's comparative advantage regarding technical and scientific questions does not extend to all other issues that would arise in product safety litigation.

The political influence of the pharmaceutical industry is widely noted.²⁸⁷ Critics also charge that recent changes have rendered the FDA, in particular, more vulnerable to industry influence,²⁸⁸ and that the agency, "although quick to approve drugs, . . . is slow to take them off the market when they prove dangerous."²⁸⁹ Given that each of the options discussed here would accord the FDA significant power over industry liability, care should be taken, in crafting the system to improve the FDA's independence. As Marcia Angell suggests, measures could include enhanced public funding for the agency, as well as enforcement of conflicts prohibitions for those serving on the FDA's advisory committees.²⁹⁰ Even assuming such measures were adopted, however, there would remain a risk that the FDA could unduly defer to the interest of industry players. It is therefore useful to assess the degree to which the options discussed here would protect against the possibility of agency capture.²⁹¹

Despite the internal separation of functions within the FDA—which would insulate an FDA hearing officer to some extent from the other parts of the agency—that officer might be subject to some pressures to accommodate industry by limiting the nature and scope of discovery against a products liability

might prove more cumbersome—than individual private claims within the agency.

287. See, e.g., ANGELL, *supra* note 47, at 198 ("The pharmaceutical industry has by far the largest lobby in Washington According to the consumer advocacy group Public Citizen, from 1997 through 2002, the industry spent nearly \$478 million on lobbying."); *id.* at 200 ("In the 1999-2000 election cycle, drug companies gave \$20 million in direct campaign contributions plus \$65 million in 'soft' money."); Sheryl Gay Stolberg & Gardiner Harris, *Re-Examining Medicare—The Drug Industry's Muscle: Industry Fights To Put Imprint on Drug Bill*, N.Y. TIMES, Sept. 5, 2003, at A1.

288. Marcia Angell has observed that the Prescription Drug User Fee Act, which seeks to speed the processing of drug applications by providing for payment of user fees, "makes the FDA dependent on an industry it regulates." ANGELL, *supra* note 47, at 208.

289. *Id.* at 209.

290. See *id.* at 243-44.

291. Obviously, federal courts could also be staffed with judges who are predisposed to favor industry defendants over products liability plaintiffs. However, Article III tenure protects judges from suffering repercussions as a result of decisions adverse to industry—which makes it likely that judges would be more willing, overall, to make such decisions.

defendant.²⁹² This might be especially true if the discovery in question had the potential to embarrass the agency—for example, by showing that the agency had failed adequately to respond to earlier evidence of safety concerns.²⁹³ Indeed, David Graham, the Associate Director for Science in CDER's Office of Drug Safety, recently raised a very similar concern with respect to the FDA's current practice of postmarketing surveillance.²⁹⁴ In the litigation context, permitting the plaintiff to develop evidence through discovery proceedings under the aegis of the court could prove a more effective information-gathering tool.

Of course, to the extent that there is agency bias of the sort suggested here, the problem will extend well beyond discovery. If the FDA's decisionmaking were unduly influenced by certain stakeholders, that could distort its review of safety and causation issues under the hybrid system as well. Such a concern may lead policymakers to reject any determinative role for the FDA in resolving tort claims, and I turn to that argument below. For the moment, it is significant to note that if policymakers were to grant the FDA a determinative role, the choice of institutional configuration would be important. Problems of FDA bias would be mitigated in the hybrid system by the fact that the record on which the agency made the safety and causation decisions would include evidence developed in a separate federal court proceeding. If that discovery process uncovered safety problems with a product, the resulting publicity could provide an inducement for the FDA to scrutinize the product carefully and to find liability where appropriate. By contrast, if the discovery process unfolded within the agency, a captured agency might prevent the plaintiff from ever developing certain evidence concerning liability.

Moreover, the reasons for submitting the safety and causation questions to the FDA do not extend to the discovery process. Even if it is taken as given that scientific and policy judgments concerning medical products should be left to FDA experts, there is nothing about the discovery process that requires a similar

292. *Cf.* McGarity, *supra* note 44, at 564 ("The very real possibility of agency capture by the regulated industry means that federal officials are not always eager to eliminate wrongful attempts to manipulate the regulatory process . . .").

293. *Cf.* Noah, *supra* note 66, at 503 (noting "the FDA's natural hesitancy to confess error when a drug it just approved generates unusual and unexpected rates of adverse reactions").

294. In testimony to a Senate committee, Graham warned:

[T]he new drug reviewing division that approved the drug in the first place and that regards it as its own child, typically proves to be the single greatest obstacle to effectively dealing with serious drug safety issues. The second greatest obstacle is often the senior management within the Office of Drug Safety, who either actively or tacitly go along with what the Office of New Drugs wants.

Vioxx Hearings, *supra* note 84 (statement of David J. Graham, Assoc. Dir. for Sci., Office of Drug Safety, FDA).

approach. Federal district judges and magistrate judges handle discovery disputes in complex litigation on a regular basis. They are expert at it. Indeed, there is some reason to think that a generalist district judge or magistrate judge might be better situated to handle discovery in a complex products liability case than a specialist hearing officer within the FDA: The experience that the generalist judge gains with discovery disputes in other types of cases could help to ensure that the discovery permitted in products liability litigation was calibrated at the level thought to be appropriate in general civil litigation.²⁹⁵

Similar considerations apply to the determination of damages. Once the FDA had settled the issue of safety and had provided guidance concerning the factors that should determine causation,²⁹⁶ the federal court could handle the question of damages at least as competently as the agency. Indeed, the expertise gained by the court in assessing damages in other types of cases would provide a useful source of cross-pollination. There is also some reason to think that the FDA itself might prefer not to be tasked with determining damages. Such determinations are likely to be fraught with controversy, and the agency might well prefer to leave the question to a separate institution.²⁹⁷

c) Litigating: Incentives and Expertise

The three proposals differ with respect to the entity pressing the claim as well as the entity deciding it. A comparison of the options therefore should consider the relative expertise of the litigator in each system, as well as that

295. Questions relating to the scope of, and limits on, discovery are not uncontroversial. See, e.g., Stephen B. Burbank & Linda J. Silberman, *Civil Procedure Reform in Comparative Context: The United States of America*, 45 AM. J. COMP. L. 675, 701 (1997) (noting that “[t]he responses of practicing lawyers to . . . the 1993 amendments to Rule 26 (discovery), were very seriously negative”); Thomas D. Rowe, Jr., *A Square Peg in a Round Hole? The 2000 Limitation on the Scope of Federal Civil Discovery*, 69 TENN. L. REV. 13 (2001) (critiquing the 2000 amendments to the discovery provisions in the Federal Rules of Civil Procedure). My point is merely that the experience with discovery in other types of complex litigation can profitably be applied to questions concerning discovery in products liability litigation concerning FDA-approved products; and situating the discovery process within the federal courts, rather than within the FDA, would make possible the application of that experience.

296. The FDA would determine the question of product safety—i.e., whether the product is safe enough to remain on the market and, if so, whether additional warnings are needed. In a case where the FDA found the product unsafe or the warnings inadequate, the FDA would also enumerate the factors that the district court should apply in order to determine specific causation—i.e., whether a particular claimant’s injury should be deemed to arise from the safety problem or inadequate warning.

297. Obviously, liability determinations would often be controversial as well. But the FDA’s relative expertise with respect to safety and causation issues would counterbalance this concern.

litigator's incentives to press valid claims. Under this analysis, the government-enforcement model may fare less well, both because the government's litigation resources would be limited²⁹⁸ and because the private plaintiffs' bar may add useful expertise. By contrast, the *qui tam* proposal might help to ensure that claims are litigated by lawyers with appropriate expertise, and would provide structured incentives for industry insiders to bring forward nonpublic information concerning safety.

As noted above, the government-enforcement model would require Congress to dedicate significantly more resources to enforcement.²⁹⁹ In theory, such a system could be financed by the private sector, through an ex ante system of user fees exacted during the product approval process³⁰⁰ or through an ex post system of fee-shifting.³⁰¹ Such an innovation, however, would likely be controversial. Absent such a measure, the government might well be unwilling to invest the significant resources that might be necessary to establish liability in a complex products liability case.³⁰²

In any event, excluding the private bar from the enforcement of product safety standards would raise issues apart from the question of resources.³⁰³ On

298. Cf. McGarity, *supra* note 44, at 564 (noting that "the reality of very limited agency resources means that even those officials who are committed to seeking out and eliminating fraud are generally not able to do so").

299. As Michael Green noted in 1997, "The FDA is woefully underfunded for its mandate, which includes regulatory oversight of products that account for more than twenty-five percent of all American consumer purchases." Green, *supra* note 46, at 476.

300. Congress has enacted legislation requiring companies to pay a fee when they apply for approval of a new drug or biologic; the fees help to pay the costs of speedier FDA review. See Prescription Drug User Fee Act of 1992, Pub. L. No. 102-571, Title I, 106 Stat. 4491 (codified as amended at 21 U.S.C. §§ 379g, 379h (2000)).

301. A fee-shifting provision could provide, for example, that the government could recover the reasonable costs of a successful lawsuit. Such a provision could provide for either one-way or two-way fee-shifting. Currently, the Equal Access to Justice Act authorizes fee-shifting in favor of respondents in administrative adversary proceedings where the agency's position in bringing the proceeding was not substantially justified. 5 U.S.C. § 504(a)(1) (2000). However, most device makers, and probably all pharmaceutical companies, would be ineligible to receive fees under the Act because their net worth exceeds the Act's limitations. See *id.* § 504(b)(1)(B).

302. Cf. Herbert M. Kritzer, *From Litigators of Ordinary Cases to Litigators of Extraordinary Cases: Stratification of the Plaintiffs' Bar in the Twenty-First Century*, 51 DEPAUL L. REV. 219, 235 (2001) (noting, with respect to the plaintiffs' bar's involvement in state tobacco litigation, that "[s]tates turned to contingency fee arrangements as a way of eliminating their own risks of having to devote substantial dollars or other resources to the litigation").

303. Cf. 2 ALI REPORTERS' STUDY, *supra* note 114, at 86 ("Regulatory agency 'failure' may occur because of inadequate resources or on account of political or bureaucratic pressures. A system of privately initiated tort remedies, administered through the decentralized, general purpose

one hand, the agency might fail to initiate proceedings that should be brought. Enforcement personnel within the FDA might not always be quick to question the FDA's own prior safety determinations—yet that questioning would be desirable in some instances, when postmarketing experience discloses a previously unknown safety issue. On the other hand, agency personnel would not have the direct personal stake in the outcome that leads plaintiffs' lawyers to be careful in selecting which cases to take: Agency personnel would be paid their salary whether or not a given enforcement action resulted in a victory for the agency.

By contrast, the private plaintiffs' lawyers who would litigate actions in the other two models would have a strong incentive to screen cases, because they would recover fees only if they obtained a judgment or settlement. The more expertise a plaintiff's lawyer possesses in the field of medical products liability, the more likely the lawyer is to assess accurately a claim's potential for success. Not only has the plaintiffs' bar generally become more specialized in recent years,³⁰⁴ but the firms handling complex, high-end cases have increased both their expertise and their resources.³⁰⁵ Such firms, when they specialize in products liability cases, are likely to possess high concentrations of both procedural expertise and substantive medical expertise.³⁰⁶ They also have the resources to invest in medical and scientific experts and to commission the type of data-mining projects that could disclose emerging safety issues.³⁰⁷

A system that employed *qui tam* suits to litigate product safety issues might help to ensure that cases were litigated by firms that possessed the necessary expertise.³⁰⁸ Individual plaintiffs may fail to select the most experienced counsel

court system, can serve as a corrective for these shortcomings.”).

304. See, e.g., John P. Heinz et al., *The Changing Character of Lawyers' Work: Chicago in 1975 and 1995*, 32 LAW & SOC'Y REV. 751, 753-54, 761-62 (1998) (reporting, based on a mid-1990s survey of Chicago lawyers, that lawyer specialization “appears to have increased substantially,” especially among plaintiffs' personal injury lawyers).

305. See Nagareda, *supra* note 15, at 319-20.

306. See Kritzer, *supra* note 302, at 231 (“Law firms that litigate huge, complex cases, such as tobacco, breast implant, and the like, require staff and financial resources beyond the scale of the traditional plaintiffs' firms.”); *id.* at 232 (noting the emergence of “repeat player” plaintiffs' firms with “the ability to bring to bear substantial legal effort and to deal with the cost of extended, monster-scale litigation”).

307. Although data mining could provide a powerful tool to identify emerging issues, there is some question whether the private bar would have access to useful databases, in the light of patient privacy concerns. It seems likely, however, that there will exist at least some relevant databases that are available for private analysis.

308. Defendants, as repeat players in products liability litigation, are likely to retain lawyers with substantive and procedural expertise. See Susan Brodie Haire et al., *Attorney Expertise*,

to represent them; though informal networks—such as referrals by generalists to specialist attorneys—may help to bridge the informational gap, some plaintiffs with valid claims may select less than expert representation.³⁰⁹ By contrast, in a *qui tam* setting, a potential claim could attract more than one set of plaintiffs' attorneys, and the court could select among them based upon their resources and expertise.³¹⁰

Another advantage of the *qui tam* mechanism is that it would provide an incentive for industry insiders to act upon information concerning safety problems: An insider possessing such information could bring a *qui tam* claim and share in the resulting recovery. Admittedly, there are other ways to provide incentives for the disclosure of such information. For example, Congress could provide a bounty for the provision of information that leads to a successful government penalty action.³¹¹ However, some insiders might mistrust a reward system in which the availability of the reward would depend on the government's decision to litigate, and success in establishing the claim; such insiders might be more likely to come forward if they could bring suit themselves as *qui tam* relators.³¹²

Policy considerations, then, suggest that the *qui tam*/hybrid scheme provides the best alternative for linking the litigation and regulatory systems. Considerations of cost and speed are inconclusive: Agency proceedings may be

Litigant Success, and Judicial Decisionmaking in the U.S. Courts of Appeals, 33 L. & SOC'Y REV. 667, 674-77 (1999) (reporting results of study that analyzed degrees of specialization of lead counsel for plaintiff and defendant in sample of products liability cases drawn from federal appellate opinions on Westlaw; study indicated that defendants' counsel tended to have more procedural experience, and somewhat more substantive expertise, than plaintiffs' counsel).

309. See *id.* at 668 ("Although the high stakes of products liability litigation has created a financial incentive for many plaintiffs' lawyers and firms to orient their practice in this area, individual plaintiffs may not be capable of making informed judgments when selecting firms or attorneys best suited to represent their interests.").

310. See *supra* note 174 and accompanying text.

311. At least one such provision already exists. See 21 U.S.C. § 333(b)(5) (2000) (providing bounty for information leading to criminal convictions for violations of certain laws concerning sales of drug samples).

312. Such concerns supported Congress's provision of a *qui tam* mechanism in the FCA. As Evan Caminker has explained:

Congress determined that potential rewards alone would not provide sufficient incentive for disclosure; many potential informers are reluctant to come forward because they refuse to accept the 'personal and financial risk' involved absent any 'confidence in the Government's ability to remedy' the misconduct, a fear rectifiable only by allowing for participation in the litigation.

Caminker, *supra* note 159, at 351 (quoting S. REP. NO. 99-345, at 6-7, reprinted in 1986 U.S.C.C.A.N. 5266).

cheaper than hybrid proceedings, but the savings would likely result from streamlined procedures that would diminish the investigative power of the discovery process. Quality of decisionmaking favors the use of the hybrid system, because there are reasons to think that the federal courts could do a more reliable job of supervising discovery and assessing damages. Quality of advocacy weighs in favor of the *qui tam* mechanism, because the incentives and expertise of *qui tam* relators and their counsel could improve the investigation and presentation of potentially valid claims. Having thus suggested that the *qui tam*/hybrid system may be the best option for providing a structural connection between litigation and FDA decisionmaking, I proceed, in the next Part, to consider how such a mechanism would work.

III. STRUCTURING A HYBRID SYSTEM

In this Part, I sketch in somewhat more detail the *qui tam*/hybrid option described above. A distinctive feature of the proposal is that a company would have the option to select the federal *qui tam* system at the time it submitted a product for FDA approval. Opting in would preempt state tort claims concerning the product; in return, the company would be required to submit to a rigorous set of federal products liability standards. In Section III.A, I describe the opt-in system; Section III.B considers ways in which the *qui tam* mechanism could improve postmarketing surveillance with respect to companies that opted in.

A. An Opt-In and a *Quid pro Quo*

A central feature of the proposed scheme is that a company would choose, when submitting a product for FDA approval, whether to opt in to the *parens patriae* system with respect to that product. By choosing to opt in, the company would disclaim any constitutional objections to the adjudicatory scheme described in Part II.³¹³ Companies' choices concerning the opt-in could also shed light on their true assessments of the jury system. In addition, the scheme offers a chance to obtain a *quid pro quo*: If the jury system imposes high uncertainty costs on companies, companies should be willing to opt in to the proposed scheme even if it broadens the range of situations in which some amount of compensation must be paid.

1. Revealing Companies' Views About the Tort System

Many in the corporate sector are quick to complain about the tort system. Critics frequently assert that juries are incompetent to handle technical or

313. See *supra* note 268 and accompanying text.

scientific questions, that they favor plaintiffs, that they award excessive damages, and that their determinations are irrational and unpredictable. Such contentions loom large in the preemption debate: Lay juries, defendants assert, should not be permitted to second-guess the expert determinations made by the FDA. The proposal outlined here offers a chance to illuminate companies' perceptions of litigation risk.³¹⁴ Given the chance, will companies exchange exposure to the jury system for a scheme that subjects liability to expert agency determination and imposes scheduled damages assessed by a judge?

In this regard, the opt-in system would address one of the dilemmas of the preemption debate. On one hand, preemption is disfavored because it deprives plaintiffs of compensation and displaces state law in a traditional area of state regulation. On the other, defendants—and now the FDA—argue that without preemption, the threat of liability will deter innovation. The problem is that it is difficult to know when, and to what extent, preemption is necessary in order to promote innovation: Is it really true that the threat of state tort liability will cause Company X to exit a line of research? After all, companies may make such arguments—whether or not they are true—in an attempt to decrease their liability exposure.³¹⁵ Research and development decisionmaking is particularly hard to study because it centers on nonpublic information.³¹⁶ Moreover, numerous factors may influence the incentive effects of liability risk. New drug development occurs mostly within large pharmaceutical companies, while new

314. Of course, the criticisms of the civil justice system described in the text are highly contested. But if the concern is whether innovation will be deterred or promoted, policymakers may validly consider the industry's perception of litigation risk, whether or not that perception is accurate. See Steven Garber, *Product Liability, Punitive Damages, Business Decisions and Economic Outcomes*, 1998 WIS. L. REV. 237, 250-51 ("To dismiss misperception by company decisionmakers (as many do) as 'their problem' misses the key point: If misperception contributes to manufacturer decisions and economic outcomes that are socially undesirable, that is also *our* (i.e., society's) problem.").

The notion of "revealed preferences"—i.e., the theory that an actor's choices can reveal the actor's preferences—has been criticized. See, e.g., Richard H. Pildes & Cass R. Sunstein, *Reinventing the Regulatory State*, 62 U. CHI. L. REV. 1, 76-80 (1995) (noting, among other things, that choices can be highly contextual). However, companies' choices concerning the opt-in system described in the text would accurately reveal their views concerning the precise question at issue in the preemption debate: the extent of the risk companies perceive to flow from the civil justice system.

315. See GARBER, *supra* note 96, at 3 n.2 (noting, with respect to "surveys of the business community," that "respondents . . . have incentives to exaggerate detrimental effects of liability and understate beneficial ones").

316. See *id.* at 142 ("Because innovation is so crucial to private performance, R&D strategies and activities are typically closely guarded secrets. As a result, very little systematic information about the innovative efforts of individual companies is publicly available.").

medical devices often are developed by smaller enterprises.³¹⁷ The effects of liability exposure on innovation are likely to vary depending on the type of product.³¹⁸

My proposal would put the companies' assertions to the test. A business that chooses to remain subject to state tort law (rather than opting in to the federal liability system) would have a more difficult time establishing that state tort law deters it from societally useful innovations. Thus, for a company that fails to opt in to the federal liability system, there should be no preemption of state tort law. The proposal described here would offer a choice between state tort law (applied by juries) and a federal liability system (applied by an expert panel, the FDA, and a federal judge). Companies that did not opt in to the federal system would be subject to state law to a greater degree than they currently are, because there would be no preemption of state law claims. But they could not claim that they were deterred from innovating, unless they also argued that the compensation scheme provided in the opt-in system deterred innovation as well.³¹⁹ That, at least, would have the salutary effect of moving the debate away from complaints about jury incompetence and toward a focus on the appropriate balance between innovation and compensation. Accordingly, I turn next to a discussion of the shape of liability under the opt-in scheme.

317. *See id.* at 22 (noting that “diversified, and hence relatively large, R&D operations” have an edge in developing new drugs, while “smaller companies are a more typical source of innovation in medical devices”).

318. *See id.* at 144 (noting reports “that product liability has substantially discouraged innovation efforts in vaccines, contraceptives, and orphan drugs”).

319. Relevant judgments about litigation risk will be made by the company itself in some instances; in others, the judgments of liability insurers will be relevant as well. As Steven Garber has noted:

For some companies, some direct costs [of liability] are covered by commercial insurance. But the existence of commercial product liability insurance hardly makes direct liability costs irrelevant. Large companies tend to be self-insured (i.e., uninsured) for product liability. In addition, liability costs paid or reimbursed by insurance companies are costly to insured companies because adverse liability experience is likely to lead to higher insurance premiums—or lack of insurance coverage—in the future. Moreover, punitive damages payments are not insurable or are only partially recoverable in many states.

Garber, *supra* note 314, at 243–44. The opt-in mechanism described here would provide information concerning decisionmakers' views of the relative litigation risks of the state-law tort system and the opt-in system—whether the decisionmaker in question is the company itself or the company's liability insurer.

2. Broadening Compensation

From a procedural and institutional standpoint, the opt-in scheme would promise significant advantages to a potential defendant. Instead of multiple proceedings in various jurisdictions, the *parens patriae* proceeding would be a single action in a single federal district court.³²⁰ In place of liability determinations under varying state tort doctrines, applied by lay juries and generalist judges, the *parens patriae* regime would employ a federal standard of liability applied by a specialist panel (with review by the FDA). The opt-in scheme thus would likely provide a significant reduction in litigation costs and could also promise both uniformity³²¹ and greater predictability.³²²

In return, the scheme could require companies that opted in to submit to more rigorous standards of liability than they might encounter under some state tort regimes. This Subsection sketches the possible outlines of such a liability framework. I do not attempt to show that the measures outlined here constitute optimal levels of liability for the opt-in scheme; rather, I use them to illustrate the notion that the opt-in scheme could impose somewhat more rigorous standards, relative to state tort law, without necessarily deterring innovation. The proof of this, of course, would come in the execution: If companies failed to opt in to the system, that could be a sign that the opt-in liability rules had overreached.

Like state tort law,³²³ the opt-in system would impose strict liability for

320. The possibility that more than one *qui tam* relator might bring a particular claim is addressed above. See *supra* text accompanying notes 171-174.

321. Cf. GARBER, *supra* note 96, at 57 (noting that the tort system's "complexity and interjurisdictional variation in doctrine" contribute to the uncertainty faced by potential defendants).

322. Some potential defendants will also like the opt-in system because it eliminates the possibility of jury trials. Empirical research has rebutted many of the complaints about jury performance, and, indeed, has suggested that products liability defendants may fare no worse before juries than before judges. See Kevin M. Clermont & Theodore Eisenberg, *Trial by Jury or Judge: Transcending Empiricism*, 77 CORNELL L. REV. 1124, 1162 (1992) (examining data from federal court cases from 1979 to 1989 and finding that products liability plaintiffs have higher win rates in bench trials than in jury trials). Obviously, the win rates in bench trials and jury trials are affected by the mix of cases heard by judges and by juries, see *id.* at 1162-63, so that the differing win rates do not in themselves prove that juries are friendlier to defendants than judges are. However, the contrast does suggest that juries are not as credulous concerning plaintiffs' claims as critics of the tort system suggest. Nonetheless, corporate decisionmakers may still believe that jury trials are undesirable. See Clermont & Eisenberg, *supra* note 281, at 146 ("Despite years of research that rebuts stereotypes about juries, every day lawyers and policymakers act on the basis of those stereotypes.").

323. See, e.g., RESTATEMENT (THIRD) OF PRODUCTS LIABILITY, §§ 2(a), 6(b)(1) (1998) (imposing strict liability for manufacturing defects); *id.* § 6 cmt. a (noting that Section 6(b)(1) states a

manufacturing defects. This is uncontroversial, and (from the viewpoint of drug manufacturers), relatively inconsequential: The FDA tightly regulates manufacturing practices, and as a result, cases of manufacturing defects are rare in the prescription drug context.³²⁴ (Manufacturing defects are, however, more common with respect to medical devices.³²⁵) A manufacturing defect defendant would be liable for medical costs, other costs of care, lost wages, and scheduled amounts for pain and suffering.

The treatment of design defects would be more significant, and here the use of the hybrid system could provide a significant benefit compared with ordinary litigation. As I have noted,³²⁶ the courts are divided over the question of pharmaceutical design defects: The question is whether the FDA's risk-benefit determination (in approving a drug for marketing) should ever be second-guessed by courts in the light of later-surfaced information. The hybrid system could avoid this dilemma, by requiring the FDA itself to revisit its safety determination. The system could direct the FDA to apply the same risk-benefit standard it had employed during the premarketing approval phase,³²⁷ but to update the analysis to take account of later-discovered information.³²⁸ If the later-discovered

“traditional rule[.]”).

324. See 5 FRUMER & FRIEDMAN, *supra* note 102, § 50.03A[3], at 50-29 (noting that “prescription drugs are manufactured to stringent standards, overseen by the FDA”). Frumer and Friedman also note that manufacturing defects—to the extent that they occur—would often be difficult to prove, because the patient may consume all of the drug, and because other portions of the same batch “may be used up, contaminated, discarded or changed through age in such a way as to defy any meaningful scientific evaluation.” *Id.* at 50-20 to 50-29.

325. See GARBER, *supra* note 96, at 39 (“Manufacturing defects seem relatively unimportant in the pharmaceutical industry, but they appear more important for medical devices.”). The Shiley heart valve provides a notorious instance. See *id.* at 39 n.19.

326. See *supra* notes 99-104 and accompanying text.

327. See RISK MANAGEMENT REPORT, *supra* note 23, at 21-22 (“Although medical products are required to be safe, safety does not mean zero risk A *safe medical product* is one that has reasonable risks, given the magnitude of the benefit expected and the alternatives available.”).

328. Alternatively, the opt-in system could impose a more stringent standard. For example, the FDA's Task Force on Risk Management noted that one reason why problems that surface in the postmarketing period can affect large numbers of people is that the product rollout may extend to many types of patients. “[I]f use of a new product were evaluated comparatively, the potential extent of injury from an unknown risk might be reduced because the product's initial postmarketing use could be limited to those patients who have been shown to experience a clear therapeutic benefit over an alternative product.” *Id.* at 49. The FDA's premarketing review does not generally require a comparison of the product's safety and efficacy with those of competitor products. However, it might be possible to take comparative efficacy and safety into account *ex post*, in determining the appropriate extent of liability for damages when patients have been harmed by the product.

information altered the risk-benefit analysis to such a degree that the product should no longer be marketed, liability under a product defect theory would be appropriate. Damages could vary depending on when and how the relevant information came to light. If the product's riskiness was both unknown and unknowable at the time of sale, damages might be limited to medical expenses and cost of care up to a capped amount.³²⁹ By contrast, if the relevant information could have been uncovered by the company had it been proactive in self-regulating with regard to safety, then damages could include uncapped medical expenses and cost of care, plus lost wages and scheduled amounts for pain and suffering.

In many instances, later-discovered information may not justify withdrawal of the product, but may require additional warnings. State tort law generally holds defendants liable for failure to warn of known or knowable risks.³³⁰ If information that came to light after premarket approval were found to justify a warning, the opt-in system could impose liability as to claimants who were sold the product without that warning.³³¹ As with product defect claims, damages for failure to warn could vary depending on when the relevant information surfaced and whether a proactive, self-regulating company should have been aware of the need for the warning at the time of the sale.³³²

In all the categories discussed above, the determination could also take into

329. These damages could be considerable. For this reason, in cases where the product defect was unknowable at the time of sale, the statute could provide for a reduction in damages to the extent that such costs were covered by a collateral source such as health insurance.

330. See 5 FRUMER & FRIEDMAN, *supra* note 102, § 50.04[1], at 50-36 (“[I]n the vast majority of states, a manufacturer of prescription drugs has a duty to provide adequate warnings of only those dangers of which the manufacturer knew or should have known at the time of marketing . . .”).

331. Under the “learned intermediary” rule, a defendant usually satisfies its duty to warn by providing appropriate warnings to the health care provider rather than the patient. See GARBER, *supra* note 96, at 40. *But cf.* RESTATEMENT (THIRD) TORTS: PRODUCTS LIABILITY § 6 cmt. e (1998) (arguing that “direct warnings and instructions to patients are warranted for drugs that are dispensed or administered to patients without the personal intervention or evaluation of a health-care provider”). The opt-in system could adopt the “learned intermediary” rule, with appropriate exceptions. *But see* GARBER, *supra* note 96, at 43 (asserting that “liability costs are especially unpredictable where the learned intermediary rule is vulnerable to exception”).

332. Failure-to-warn cases would involve more than one type of causation question. As in all of the cases discussed in the text, the panel would make a general determination concerning whether the product causes the type of injuries at issue, and would provide guidelines for the court to use in determining whether a specific claimant's injuries were caused by the product. Failure-to-warn cases would also require a determination whether the provision of the warning would have prevented the harm to the claimant. This type of causation issue need not be referred to the panel; it could be decided, on a claim-by-claim basis, or with respect to particular classes of claims, by a special master under the direction of the district court.

account whether the company had engaged in misbehavior. Fraud on the FDA or other violations of FDA requirements could help to establish that the product was unsafe.³³³ Moreover, in appropriate cases, fraud, serious violations of FDA requirements, or other egregious behavior could result in an award of punitive damages in addition to the compensatory damages discussed above.

A liability framework along the lines sketched above would extend the scope of liability, in some respects, beyond the boundaries set by some or all states. It would, for example, provide some level of compensatory damages in cases where the company could not yet have known of the relevant danger—such as when a safety issue first surfaces during the marketing of a new product.³³⁴ And it would provide for design defect liability in cases where the FDA later concluded an approved product should be withdrawn from the market, though damages would vary depending on whether the risk was knowable at time of sale.³³⁵

On the other hand, the opt-in framework could promise increased predictability in damages awards. Pain and suffering damages, when available, would be calculated based on a matrix that took into account factors such as the type and extent of injury.³³⁶ This approach would render non-economic damages awards under the opt-in system less variable than comparable awards in jury trials.³³⁷ Though punitive damages would be available under both systems, the opt-in system would ensure that there would be no duplicative punitive awards.³³⁸

333. Under state tort law, “failure to comply with [FDA] regulations is often taken as evidence of negligence per se.” GARBER, *supra* note 96, at 43.

334. As discussed above, state tort law generally does not impose liability for risks that were not knowable at time of sale.

335. As noted above, some states consider all prescription drugs to be “unavoidably unsafe,” which in effect precludes design defect liability (though liability can still be imposed on other theories, such as failure-to-warn).

336. See Randall R. Bovbjerg et al., *Valuing Life and Limb in Tort: Scheduling “Pain and Suffering,”* 83 NW. U. L. REV. 908, 939 (1989) (suggesting a matrix for non-economic damages determinations based on “the severity of injury, the injured person’s age, and the body part affected”).

337. Studies suggest that the non-economic components of jury awards may be more variable than the components that reflect economic damages (such as medical expenses). See *id.* at 937 tbl.3 (recounting results of study of personal injury jury verdicts in Kansas City and Florida in 1970s and 1980s); Shari Seidman Diamond et al., *Juror Judgments About Liability and Damages: Sources of Variability and Ways To Increase Consistency*, 48 DEPAUL L. REV. 301, 317 (1998) (describing results of jury experiment).

338. An ALI Reporters’ Study explained the issues that arise from the possibility of punitive awards in product liability mass torts:

If liability for punitive damages can be established for *any* of the resulting tort claims, then such an award should be available for *all* the claims arising out of the single

The opt-in system would also constrain awards, relative to the largest awards in the jury system. This would be true, for example, when awards in the opt-in system were compared to jury awards that include large components of pain and suffering.³³⁹ It should be noted that media coverage tends to overplay such large awards and to underemphasize the extent to which they are reduced post-verdict by settlement or judicial review.³⁴⁰ Nonetheless, the highest awards under the opt-in system would likely fail to approach the highest awards that might occur in the jury system.

From the claimants' perspective, these changes might not have as great an effect on net compensation as might at first appear. Plaintiffs' net recoveries in the tort system are substantially reduced by costs and contingent attorney fees; in the opt-in system, by contrast, successful relators would recover a reasonable attorney's fee as a separate element of damages.³⁴¹ From the potential

corporate misdeed. Yet the consequence is that beyond . . . compensatory damages . . . the firm will be penalized again and again for a single wrongful judgment or action . . . [S]ubstantial payments for the earlier punitive awards may strip the firm of its insurance coverage and assets, thus endangering the ability of later claimants to [obtain] compensatory redress.

2 ALI REPORTERS' STUDY, *supra* note 114, at 260-61. As this passage indicates, the possibility of multiple punitive awards for the same course of conduct has caused concern. It is far less clear, however, that the actual incidence of punitive awards poses a substantial problem. *See supra* note 11. Nonetheless, to the extent that industry decisionmakers fear the potential for multiple punitive awards, the opt-in system could provide a valuable alternative.

339. In the case of pain and suffering damages, the difference would result from the fact that, in the opt-in system, such damages would be scheduled. With respect to some products, a difference might also arise between the aggregate punitive damages awarded under each system. Empirical research suggests that juries do not differ substantially from judges in awarding punitive damages:

Juries and judges award punitive damages at about the same rate, and their punitive awards bear about the same relation to their compensatory awards. Jury punitive awards have a bit more spread than judge awards, but the effect is not robust and leads to few jury punitive awards outside the range of what judges are expected to award.

Eisenberg et al., *supra* note 11, at 780 app. tbl.1 (reporting results of study of data from trials in 1996 in selected state courts). (As the authors note, the conclusions that can be drawn from these findings are limited by the fact that case characteristics may differ as between bench trials and jury trials; but still the data are suggestive. *See id.* at 746.) However, a difference could arise from the fact that, under the opt-in system, punitives would be determined and awarded, if at all, in a single action, rather than (potentially) in multiple actions concerning the same product. *See supra* note 338 and accompanying text.

340. *See, e.g.,* GARBER, *supra* note 96, at 60 ("[M]ass media seem to provide more complete coverage of plaintiff victories and large and punitive awards than defendant victories, small awards, damages reduced by the judge, or cases overturned on appeal.").

341. *Cf.* 2 ALI REPORTERS' STUDY, *supra* note 114, at 229 (warning that limits on pain and

defendant's perspective, however, even if the mean payout under the opt-in system (taking attorneys' fees into account) were equivalent to that under the tort system, the opt-in system would eliminate the possibility of truly high-end awards. This feature might have significant appeal for decisionmakers to the extent that their deliberations focus on the magnitude of "worst case scenarios" rather than on their probability.³⁴²

As this discussion suggests, the opt-in system might prove attractive to potential defendants while still serving compensatory (and, where appropriate, punitive) goals. As I explain in the next Subsection, the system could also improve the FDA's ability to conduct postmarketing surveillance with respect to the products of companies that opted in.

B. Improving Postmarketing Surveillance

As discussed in Part I, the FDA's postmarketing surveillance of drugs and devices is less than optimally rigorous. Though companies must report adverse events, the data are reported in a summary format that may not disclose all relevant information. The FDA lacks the resources to adequately analyze postmarketing data and lacks sufficient ability to obtain further information from companies when needed. In this Section, I discuss ways in which a *parens patriae* litigation system could supplement the FDA's scarce postmarketing surveillance resources. The system could help FDA regulators to focus their investigative resources, by flagging emerging safety problems. Information unearthed during discovery could provide insights that otherwise might not reach the FDA. And for claims that survived summary judgment, the referral of safety and causation issues to the panel would provide the FDA with a formal occasion

suffering damages should "be adopted only as part of a multifaceted tort reform that would also make the successful plaintiff's attorney fees an independently compensable head of damages").

342. Two studies of corporate executives in the 1970s and 1980s found that the executives based their decisions more on "the magnitudes of possible bad outcomes" than on their probability. James G. March & Zur Shapira, *Managerial Perspectives on Risk and Risk Taking*, 33 MGMT. SCI. 1404, 1407 (1987). Thus, for example, in one study, eighty percent of the executives "asked for estimates of the 'worst outcome' or the 'maximum loss'" when evaluating a possible course of action. *Id.* The focus on the magnitude rather than the probability of the worst-case scenario "leads to a propensity to accept greater risk (in the sense of variance) when the probability distribution of possible outcomes is relatively rectangular than where there are relatively long tails." *Id.* at 1411; see also GARBNER, *supra* note 96, at 71-72 (employing March and Shapira's findings to assess the likely effects of products liability exposure on industrial decisionmaking and concluding that "the possibility of extremely bad outcomes is particularly salient in the decision process"); cf. Sage, *supra* note 46, at 1004 (noting that "managerial risk aversion exists regardless of the availability of insurance").

for taking a hard look at its prior approval decision.

1. *Providing an Early Alert System*

When a *qui tam* relator filed a *parens patriae* suit, it would be required to serve the complaint and related information on the government. In addition to providing the government with an opportunity to take over the litigation, this notice would provide the FDA with a systematic source of information concerning potential problems that are ripening into litigation. In instances where a safety problem has been publicly discussed prior to suit, the filing would not provide the first indication of a problem; in other instances, as where the *qui tam* relator is a former employee who sues based on nonpublic information, the filing might provide the first concrete evidence of a safety issue. In either event, the filing would flag the problem as potentially significant, and it would alert the FDA to the need to monitor the litigation so that regulators could promptly assess information uncovered through discovery.

2. *Using Civil Discovery To Supplement Reporting*

Discovery in the *qui tam* litigation may uncover information that otherwise would not make its way to the FDA.³⁴³ A *qui tam* relator can use the civil discovery tools to obtain information that would not appear in reports submitted to the FDA.³⁴⁴ Though private tort suits may already uncover such information, the phenomenon of “secret settlements” may limit the extent to which information obtained in private suits reaches the FDA; in *qui tam* suits, by contrast, the FDA automatically would have access to the fruits of the discovery process.

Some critics have complained that the FDA’s information-gathering capabilities are largely passive.³⁴⁵ “[T]he FDA lacks the general subpoena power

343. Cf. Green, *supra* note 46, at 499 (noting that drug manufacturers’ reporting of adverse events “has been less than perfect,” and that “[s]ome notable examples of flagrant manufacturer disregard for [the reporting] requirement have been documented, sometimes as the result of a tort suit”).

344. Cf. McGarity, *supra* note 44, at 571 (“Private attorneys are adept at uncovering evidence of fraud and misrepresentation in the discovery that precedes common law trials, and they are willing to spend the resources necessary to copy and organize documents, take depositions, and fight the company’s efforts to resist discovery.”).

345. Cf. Rabin, *supra* note 12, at 2069 (“Even in the case of a comprehensive regulatory regime like FDA certification of new drugs, the agency process is noninvasive: the burden is on the company to produce evidence in support of its new drug application, and the agency does not conduct its own testing and experimentation.”).

that other agencies have, and therefore, in most instances cannot compel the disclosure of information about product risks.”³⁴⁶ Reporting requirements provide the FDA with basic information concerning adverse events (so long as companies comply with their reporting obligations). But discovery in a *qui tam* suit could shed light on problems that might not be as readily apparent in the summary reports. For example, a lawyer for the plaintiffs in a suit involving Paxil asserts that “court-ordered discovery allowed her to see raw data on safety and efficacy, while the FDA saw only the completed write-ups,” and that discovery also produced “the company’s internal communications about how to approach the agency, which the FDA never saw.”³⁴⁷

In addition to obtaining documents that would not be turned over in routine reporting to the FDA, a *qui tam* relator could follow up on promising avenues by deposing company employees. In the Bjork-Shiley heart valve litigation, for example, employee depositions revealed that workers “disguised cracks in defective valves,” and document discovery unearthed a plant supervisor’s memorandum that “complain[ed] of a company policy of disguising cracked valves as intact [and stated] ‘I feel we are hiding our most serious defect.’”³⁴⁸ Other employees may have equally pertinent information; for instance, sales representatives who are responsible for marketing a product to physicians may often have early warning of safety issues with the product.³⁴⁹

346. Teresa Moran Schwartz, *Prescription Products and the Proposed Restatement (Third)*, 61 TENN. L. REV. 1357, 1386 (1994). Schwartz notes that the FDA does have “power to demand documents where statutory provisions specifically provide, such as those governing factory inspections and new drug approvals,” and that companies may “disclose information voluntarily to the FDA to create good relations or to avoid an enforcement action.” *Id.* at 1386 n.177.

347. Gary Young, *FDA Strategy Would Pre-empt Tort Suits: Does It Close Off Vital Drug Data?*, NAT’L L.J., Mar. 1, 2004, at 1, 12 (discussing statements by Karen Barth Menzies); see also Gina Kolata, *Questions Raised on Ability of F.D.A. To Protect Public*, N.Y. TIMES, Jan. 26, 1992, § 1, at 1 (“In the case of Halcion, critics who have examined case report forms in connection with a lawsuit against Upjohn charge that the company left out information about adverse reactions reported on those forms when it prepared its data analyses for the F.D.A. Upjohn denies the charges . . .”). Likewise, discovery in the Vioxx litigation has apparently brought to light documents that bear upon Merck’s knowledge of safety problems with the drug. See Harris, *supra* note 70.

348. Gina Kolata, *Manufacturer of Faulty Heart Valve Barred Data on Dangers*, F.D.A. Says, N.Y. TIMES, Mar. 21, 1992, § 1, at 50.

349. See Kit R. Roane, *Replacement Parts: How the FDA Allows Faulty, and Sometimes Dangerous, Medical Devices onto the Market*, U.S. NEWS & WORLD REP., July 29, 2002, at 57, 59 (describing instances in which sales representatives became aware of concerns about product safety). “Direct contact between physicians (and other health-care professionals) and sales representatives of the companies is often the primary form of sales promotion,” though other methods include ads in medical publications, mailings to physicians, and direct-to-consumer ad

Parens patriae suits will also provide a more effective means of putting information uncovered during discovery into the hands of the FDA. When discovery in private civil suits yields information relevant to product safety, protective orders may sometimes prevent the plaintiffs' lawyers from sharing that information with the FDA.³⁵⁰ Defendants that settle such cases may be able to obtain a return of damaging discovery materials and a court order that the terms of settlement remain confidential.³⁵¹ Plaintiffs may accede to such secrecy provisions in return for a higher settlement payment.³⁵² Commentators have raised concerns that secrecy provisions may prevent the disclosure to the FDA of

campaigns. GARBER, *supra* note 96, at 21; *see also* ANGELL, *supra* note 47, at 127 ("Drug reps are allowed to attend medical conferences, may be invited into operating and procedure rooms, and sometimes are even present when physicians examine patients . . .").

350. *See* Kolata, *supra* note 348 (discussing protective order in case involving Bjork-Shiley heart valve and stating that company did not disclose to the FDA certain information, covered by the protective order, until the suit was dismissed); Kolata, *supra* note 347 ("[T]he data that caused the Commissioner . . . to ban [silicone breast] implants this month pending a review of their safety . . . were disclosed to trial lawyers eight years ago, but the [FDA] learned about them only recently because a court agreement had kept them confidential.").

351. *See* Joseph F. Anderson Jr., *Hidden from the Public by Order of the Court: The Case Against Government-Enforced Secrecy*, 55 S.C. L. REV. 711, 713-14 (2004) (describing aspects of secret settlements). As Judge Weinstein has explained, damaging discovery material can include:

"smoking gun" documents that indicate defendants knew of the danger but suppressed the information. Oral material obtained in depositions is also often highly useful to plaintiffs and devastating to defendants. Documents showing cover-ups or early knowledge by defendants of defects can lead to billions of dollars in punitive damages as well as extensive liability for ordinary damages, so there is strong reason for defendants to try to keep them secret.

Jack B. Weinstein, *Ethical Dilemmas in Mass Tort Litigation*, 88 NW. U. L. REV. 469, 512 (1994). Chief Judge Anderson notes that although the parties could reach a secrecy agreement without involving the court, defendants often "want the judge's signature, and the corresponding contempt power of the court, to legitimize their conduct and to have assurance that a violation will be summarily dealt with by the court." Anderson, *supra*, at 732.

Secret settlements are controversial. *Compare, e.g.,* Susan P. Koniak, *Are Agreements To Keep Secret Information Learned in Discovery Legal, Illegal, or Something in Between?*, 30 HOFSTRA L. REV. 783, 787 (2002) ("[A]ny legal regime that facilitates the keeping of secrets as lethal as the secrets Firestone was allowed to keep [concerning defective tires] may be a legal regime in need of serious repair. Certainly, the public is likely to feel that way . . ."), *with* Arthur R. Miller, *Confidentiality, Protective Orders, and Public Access to the Courts*, 105 HARV. L. REV. 427, 431-32 (1991) (arguing that "promoting increased public access to information by restricting the discretion of the courts to protect confidential information is ill-advised").

352. *See* Anderson, *supra* note 351, at 731 (noting statements by "some plaintiffs' lawyers . . . that court-ordered secrecy gives them the opportunity to leverage a little more money out of the defendant at settlement time").

information concerning product safety.³⁵³ In a *parens patriae* suit, by contrast, the FDA would have the right to review all information made available in discovery; the statutory framework could require that any protective order in the case provide the FDA's lawyers with the same status, under the protective order, as the lawyer for the *qui tam* relator.

3. Revisiting FDA Approval in Light of Later Information

In addition to uncovering or highlighting information on which the FDA may not previously have focused, the litigation process would provide a formal occasion for the FDA to revisit its prior safety assessments. A referral from the district court in a *qui tam* suit³⁵⁴ would prompt an advisory panel to evaluate the issues of safety and causation in the light of the information developed during discovery. The FDA would then be required to review the panel's findings and would be aided in its review both by the parties' presentation of the issues and by the panel's views. Obviously, the FDA could revisit its safety determinations in any event. But the *qui tam* litigation could enhance the record on which the FDA based its reevaluation and could provide added incentives for the FDA to take a harder look at a questionable product.³⁵⁵ In addition to answering the questions referred by the district court, the FDA would also have the opportunity to consider regulatory action concerning the product. The agency could require labeling changes, or—in extreme cases—direct the company to pull the product from the market.

CONCLUSION

In sum, the system described in Part III could offer benefits. In addition to providing expert agency views on questions of product safety and causation, it

353. See Dorothy J. Clarke, *Court Secrecy and the Food and Drug Administration: A Regulatory Alternative to Restricting Secrecy Orders in Product Liability Litigation Involving FDA-Regulated Products*, 49 FOOD & DRUG L.J. 109, 111 (1994) (advocating amendment of the FDCA "to require drug and device manufacturers to submit information to the FDA regarding product liability litigation and settlements").

354. Such referrals would not cause undue inconvenience for the FDA, because they would only occur in cases where the relator successfully resisted the defendant's summary judgment motion. Because summary judgment would be granted unless the relator showed either a violation of existing FDA requirements or the existence of material information that the FDA had not previously considered, the summary judgment stage would screen out claims that did not merit consideration by the FDA.

355. See GARBNER, *supra* note 96, at 36 (noting that "[n]ew information and publicity [generated by products liability suits] can generate substantial pressure on the FDA to reevaluate its previous decisions").

could shed new light on companies' true views of the tort system and could preserve litigation's role in generating information on product safety. Potentially, the system could broaden the availability of compensation for persons injured by a medical product manufactured by a company that had opted in.³⁵⁶

Would that system be preferable to the status quo? The answer depends largely on one's view of the civil jury's capabilities. If juries truly are irremediably ill-suited to the task of assessing product safety and causation, then the *qui tam* system would provide a benefit by sending those issues to the FDA. As I have noted, however, commentators are divided in their assessment of the jury's capacities.³⁵⁷ Moreover, it is possible to improve the performance of judges and juries within the current tort system³⁵⁸—an approach that holds the promise of addressing the critiques of the current tort system without embarking on radical systemic change.³⁵⁹

Equally important, the aim of obtaining expert agency resolution of products liability questions should be balanced against the risk of agency ineffectiveness or capture.³⁶⁰ It should be readily apparent from the discussion above that in

356. However, the details of the compensation model adopted for the opt-in system would be key: If the opt-in system simply tracked the remedies available under current tort doctrine (rather than providing broader compensation), the chance to obtain a quid pro quo from the companies that opted in would be squandered.

357. See *supra* notes 9-11 and accompanying text.

358. Measures such as crafting better jury instructions, providing those instructions before as well as after the presentation of evidence, permitting jurors to take notes and submit questions to witnesses, and permitting interim arguments by lawyers during a complex trial may improve juror comprehension and performance. The presentation of expert testimony could be improved in appropriate cases by directing opposing sides' experts to testify back-to-back or by including testimony from a nonpartisan expert. Judges could provide juries with better guidance on assessing noneconomic damages, and could engage in more searching review of awards of such damages. And improved judicial training could better enable judges to perform all these tasks.

359. Cf. Rabin, *supra* note 12, at 2067 (raising "the question whether . . . institutional reforms of the tort process offer promise of addressing satisfactorily the criticisms of those who would displace tort in cases where scientific evidence is in play and a regulatory agency has independently assessed the risks associated with a product").

360. Other costs should also be considered. The *qui tam* system would in effect require aggregate determination of all covered products liability claims, thus eliminating the ability of many claimants to control the presentation of their claims. (Though aggregation would occur in many instances under the existing system, it would not always be necessary.)

Because the *qui tam* mechanism would create a special procedure for liability claims concerning FDA-regulated medical products, a question of boundaries would arise: What should be done with non-products liability claims arising from the same set of facts? For example, many failure-to-warn cases may also include malpractice claims against a health care provider. See, e.g., *Marks v. Ohmeda, Inc.*, 871 So. 2d 1148, 1151, 1156 (La. Ct. App. 2004) (sustaining judgment

order safely to privilege the agency's views on safety and causation, it will be necessary to ensure that those views are accurate and free of improper bias.³⁶¹ Recent events have underscored the difficulties with the FDA's current postmarketing surveillance system: Resource constraints, and possibly an unwillingness to question prior determinations of product safety, have impaired the FDA's ability to respond to emerging problems. Those difficulties would also plague any system that attempted to rely on the FDA to resolve retrospective liability questions. Proposals for the creation of an independent postmarketing surveillance agency might help to address this issue: If Congress were to create and adequately fund such an agency, and protect it from pressure by the FDA and by stakeholders, the independent agency might be able both to monitor product safety and to resolve appropriately safety and causation issues referred to it by a court. If Congress does not create an independent safety monitor, however, recent experience provides strong reason to question the wisdom of giving the FDA (as currently structured and funded) and its advisory panels (as currently staffed) a dispositive role in products liability actions.

This Article, accordingly, has failed to demonstrate that the hybrid system

against manufacturer of anesthetic and anesthesia machine, in failure-to-warn case that also involved malpractice claims against hospital and nurse anesthetist); *see also* RISK MANAGEMENT REPORT, *supra* note 23, at 26 (noting that "[s]ubstantial numbers of injuries and deaths occur annually" due to "incorrect administration of the prescribed product or incorrect operation or placement of a medical device"). Such malpractice claims generally would raise issues specific to the particular claimant and physician, and would be unsuitable for resolution within the *qui tam* system (which would focus on aggregate determination of the products liability issues). Thus, any benefits of the new system would be offset to some degree by the costs of parallel litigation.

361. Richard Nagareda has suggested that one advantage of referral to the agency is that the agency—unlike a jury—is politically accountable for its decisions. *See* Nagareda, *supra* note 15, at 299 ("Although agencies have long been considered repositories of technical expertise, commentators have neglected an equally powerful justification for agency involvement in the mass tort area: the political accountability of such bodies."). As he argues, "[t]he conditions of scientific uncertainty that plague the handling of [mass torts] within the tort system cry out for the application of political judgment and deliberation through administrative channels in a manner susceptible to public scrutiny." *Id.* at 313. The downside of agency accountability, however, is the potential for bias in favor of the regulated industry. Nagareda notes that "the highly concentrated interests typified by regulated industries . . . may be better positioned to sway a single regulatory agency at the federal level than to exert influence over the multitude of courts within which lawsuits would proceed in the tort system." *Id.* at 364. He argues, however, that in the context of a referral to the agency of issues raised in mass tort litigation, the high-profile nature of the dispute and the existence of injured victims could counterbalance any tilt in favor of industry. *See id.* ("Agency action in the aftermath of thousands of individual tort suits—so many as to call for consolidation of litigation within the federal courts by the MDL Panel—is far less susceptible to influence by industry.").

considered here is preferable to the status quo (or to less drastic options for reform). But it has shown that the *qui tam* system would be preferable to preemption. By privileging FDA determinations on safety and causation, the *qui tam* system would address preemption advocates' central criticisms of the current tort system. But unlike preemption, the *qui tam* system would address those criticisms while preserving some compensatory and monitoring role for litigation. Where preemption advocates tend to accept uncritically the industry's contention that the specter of tort liability chills innovation, the *qui tam* system would provide a means for measuring that contention against companies' actual choices concerning the opt-in. And where preemption would remove entirely the role of the states in regulating product safety, the *qui tam* proposal would displace that role only in instances where the manufacturer opted in to the *qui tam* system with respect to the relevant product.

Thus, this Article establishes that advocates of preemption who cast the debate in binary terms have failed to address the full range of options. Those advocates should be required to carry the burden of demonstrating the need for change, but they also should be required to show that, if change is warranted, preemption is the best choice. As this Article illustrates, a range of options short of preemption would address the asserted defects in the tort system without eliminating the ability to hold companies responsible for harm caused by safety problems with medical products.

NOTE

Juvenile Mental Health Courts and Therapeutic Jurisprudence: Facing the Challenges Posed by Youth with Mental Disabilities in the Juvenile Justice System

Patrick Geary*

INTRODUCTION

Ever-increasing numbers of children struggle to live and develop under the burden of mental disability.¹ Yet the juvenile justice system—an institution created in large part to look after these very children—has often failed to meet, address, or fully realize their mental health needs.² As children's mental health issues have entered the spotlight in recent years, the juvenile court's gross inadequacy as a guardian of child development and gatekeeper of treatment services has become clearer. Indeed, many have concluded that "the inadequate and uneven delivery of mental health services to children and families in the juvenile justice system is a national crisis."³

The ideas behind the juvenile mental health court movement, however, may

* J.D. candidate, Yale Law School.

1. See W. John Thomas et al., *Race, Juvenile Justice, and Mental Health: New Dimensions in Measuring Pervasive Bias*, 89 J. CRIM. L. & CRIMINOLOGY 615, 627 (1999); Lois A. Weithorn, Note, *Mental Hospitalization of Troublesome Youth: An Analysis of Skyrocketing Admission Rates*, 40 STAN. L. REV. 773, 773 (1988); U.S. DEP'T OF HEALTH & HUMAN SERVS., REPORT OF THE SURGEON GENERAL'S CONFERENCE ON CHILDREN'S MENTAL HEALTH: A NATIONAL ACTION AGENDA (2000), <http://www.hhs.gov/surgeongeneral/topics/cmh/cmhreport.pdf> [hereinafter CONFERENCE].

2. See CONFERENCE, *supra* note 1.

3. NAT'L COUNCIL OF JUVENILE & FAMILY COURT JUDGES, ENHANCING THE MENTAL HEALTH AND WELL BEING OF INFANTS, CHILDREN AND YOUTH IN THE JUVENILE AND FAMILY COURTS: A JUDICIAL CHALLENGE 1 (2000).

offer the juvenile justice system new hope. This movement advocates the creation of separate juvenile courts for youth with pronounced mental health needs and brings renewed attention to the rehabilitative goals of the juvenile justice system. While the aims of the juvenile mental health court movement are laudable, its greatest influence may lie beyond the formation of specialized juvenile courts that serve only a limited number of youthful offenders. Raising awareness of mental health needs in the broader juvenile justice system presents a wider opportunity to improve the treatment of youth with mental disabilities in all juvenile courts.

As outlined in Part I of this Note, the progressive foundation of the juvenile court serves as a uniquely appropriate base from which to address the needs of youth suffering from mental disabilities. Part II highlights the diverse and expanding nature of these needs and outlines the scope of the issues facing the juvenile justice system today. Part III follows the proliferation of specialty "problem-solving" courts in the adult and juvenile justice systems and describes the principles of therapeutic jurisprudence that encourage increased sensitivity to youth's mental health needs in courtroom procedures. Given the extraordinary prevalence of mental health needs among youthful offenders, Part IV suggests that it would be impractical to address these needs through smaller, specialized courts and argues that it would be better to apply the approach adopted in juvenile mental health courts throughout the entire juvenile court system. The details of potential mental health oriented reforms are described in Part V, and both existing and potential funding initiatives designed to support these reforms are discussed in Part VI.

Dealing with youthful offenders plagued by mental disabilities will always be difficult. The juvenile justice system may not be the ideal place to address these children's mental health needs, but we should not overlook the contributions that it can make. Rehabilitative treatment remains a fundamental tenet of the juvenile court, and youthful offenders must not be denied access to mental health services in the name of retribution or inadequate funding. The allocation of additional resources to juvenile mental health needs today would not only fulfill the original mandate of the juvenile court to provide treatment, rather than punishment; it would also save society money in the long run by reducing the need to expend resources on these juveniles later in their lives.⁴ The time for juvenile justice reform is now, and the present support for juvenile mental health courts demonstrates a public and judicial readiness to recognize the importance of mental health concerns and rediscover individualized treatment in

4. Gail B. Nayowith, *A Window of Opportunity for Children Who Stay Too Long*, in CHILDREN'S LAW INSTITUTE 2000, at 355 (PLI Litig. & Admin. Practice Series, Criminal Law & Urban Problems Course Handbook Series No. C-185, 2000).

juvenile court dispositions. This rediscovery may be just the answer for a juvenile court under fire. By embracing young offenders and their mental health needs, the flailing juvenile justice system could complete a return to its legitimate roots in the rehabilitative ideal.

I. A BRIEF HISTORY OF JUVENILE JUSTICE

The first juvenile court opened its doors to wayward children just over a century ago.⁵ Under the state's *parens patriae* power, the juvenile court had wide discretion to "rescue" young offenders⁶ and further the "best interests" of these children.⁷ The leaders of the juvenile justice revolution saw youth as developmentally sensitive and largely amenable to intervention and treatment and, accordingly, rejected the adult system's emphasis on accountability and culpability.⁸ Progressive criminal justice reformers sought to create "a space to protect, to rehabilitate and to heal children, a site of nurturance and guidance, understanding and compassion."⁹ A separate juvenile justice system provided the opportunity to eliminate the harms of contact with the adult criminal courts and to improve offenders' well-being.¹⁰ Juvenile court dispositions focused on the child's need for specialized treatment, rather than her culpability. Informal, paternalistic, and non-adversarial courtroom procedures facilitated expedient

5. Illinois Juvenile Court Act of 1899, 1899 Ill. Laws 131. The Act "marked the end of a penal approach to juvenile delinquency and the beginning of what was perceived as a preventative approach." Janet Gilbert et al., *Applying Therapeutic Principles to a Family-Focused Juvenile Justice Model (Delinquency)*, 52 ALA. L. REV. 1153, 1159 (2001).

6. Franklin E. Zimring, *The Common Thread: Diversion in Juvenile Justice*, 88 CAL. L. REV. 2477, 2480 (2000).

7. Barry C. Feld, *The Juvenile Court Meets the Principle of Offense: Punishment, Treatment, and the Difference It Makes*, 68 B.U. L. REV. 821, 821 (1988); see also BARRY NURCOMBE & DAVID F. PARTLETT, *CHILD MENTAL HEALTH AND THE LAW* 274 (1994) ("The state's assertion of its *parens patriae* power was based upon the morally benevolent premise that youth do not deserve to be punished for violations of the law."); Shay Bilchik, *Juvenile Justice: A Century of Change*, 1999 NAT'L REP. SERIES: JUV. JUST. BULL. 1.

8. Bilchik, *supra* note 7, at 1; Feld, *supra* note 7, at 833.

9. David C. Anderson, *When Should Kids Go to Jail?*, AM. PROSPECT, May-June 1998, at 72, 72-73 (citation omitted). Indeed, "the role of the juvenile court judge was to strengthen the child's belief in himself and make available to him all of the support and encouragement from outside the court that the judge could harness on his behalf." *Id.* at 73 (citation omitted).

10. Bilchik, *supra* note 7, at 1; Feld, *supra* note 7, at 833; Rhonda Gay Hartman, *Adolescent Autonomy: Clarifying an Ageless Conundrum*, 51 HASTINGS L.J. 1265, 1274 (2000) (noting that the juvenile court meant not to stigmatize and later reform children, but to shield them from stigma altogether).

delivery of the best-suited rehabilitative services.¹¹

Unfortunately, the traditional juvenile court failed to maintain its rehabilitative aspirations. The courts had been given extensive judicial discretion to tailor proceedings to the needs of individual offenders, but by the 1960s this discretion was often abused.¹² In the Supreme Court's landmark decision *In re Gault*,¹³ the Court scaled back the juvenile court's dispositive and procedural flexibility. Responding to the arbitrariness of juvenile court dispositions and "an absence of the rehabilitation that the system had promised,"¹⁴ the Court granted juvenile defendants certain safeguards available in the more formalized adult criminal courts.¹⁵ In the years following *Gault*, the increasing procedural convergence of juvenile and adult criminal courts began to erode the juvenile court's focus on rehabilitation.¹⁶ The juvenile justice system's trademark individualized treatment plans gave way to a focus on young offenders' culpability.¹⁷ These changes were reinforced by public dissatisfaction with the perceived leniency of the traditional juvenile court, a sentiment fueled by rising juvenile crime rates and a growing public fear of adolescent criminality.¹⁸

Public support for a separate justice system for children continues to wane today, and as many jurisdictions begin "to shift more resources into monitoring

11. Loren M. Warboys & Shannan Wilber, *Mental Health Issues in Juvenile Justice*, in LAW, MENTAL HEALTH, AND MENTAL DISORDER 503 (Bruce D. Sales & Daniel W. Shuman eds., 1996).

12. Melissa Moon et al., *Is Child Saving Dead? Public Support for Juvenile Rehabilitation*, 46 CRIME & DELINQ. 38, 39 (2000); see also FRANKLIN E. ZIMRING, THE CHANGING LEGAL WORLD OF ADOLESCENCE 128 (1982) (noting that "[t]he tyranny of unguided discretion is why we have retreated from the lawlessness of the original juvenile court"); N. Dickon Reppucci, *Adolescent Development and Juvenile Justice*, 27 AM. J. COMMUNITY PSYCHOL. 307, 313 (1999).

13. 387 U.S. 1 (1967).

14. Thomas Grisso, *Why Juvenile Justice Will Survive Its Centennial*, in THE EVOLUTION OF MENTAL HEALTH LAW 169 (Lynda E. Frost & Richard J. Bonnie eds., 2001).

15. The *Gault* Court held that juveniles have the right to notice of charges, a fair and impartial hearing, assistance of counsel, and to protection against self-incrimination. 387 U.S. 1.

16. Barry C. Feld, *The Transformation of the Juvenile Court*, 75 MINN. L. REV. 691, 691-92 (1991). Following *Gault*, the court increasingly found procedural safeguards necessary in the juvenile court. See, e.g., *Breed v. Jones*, 421 U.S. 519 (1975) (extending the protections of the Double Jeopardy Clause of the Fifth Amendment to the juvenile justice system); *In re Winship*, 397 U.S. 358 (1970) (setting the standard of proof for delinquency as "beyond a reasonable doubt"). But see *McKeiver v. Pennsylvania*, 403 U.S. 528 (1971) (denying the right to a jury trial in a state delinquency proceeding).

17. Feld, *supra* note 7, at 830; see also Thomas et al., *supra* note 1, at 621.

18. Bilchik, *supra* note 7, at 3; see also ZIMRING, *supra* note 12, at 1-8 (1998). Zimring notes both that "[j]uvenile violence in the United States is frequently depicted as a difficult current problem that will inevitably get worse," *id.* at 4, and that "lenient treatment by the juvenile justice system [is seen as] a major cause of high rates of youth crime," *id.* at 7.

and incarcerating the most serious juvenile offenders for longer periods of time[,] . . . [f]ewer resources are left to deal with . . . those youth most amenable to rehabilitation.”¹⁹ Legislatures—responding to the calls to “crack down” on juvenile crime—brought punitive reforms to the juvenile courts through much of the 1970s, 1980s, and 1990s.²⁰ Statutes enabling juvenile transfer to adult court,²¹ mandatory minimum sentences,²² and reduced confidentiality provisions²³ have continued to move the juvenile court farther away from the rehabilitative ideal.²⁴ Some scholars have even called for the abolishment of the juvenile justice system altogether.²⁵

19. Edward Humes, A Brief History of Juvenile Court, at <http://www.edwardhumes.com/articles/juvhist.shtml> (last visited Apr. 8, 2003); see also Elizabeth S. Scott & Thomas Grisso, *The Evolution of Adolescence: A Developmental Perspective on Juvenile Justice Reform*, 88 J. CRIM. L. & CRIMINOLOGY 137, 147 (1997) (finding that “[t]he emphasis today is on social control—on protecting society from the harms inflicted by youth offenders—and the clear trend has been toward imposing penalties on adolescents (especially those who commit violent crimes) that approximate sanctions imposed on adults”); Francine T. Sherman & William Talley, Jr., *Foreword to Symposium, Struggling for a Future: Juvenile Violence, Juvenile Justice*, 36 B.C. L. REV. 889, 893 (1995) (noting that juvenile justice policy has effectively created “[a] war on children”). The public appears to favor this move toward a more offense-based system. See T. Hart, *Causes and Consequences of Juvenile Crime and Violence: Public Attitudes and Question-Order Effect*, 23 AM. J. CRIM. JUST. 129 (1998).

20. Anderson, *supra* note 9, at 73-74.

21. Indeed, prosecutorial discretion (concurrent jurisdiction over serious offenders in both the adult and juvenile courts), legislative offense exclusion (exclusive criminal court jurisdiction for serious offenses committed by juveniles of a certain age), and judicial waiver (automatic, presumptive, or discretionary transfer of juveniles into adult criminal court) have all but stripped today’s juvenile courts of their broad jurisdiction. *Id.* at 74; see also Feld, *supra* note 16, at 701-08; Grisso, *supra* note 14, at 173 (“For judicial transfer to criminal court, offense-based criteria were broadened, age-based criteria were lowered, transfer hearings were mandated, relevant criteria were expanded, burdens of proof were shifted to the defense, and standards of proof for transfer were reduced.”); Earl F. Martin & Marsha Kline Pruett, *The Juvenile Sex Offender and the Juvenile Justice System*, 35 AM. CRIM. L. REV. 279, 326-27 (1998).

22. Feld, *supra* note 7, at 717; Grisso, *supra* note 14, at 171.

23. Bilchik, *supra* note 7, at 5.

24. Over a quarter of the states have tolled the death knell for the rehabilitative ideal by amending juvenile court purpose clauses to include language “emphasizing offender accountability, public safety, and competency development.” Bilchik, *supra* note 7, at 3; see also Feld, *supra* note 16, at 709 (“These amendments de-emphasize rehabilitation and the child’s ‘best interests,’ and emphasize the importance of protecting public safety, enforcing children’s obligations to society, applying sanctions consistent with the seriousness of the offense, and rendering appropriate punishment to offenders.”).

25. Feld, *supra* note 16, at 723-24. Feld favors “an adult criminal court that administers justice for young offenders[,] . . . provid[ing] children with all the procedural safeguards already available

Any consideration of children's mental health needs in the contemporary juvenile justice system must occur against the background of these uneasy circumstances. At the inception of the juvenile court, the focus on rehabilitative treatment programs for individual children led founders to look beyond young offenders' delinquent acts.²⁶ Juvenile court judges examined all probable causes of delinquency, and the mental health needs of the children before them figured prominently in their decision-making.²⁷ While today's juvenile courts operate on dramatically different terms than did their century-old predecessors, the juvenile justice system continues to face the challenges associated with handling youthful offenders with mental disabilities. Even if these challenges must ultimately be resolved in the adult criminal justice system, "we will always need a special legal mechanism to respond to children in need of services"²⁸

II. THE SCOPE OF THE PROBLEM: MENTAL DISABILITY IN THE JUVENILE JUSTICE SYSTEM

Despite the early juvenile courts' focus on rehabilitative treatment programs, these courts were strongly criticized for the way in which they handled youths with mental disabilities.²⁹ Surprisingly little has changed. In the past century, the mental health field has made dramatic advances,³⁰ yet many of the mental health problems of young offenders in today's juvenile courts remain undiagnosed and

to adult defendants and additional enhanced protections because of the children's vulnerability and immaturity." *Id.* Although no states to date have elected to merge the juvenile and adult courts into a unitary criminal justice system, such a move is certainly not beyond the realm of possibility: "The legal response to juvenile crime is undergoing revolutionary change, and its ultimate shape is uncertain." Scott & Grisso, *supra* note 19, at 137.

26. Monrad G. Paulsen, *Children's Court: Gateway or Last Resort?*, 10 COLUM. U.F. 4 (1967). Court dispositions were fashioned to poison the roots of delinquency and thereby foster maturity into productive adulthood by encouraging youths' continuing stability. *Id.*

27. The juvenile court often invoked progressive guidance to mandate that these needs be met through ordering and applying "[p]sychological techniques . . . to the mentally disturbed." *Id.* at 5.

28. Grisso, *supra* note 14, at 174.

29. See Thomas et al., *supra* note 1, at 616 ("Many of these Juvenile Offenders need the services of a good physician more than they do those of the jailor." (quoting WILLIAM MACDONALD, A STORY OF JUVENILE COURTS FROM THEIR INCEPTION TO THE PRESENT DAY, WITH COMMENTS UPON THE EXTENSION OF THE PROBATION SYSTEM AND A HISTORY OF THE JUVENILE COURT MOVEMENT 27 (1912))).

30. COALITION FOR JUV. JUST., HANDLE WITH CARE: SERVING THE MENTAL HEALTH NEEDS OF YOUNG OFFENDERS, 2000 ANNUAL REPORT 40 (2000) [hereinafter HANDLE WITH CARE] (remarking that we continue to move "towards a clearer and more sophisticated understanding of the underlying causes [of delinquency]").

untreated.³¹ Concern with the increasingly punitive nature of the juvenile justice system has prompted many juvenile and mental health professionals to decry what they view as an insufficient emphasis on treatment and rehabilitation.³² Critics—although substantially ignored in the present juvenile justice system—maintain that there is a strong nexus between delinquency, mental illness, and the overall need for mental health treatment to prevent recidivism.³³

Youths in contact with the juvenile justice system are significantly more likely than other youths to have mental disabilities.³⁴ The juvenile justice system has in some ways become a “‘dumping ground’ for mentally ill, learning disabled, [and] behaviorally disordered juveniles. Many juvenile offenders have a history of involvement with the mental health system but migrate to the juvenile justice system because the mental health system has failed to serve their needs.”³⁵ Although many delinquents are deemed simply socially maladjusted by the juvenile justice system, a considerable portion of these children have serious, diagnosable emotional disturbances.³⁶ While estimates of these disturbances in the general population of children and adolescents range from two to seven percent, estimates for the delinquent population range from sixteen to fifty

31. Richard E. Redding, *Barriers to Meeting the Mental Health Needs of Offenders in the Juvenile Justice System* (2000), at http://www.ilppp.virginia.edu/Publications_and_Reports/BarrMeet.html; see also JANE KNITZER, *UNCLAIMED CHILDREN: THE FAILURE OF PUBLIC RESPONSIBILITY TO CHILDREN AND ADOLESCENTS IN NEED OF MENTAL HEALTH SERVICES* (1982) (noting that children in the juvenile justice system with mental disorders are largely neglected and ignored).

32. Redding, *supra* note 31.

33. Tamera Wong, Comment, *Adolescent Minds, Adult Crimes: Assessing a Juvenile's Mental Health and Capacity To Stand Trial*, 6 U.C. DAVIS J. JUV. L. & POL'Y 163, 165 (2002).

34. More specifically, Scott and Grisso list emotional disturbances and learning and attention deficit disorders among those disabilities more prevalent in the delinquent population. Scott & Grisso, *supra* note 19, at 169.

35. Redding, *supra* note 31.

36. Warboys & Wilber, *supra* note 11, at 506 (citing D.M. Murphy, *Prevalence of Handicapping Conditions Among Juvenile Delinquents*, 7 REMEDIAL & SPECIAL EDUC. 7 (1986)); Nat'l Mental Health Ass'n, *Children with Emotional Disorders in the Juvenile Justice System* (1998), at <http://www.nmha.org/position/ps37.cfm>. As a report by the Coalition for Juvenile Justice explains, while laws vary among states:

[A] youth is generally considered to be seriously emotionally disturbed when:
Emotional and/or social impairment disrupts his or her academic and/or developmental progress; [such impairment] [d]isrupts family and/or other interpersonal relationships;
[s]uch impairment of functioning has continued for a period of at least one year; [o]r
such impairment is of short duration and high severity.

HANDLE WITH CARE, *supra* note 30, at 8-9.

percent.³⁷ Among delinquent youth, between one and six percent suffer from psychotic disorders,³⁸ and at least twenty percent are estimated to suffer from serious mental disorders generally (including schizophrenia, major depression, and bipolar disorder).³⁹ In addition, fifty-five percent of youth in the juvenile justice system show symptoms of clinical depression, and up to nineteen percent of youth may be suicidal.⁴⁰

Overall, the prevalence of psychiatric disorders among those detained in the juvenile justice system is between fifty and seventy-five percent.⁴¹ Put simply, “a far greater proportion of children in the juvenile justice system suffer from a serious emotional disturbance than in the general population.”⁴² If not only serious emotional disturbances but also other mental disabilities—attention deficit disorder, attention deficit hyperactivity disorder, substance abuse and dependence, learning disabilities, mental retardation, anxiety disorders, and conduct disorders—are considered, an even higher proportion of children before the juvenile court present substantial mental health needs.⁴³ For many of these less severe conditions, the estimated prevalence among youthful offenders exceeds eighty percent.⁴⁴ Among this population, the number of children with

37. NAT'L MENTAL HEALTH ASS'N, ALL SYSTEMS FAILURE (1993); Warboys & Wilber, *supra* note 11, at 506 (citing N.A. Brandenburg et al., *The Epidemiology of Childhood Psychiatric Disorders: Recent Prevalence Findings and Methodological Issues*, 29 J. AM. ACAD. CHILD & ADOLESCENT PSYCHIATRY 76 (1990)); see also Dana Royce Baerger et al., *Responding to Juvenile Delinquency: Mental Health Service Needs of Male and Female Juvenile Detainees*, 3 J. CENTER FOR FAM. CHILD. & CTS. 21, 21 (2001) (finding that more than one-third of adolescents arrested and adjudicated within the juvenile justice system exhibit symptoms of major affective disorders).

38. Randy K. Otto et al., *Prevalence of Mental Disorders Among Youth in the Juvenile Justice System*, in NAT'L COALITION FOR THE MENTALLY ILL IN THE CRIM. JUST. SYS., *RESPONDING TO THE MENTAL HEALTH NEEDS OF YOUTH IN THE JUVENILE JUSTICE SYSTEM* 21 (Joseph J. Cocozza ed., 1992).

39. Joseph J. Cocozza & Kathleen Skowrya, *Youth with Mental Health Disorders: Issues and Emerging Responses*, 7 JUV. JUST. 3, 6 (2000).

40. HANDLE WITH CARE, *supra* note 30, at 10-11.

41. Nat'l Mental Health Ass'n, *Prevalence of Mental Disorders Among Children in the Juvenile Justice System*, at <http://www.nmha.org/children/justjuv/prevalence.cfm> (last visited Mar. 24, 2005). The prevalence breaks down further along gender lines, with somewhere near two-thirds of males and three-fourths of females presenting psychiatric disorders. Linda A. Teplin et al., *Psychiatric Disorders in Youth in Juvenile Detention*, 59 ARCHIVES GEN. PSYCHIATRY 1133, 1137 (2002).

42. Warboys & Wilber, *supra* note 11, at 506.

43. Otto et al., *supra* note 39. Roughly half of the youth in contact with the juvenile system have conduct disorders, and up to forty-five percent have attention deficit hyperactivity disorder. HANDLE WITH CARE, *supra* note 30, at 11.

44. Daniel P. Mears et al., *Critical Challenges in Addressing the Mental Health Needs of*

multiple diagnoses is substantial,⁴⁵ and at least half of adolescents with mental illnesses in the juvenile justice system have co-occurring substance abuse disorders.⁴⁶

Moreover, the rates of admission to mental health facilities for juveniles *not* in contact with the court system have rapidly escalated over the past several decades as well.⁴⁷ More youth today have experienced child abuse or neglect, family dysfunction, or a host of other factors that might call for mental health intervention.⁴⁸ Not surprisingly, then, the overall prevalence of psychosocial problems among youth seem to be increasing,⁴⁹ and the demand for mental health services by some estimates nearly doubles each year.⁵⁰ As these troubled youth begin to make contact with the juvenile justice system, it will become increasingly clear that this system must find a way to address these youths' mental health needs in an appropriate manner.

III. THERAPEUTIC JURISPRUDENCE AND PROBLEM-SOLVING COURTS

Therapeutic jurisprudence—defined by one scholar as “the use of social science to study the extent to which a legal rule or practice promotes the psychological and physical well-being of the people it affects”⁵¹—offers a publicly acceptable vehicle for juvenile justice reform. This approach sits at the

Juvenile Offenders, 1 JUST. POL’Y J. 41 (2001), at http://www.childrensprogram.org/media/word/the_justice_policy_journal.doc.

45. HANDLE WITH CARE, *supra* note 30, at 11.

46. *Id.*

47. The rising number of mental disabilities in children and adolescents may be one of many factors accounting for this rise. Thomas et al., *supra* note 1, at 627; *see also* Weithorn, *supra* note 1, at 773-75 (noting that certain juvenile justice reforms and a general increase in inappropriate or unnecessary juvenile hospitalizations may be responsible for the rise in admission rates at juvenile mental health facilities).

48. Warboys & Wilber, *supra* note 11, at 507. As many as “seventy-five percent of violent juvenile offenders suffered severe abuse by a family member, eighty percent witnessed physical violence from beatings and killings, fifty percent were raised in one parent households, and over twenty-five percent had a parent who abused drugs or alcohol.” Joshua T. Rose, *Innocence Lost: The Detrimental Effect of Automatic Waiver Statutes on Juvenile Justice*, 41 BRANDEIS L.J. 977, 986 (2003).

49. *See* CONFERENCE, *supra* note 1.

50. Bree Langemo, Comment, *Serious Consequences for Serious Juvenile Offenders: Do Juveniles Belong in Adult Court?*, 30 OHIO N.U. L. REV. 141, 157 (2004).

51. Peggy Fulton Hora et al., *Therapeutic Jurisprudence and the Drug Treatment Court Movement: Revolutionizing the Criminal Justice System’s Response to Drug Abuse and Crime in America*, 74 NOTRE DAME L. REV. 439, 443 (quoting Christopher Slobogin, *Therapeutic Jurisprudence: Five Dilemmas To Ponder*, 1 PSYCHOL. PUB. POL’Y & L. 193, 196 (1995)).

nexus of mental health and law, and its adherents look optimistically for opportunities to apply recent developments in the clinical behavioral sciences in the legal field.⁵² Therapeutic jurisprudence principles emphasize the ways in which legal rules and processes may further the psychological health and emotional well-being of those in contact with the justice system. Therapeutic jurisprudence models examine the role of law as a therapeutic agent, with a mental health-focused approach to the law compatible with existing legal values.⁵³ While in many circumstances other legal considerations may trump therapeutic ones,⁵⁴ therapeutic jurisprudence ideals nonetheless promise innovation and improvement in the legal system's response to mental health concerns.⁵⁵

The therapeutic jurisprudence movement is a product of a growing impetus for change in the U.S. justice system's approach to the complex problems presented by "defendants whose substance abuse or mental disabilities appear to be related inextricably to repeated criminal [or delinquent] behavior."⁵⁶ Therapeutic jurisprudence recognizes that the courts are not manned by mental health professionals but hopes to encourage the courts to be sensitive to mental health issues: "It is unrealistic to suggest that lawmakers should be social scientists. Rather, law-makers, particularly judges, should be asked to take account of social science."⁵⁷

The therapeutic jurisprudence movement continues to mature, and its principles have already influenced the development of specialized "treatment courts" and the juvenile justice system's goals, as discussed below. In fact, therapeutic jurisprudence represents a theoretical basis for the entire "treatment court" movement and once served as the cornerstone of the juvenile justice

52. David B. Wexler, *The Development of Therapeutic Jurisprudence from Theory to Practice*, in *THE EVOLUTION OF MENTAL HEALTH LAW* 282 (Lynda E. Frost & Richard J. Bonnie eds., 2001).

53. David B. Wexler, *An Orientation to Therapeutic Jurisprudence*, 20 *NEW ENG. J. ON CRIM. & CIV. CONFINEMENT* 259, 259 (1994).

54. Among these other legal considerations, scholars have singled out the protection of defendant's rights, the protection of societal interests, and the enhancement of daily procedural interests in the legal system. David Finkelman & Thomas Grisso, *Therapeutic Jurisprudence: From Idea to Application*, 20 *NEW ENG. J. ON CRIM. & CIV. CONFINEMENT* 243, 249 (1994); Wexler, *supra* note 53, at 259-60.

55. Finkelman & Grisso, *supra* note 54, at 248 ("By applying psychological research and theory to mental health law in particular, therapeutic jurisprudence promises to reinvigorate the area and, if successful, to produce better mental health law, and better treatment for those who find themselves involved in the mental health law system.").

56. Teresa W. Carns et al., *Therapeutic Justice in Alaska's Courts*, 19 *ALASKA L. REV.* 1, 2 (2002).

57. NURCOMBE & PARTLETT, *supra* note 7, at 9.

system.⁵⁸ The formation of specialty “problem-solving” or “treatment” courts to better address specific categorical concerns and common needs of certain types of offenders is perhaps the best example of the application of therapeutic jurisprudence concepts in the justice system.⁵⁹ These courts have emerged in both the criminal and juvenile justice systems and demonstrate an institutional capacity to address the substance abuse and mental health needs of offenders.

A. “Problem-Solving” Courts in the Criminal Justice System

Specialized therapeutic courts handle a wide array of issues, ranging from family problems and domestic violence to substance abuse and mental health concerns.⁶⁰ These “holistic” courts draw together the efforts of legal and mental health professionals to fashion treatment plans and supervision models.⁶¹ Judges use innovative procedures to facilitate creative solutions to the issues presented by each individual offender. These judges are given the freedom to set aside the paternalistic leanings associated with their traditional role in the criminal justice system and demonstrate heightened respect for the dignity and autonomy of offenders.⁶² They may employ persuasive techniques to encourage defendants to complete treatment plans in the hopes of increasing compliance with programs tailored to ensure that individuals will avoid the justice system in the future.⁶³

While some have criticized the coercive and paternalistic potential inherent in specialized therapeutic courts, advocates of these courts tout as benefits the reduced recidivism rates and the greater likelihood that defendants will return to their communities as productive individuals.⁶⁴ The Conference of Chief Justices, the Conference of State Court Administrators, and the American Bar Association have all expressed support for the maintenance and formation of specialized

58. See Gene Griffin & Michael J. Jenuwine, *Using Therapeutic Jurisprudence To Bridge the Juvenile Justice and Mental Health Systems*, 71 U. CIN. L. REV. 65, 67 (2002) (“Juvenile courts, by their very nature, were designed to be more therapeutic than the adult criminal justice system.”).

59. Carns et al., *supra* note 56, at 5. Therapeutic jurisprudence offers problem-solving courts the tools to improve interpersonal skills, respect autonomy by avoiding paternalism, use persuasion, spark motivation, and increase compliance. Bruce J. Winick, *Therapeutic Jurisprudence and Problem Solving Courts*, 30 FORDHAM URB. L.J. 1055 (2003).

60. Winick, *supra* note 59, at 1068-97.

61. Carol Kessler, *Alternatives to Adjudication: Drug Courts, Mental Health Courts, Peer Courts*, in AM. ACAD. OF CHILD & ADOLESCENT PSYCHIATRY, RECOMMENDATIONS FOR JUVENILE JUSTICE REFORM 62, 62 (2001).

62. Winick, *supra* note 59, at 1066-78.

63. *Id.* at 1079-89.

64. Carns et al., *supra* note 56, at 54.

therapeutic courts.⁶⁵ Despite frequent concerns about the resources required to establish and maintain therapeutic courts,⁶⁶ the size, number, and diversity of these courts continues to grow.⁶⁷ While the oldest and most prominent specialized therapeutic treatment courts were developed primarily to handle the problems of substance abuse, a much younger mental health court movement has now emerged and expanded in step with the growing understanding of therapeutic jurisprudence and the mental health needs of offenders.

The first drug treatment court began operation the summer of 1989 in Miami, Florida.⁶⁸ In only fifteen years time, drug treatment courts have proliferated and now “apply the concepts of therapeutic jurisprudence . . . in hundreds of courtrooms across America.”⁶⁹ These drug treatment courts do more than simply expedite the judicial process in courts with crowded dockets; they seek to address the “underlying problems of drug crimes—drug use and addiction.”⁷⁰ By treating substance abuse not as a criminal failing but as a physiological condition requiring therapeutic intervention, drug treatment courts shift their orientation away from the retributive aims of the general criminal justice system.⁷¹

For drug offenders to be eligible for drug treatment courts, community officials must determine that these defendants have a substantial chance at recovery and pose a minimal threat to public safety.⁷² For these offenders, drug treatment courts use a system that is cooperative, rather than adversarial, and focus on promoting recovery through coordinated response.⁷³ Using a therapeutic lens, drug treatment courts look at offenders as clients and at potential relapses or other obstacles to recovery as an expected part of the treatment process.⁷⁴ Judges

65. *Id.* at 9-10.

66. *Id.* at 10-11.

67. See Susan Finlay, *The Changing Face of Justice: Alternative Approaches to Problem Solving*, 29 FORDHAM URB. L.J. 1981, 1982 (2002). Finlay recounts a number of recently fashioned “problem-solving courts,” including juvenile mental health courts, homeless courts, unified family courts, juvenile dependency drug courts, teen or youth courts, child support courts, and DUI courts. *Id.* at 1984-86.

68. Hora et al., *supra* note 51, at 454.

69. *Id.* at 448. The authors note that in 1996, over 125 drug courts were up and running in forty-five states; in 1997, roughly 325 drug treatment court programs were being planned or operating in forty-eight states. *Id.* at 455.

70. *Id.* at 463.

71. *Id.* at 468.

72. *Id.* at 507.

73. *Id.* at 469 (citing DRUG COURTS PROGRAM OFFICE, U.S. DEP’T OF JUSTICE, *DEFINING DRUG COURTS: THE KEY COMPONENTS* 6 (1997)).

74. *Id.*

and teams of court personnel follow clients through the full life cycle of their cases and may become intimately familiar with each defendant's particular circumstances and needs.⁷⁵

With the cooperation of local law enforcement and community drug rehabilitation services, drug treatment courts substitute supervised treatment plans for incarceration and probation. The need for immediate services is paramount, and defendants are placed into programs as soon as possible after their first drug court appearance.⁷⁶ Drug treatment courts aim to provide offenders with an opportunity to overcome addiction and thereby eliminate a significant cause of the behavior that led to their entrance into the criminal justice system.⁷⁷ While many drug courts are too new to make empirical analysis meaningful,⁷⁸ statistics and accounts assessing older drug treatment courts seem to indicate positive results. These older courts have largely demonstrated their effectiveness by reducing recidivism rates, increasing treatment program retention, and conserving criminal justice system resources.⁷⁹

Mental health courts have followed on the heels of the drug treatment court movement's success.⁸⁰ Pushed by an assortment of social and systemic factors—deinstitutionalization, the extraordinary prevalence of mental illness among the growing homeless population, prison overcrowding, and the high rates of recidivism in mentally-ill offenders⁸¹—these specialized courts were established to address a portion of the adults with mental health needs entering the criminal justice system.⁸² Today there are roughly thirty mental health courts in existence

75. *Id.* at 472.

76. *Id.* at 473. There is an important dichotomy in the timing of drug treatment court adjudication processes: Preadjudicative drug court models defer prosecution and divert more readily, while postadjudicative courts defer only sentencing or entry of judgment. *Id.* at 513. The preadjudicative model would appear to be more consonant with therapeutic goals, as it does not require the entry of a guilty plea to obtain treatment. *Id.*

77. *Id.* at 463.

78. Carns et al., *supra* note 56, at 8-9.

79. DRUG STRATEGIES, CUTTING CRIME: DRUG COURTS IN ACTION 6 (1997); Hora et al., *supra* note 51, at 502; *see also* STEVEN BELENKO, RESEARCH ON DRUG COURTS: A CRITICAL REVIEW 7 (2001).

80. Kessler, *supra* note 61, at 63.

81. *See, e.g.*, Carns et al., *supra* note 56, at 21; LeRoy L. Kondo, *Advocacy for the Establishment of Mental Health Specialty Courts in the Provision of Therapeutic Justice for Mentally Ill Offenders*, 28 AM. J. CRIM. L. 255, 272 (2001) (noting that mentally ill offenders reported high rates of homelessness, unemployment, substance abuse, and either physical or sexual abuse).

82. *See, e.g.*, John Pettila et al., *Preliminary Observations from an Evaluation of the Broward County Mental Health Court*, CT. REV., Winter 2001, at 14, 15 (2001). The first mental health court

and more are being planned.⁸³ Each of these courts maintains a separate docket and employs judges, prosecutors, and defense attorneys trained and familiar with the special needs of mentally disabled defendants.⁸⁴ There are two basic mental health court models: those in which courts drop or suspend criminal charges when a person is assigned to a treatment program (“meaningful diversion”) and those that require a guilty plea before assignment.⁸⁵ Individuals must qualify for participation in mental health courts, and while criteria are expanding in many systems, eligibility is frequently limited to those charged with less serious crimes and diagnosed with a current or previous mental health problem.⁸⁶

Continuing participation in mental health courts is often strictly voluntary.⁸⁷ Once a case enters the mental health court system, judges, counsel, other court personnel (frequently including a case manager or coordinator), and designated outside agencies begin to develop treatment strategies.⁸⁸ Clients return to the courtroom for non-adversarial court proceedings and regular meetings to assess their progress and monitor their program compliance.⁸⁹ While courts may choose

to become operational—the Broward County Mental Health Court in Broward County, Florida—recognized the importance of developing a

new strategy . . . to isolate and focus upon individuals arrested for misdemeanor offenses who are mentally ill or mentally retarded in view of the unique nature of mental illness and mental retardation, and the need for appropriate treatment in an environment conducive to wellness and not punishment, as well as the continuing necessity to insure the protection of the public.

Id. (quoting Admin. Order No VI-97-1-1A, *In re* Creation of a Mental Health Court Subdivision Within the County Criminal Division (Fla. Cir. Ct. June 6, 1997)).

83. Robert Bernstein & Tammy Seltzer, *Criminalization of People with Mental Illnesses: The Role of Mental Health Courts in System Reform*, 7 D.C. L. REV. 143, 144 (2003).

84. *Id.* at 150.

85. Bernstein & Seltzer, *supra* note 83, at 153 (“A guilty plea adds a conviction to the individual’s record, making it harder to get or keep the housing and employment that are so crucial to effective mental health treatment, community tenure and management of a long-term psychiatric disability.”); Nat’l Mental Health Ass’n, *Mental Health Courts*, Nov. 17, 2001, at <http://www.nmha.org/position/mentalhealthcourts.cfm> (revised Nov. 13 2004) (“NMHA does not support mental health courts unless a particular court provides a meaningful alternative to criminal sanctions . . .”).

86. *See, e.g.*, Petrila et al., *supra* note 82, at 18. Roughly half of mental health courts limit jurisdiction to defendants with misdemeanor charges, although many courts are beginning to accept people charged with more serious and violent offenses. Bernstein & Seltzer, *supra* note 83, at 154-55.

87. *See, e.g.*, Bernstein & Seltzer, *supra* note 83, at 150; Carns et al., *supra* note 56, at 27; Petrila et al., *supra* note 82, at 19.

88. *See, e.g.*, Carns et al., *supra* note 56, at 27; Petrila et al., *supra* note 82, at 20.

89. *See, e.g.*, Carns et al., *supra* note 56, at 27; Kondo, *supra* note 81, at 291-92; Petrila et al.,

to impose sanctions for non-compliance, most mental health courts instead respond by modifying treatment plans and ensuring that participants' needs are being met.⁹⁰

Looking to the future of the mental health court movement, the National Mental Health Association (NMHA) suggests that "mental health courts [can] play a role in convening criminal justice, mental health, substance abuse and other relevant social service agencies to facilitate diversion from the criminal justice system."⁹¹ NMHA advises that mental health courts not "risk further criminalizing people with mental illness, [or] fragmenting the mental health and criminal justice system," and notes that courts cannot and should not run the mental health system.⁹² Advocates believe that acceptable mental health court models should neither coerce nor compel treatment,⁹³ but rather work to "effectively determine individual needs and advocate for good individual treatment."⁹⁴ This individualized treatment should focus on recovery and choice, and include "mental and physical health care, case management, housing, supportive education, substance abuse treatment, and psychosocial services in the least restrictive environment possible."⁹⁵ Finally, in order for mental health courts to benefit the offender and community alike, court systems must avoid simply straining already insufficient local resources; they must promise to bring additional treatment resources into the communities where they operate.⁹⁶

Even if advocates' hopes for mental health courts are not fully realized, the role of these specialized treatment courts in the criminal justice system is likely to expand. In 2000, Congress authorized the Attorney General to make grants available for up to one hundred mental health courts in the America's Law Enforcement and Mental Health Project Act.⁹⁷ The Act synthesizes information

supra note 82, at 20.

90. See, e.g., Carns et al., *supra* note 56, at 27-29; Petrila et al., *supra* note 82, at 20. Mental health advocates have suggested that "[i]f the goal is to lessen the incarceration of people with mental illnesses, then using incarceration as punishment is a perversion of the whole idea of mental health courts." Bernstein & Seltzer, *supra* note 83, at 158.

91. Nat'l Mental Health Ass'n, *supra* note 85; see also Bernstein & Seltzer, *supra* note 83, at 149 (recommending that "a mental health court . . . coordinate not only with police, sheriff, and prosecution but also with state and local service systems").

92. Nat'l Mental Health Ass'n, *supra* note 85.

93. See Nat'l Mental Health Ass'n, *supra* note 85 ("Mental health courts should act as conveners of criminal justice and treatment resources, not as wielders of criminal justice sanctions to coerce mental health treatment."); see also Bernstein & Seltzer, *supra* note 83, at 14.

94. Nat'l Mental Health Ass'n, *supra* note 85.

95. *Id.*

96. *Id.*

97. Pub. L. No. 106-515, 114 Stat. 2399 (2000) (codified as amended at 42 U.S.C. § 3796ii).

and recommendations from mental health and criminal justice professionals and endorses mental health court models offering continuing judicial supervision of qualified, non-violent offenders with mental disabilities.⁹⁸ The Act also calls for the creation of coordinated programs to train court and law enforcement personnel to recognize offenders with mental health needs, to provide voluntary mental health treatment as a “meaningful diversion” from criminal sanctions, to centralize case management processes by coordinating mental health treatment plans with the provision of social services, and to provide continuity in psychiatric care following release.⁹⁹ Although mental health courts are by no means a panacea for the individual problems and systemic failures that have brought people with mental illnesses in contact with the criminal justice system, they might at least offer partial solutions by reducing the incarceration and recidivism rates of mentally ill offenders and facilitating their reintegration into their communities.¹⁰⁰

B. “Problem-Solving” Courts in the Juvenile Justice System

“Problem-solving” courts in the juvenile justice system implement “special strategies to address the particular risk factors that influence the growth and development of children today.”¹⁰¹ While the subject matter of these courts may vary, they share the goal of improving therapeutic outcomes for youthful offenders.¹⁰² Specialized juvenile courts are designed to intervene aggressively and immediately in the lives of troubled youth. Through early intervention and comprehensive treatment plans, “problem-solving” courts empower judges to consider the needs of individual offenders and creatively tailor dispositions.¹⁰³

Existing juvenile specialty court models tend to converge on very similar therapeutic elements to a greater degree than the diverse specialized treatment

(2000)).

98. *Id.*

99. *Id.*; see also Kondo, *supra* note 81, at 289 (“Judges and governmental task forces who wish to establish a [Mental Health Treatment Court] in their state may find it advisable to consider some of the following suggestions: begin with less complex misdemeanor cases with gradual transition to more complex felony cases; establish organized procedures for law enforcement and jail staff to recognize potential candidates for the [Mental Health Treatment Court]; devise probationary and conditional release plans and criteria for release of offenders from institutional commitment; and implement an organized system for follow-up to ensure that mentally ill offenders are regularly re-assessed and monitored.”).

100. Bernstein & Seltzer, *supra* note 83, at 148.

101. Gilbert et al., *supra* note 5, at 1202.

102. *Id.*

103. *Id.* at 1203.

courts in the adult criminal justice system. “Problem-solving” juvenile courts universally strive to use both consequences and incentives in treatment and recovery plans¹⁰⁴ and focus on “the role and functioning of the youth’s family in terms of rehabilitating the youth.”¹⁰⁵ These common goals can best be clarified through an examination of juvenile “problem-solving” courts in existence today, with a focus on the more established juvenile drug courts and the nascent juvenile mental health court movement.

Tailoring the drug court treatment model to juveniles has proven far more difficult than had been originally anticipated.¹⁰⁶ Juvenile drug courts have faced unique challenges, ranging from offenders’ lack of maturity and differing developmental stages to negative peer influences and family environments that often foster substance abuse problems.¹⁰⁷ As these and other issues have arisen, courts’ attempts to meet the needs of juvenile drug offenders have relied in large part on their own institutional flexibility. To begin, juvenile drug courts have implemented earlier and more comprehensive mental health screening assessments to identify youth and family substance abuse needs than their mainstream juvenile court counterparts. Once needs have been assessed, fashioning an individualized youth drug court treatment plan involves a much greater range of individuals and institutions than the adult substance abuse treatment model. Juvenile drug courts increasingly rely on coordination among court actors, the family, the treatment community, the school system, and various other juvenile-focused community agencies.¹⁰⁸ Through this coordination, juvenile drug courts strive to provide each child with a solid psychological, social, and educational foundation, including “an opportunity to be clean and sober; constructive support to aid them in resisting further criminal activity; support to perform well in school and develop positive relationships in the community; and skills that will aid them in leading productive, substance-free, and crime-free lives.”¹⁰⁹

104. *See id.* at 1210-11 (“The use of consequences and incentives is an important component of . . . specialty courts. Consequences must be structured to promote each juvenile’s ability to take responsibility for his or her actions. Positive rewards and incentives for compliance with program conditions are as important as negative sanctions for program compliance It is important to develop an appropriate array of both consequences and incentives and to communicate those to the family and youth early on in the process.”).

105. *Id.* at 1203.

106. DRUG COURT CLEARINGHOUSE & TECHNICAL ASSISTANCE PROJECT, OFFICE OF JUSTICE PROGRAMS, JUVENILE AND FAMILY DRUG COURTS: AN OVERVIEW (1998), <http://www.ncjrs.org/html/bja/jfcdcoview/dcpojuv.pdf>.

107. *Id.*

108. *Id.*

109. *Id.*

One of the first juvenile courts to open its doors was the Escambia County Juvenile Drug Court of Pensacola, Florida. The Escambia Court employs a typical multi-tiered approach to tackling the issues of juvenile substance abuse throughout youths' required twelve-month commitment to treatment. Juvenile offenders are screened within twenty-four hours of intake and referred to the juvenile drug treatment court between forty-eight hours (detainees) to three weeks (non-detainees) later.¹¹⁰ As in adult drug treatment courts, courtroom procedures are designed to facilitate and reinforce substance abuse treatment programs and seek to provide an "early intervention [that] serves as a meaningful alternative to incarceration."¹¹¹ However, the juvenile drug court goes a step beyond the adult courts in at least one respect—its focus includes the "family and social facets of juvenile addiction and drug abuse."¹¹² The court's program accordingly places an additional, rehabilitative emphasis on the offender's "vocational, educational, and spiritual needs" in the community.¹¹³ The court assigns "family intervention specialists" to assist in meeting both the youth's and his family's "psychiatric, psychological, social, economic, and medical" needs.¹¹⁴

Over the past three years, the focus of therapeutic "problem-solving" courts in the juvenile justice system has increasingly expanded beyond substance abuse concerns to include broader juvenile mental health concerns. This shift has led to the introduction of juvenile mental health courts. This movement, like its adult counterpart, stems from a recognition that mental disabilities often cause, or contribute to, delinquent behavior.¹¹⁵ The juvenile justice system—an institution designed to treat and rehabilitate youth—offers a unique opportunity to intervene in the lives of children with mental disabilities before any negative behavioral or psychological patterns take hold. Though the procedures employed by courts may vary, all appear to focus on the importance of developing individualized treatment programs for offenders and returning to the rehabilitative ideal.¹¹⁶

To date, the progress of the juvenile mental health court movement has been limited: The only two juvenile mental health courts in operation are in California's Santa Clara and Los Angeles counties.¹¹⁷ In Santa Clara, the Court

110. Hora et al., *supra* note 51, at 500.

111. *Id.* at 502 (quoting ESCAMBIA COUNTY, JUVENILE DRUG COURT PROGRAM 2 (1996)).

112. *Id.* at 500-01.

113. *Id.* at 500 (quoting ESCAMBIA COUNTY, *supra* note 111, at 1).

114. *Id.* at 501 (quoting ESCAMBIA COUNTY, *supra* note 111, at 3).

115. David E. Arredondo et al., *Juvenile Mental Health Court: Rationale and Protocols*, JUV. & FAM. CT. J., Fall 2001, at 1, 3.

116. Dennis E. Cichon, *Encouraging a Culture of Caring for Children with Disabilities*, 25 J. LEGAL MED. 39, 59 (2004).

117. See Michelle Guido & Yomi S. Wronge, *Juvenile Court Targets Mental Illness*, SAN JOSE

for the Individualized Treatment of Adolescents (CITA) offers one-year treatment programs to certain non-violent youth diagnosed with organic disorders “that have a clear biological cause,”¹¹⁸ such as attention deficit hyperactivity disorder, bipolar disorder, or severe depression.¹¹⁹ To identify candidates for CITA, all minors undergo initial screening for these and other mental disabilities upon arrival at the juvenile detention center.¹²⁰ Eligible youth receive further comprehensive assessments and—with the consensus of a multi-disciplinary team consisting of the district attorney, defense counsel, probation officer, and mental health coordinator—may ultimately be offered participation in the program.¹²¹

For youth who choose to accept CITA jurisdiction, the court’s mental health coordinator develops individualized treatment plans, drawing from a full range of mental health services.¹²² Though more serious offenders may still be incarcerated, the majority are placed on an electronic monitoring system and released to receive individualized treatment and rehabilitation services “designed to keep youth in their homes, schools and communities while providing comprehensive mental health services.”¹²³ While on probation, youth return to CITA for judicial review every thirty to ninety days.¹²⁴ To remain in the program, they must, at a minimum, demonstrate their willingness to participate in psychological counseling, comply with any prescribed medication regimens, and exhibit a “generally positive attitude.”¹²⁵ If all conditions are met and the treatment program is successfully completed, juveniles are then “released from the court’s jurisdiction and the pending charges are dismissed.”¹²⁶

The Los Angeles Juvenile Mental Health Court operates on a similar model for youth whom the district attorney’s office and other county agencies believe can benefit from the court’s intervention.¹²⁷ Youth eligibility for the court is

MERCURY NEWS, Feb. 24, 2001, at 1A; CAL. HEALTHCARE FOUND., LOS ANGELES COUNTY OFFERS SPECIAL COURT FOR JUVENILES WITH MENTAL HEALTH PROBLEMS (2002), <http://www.californiahealthline.org>.

118. Karen de Sa, *Court Addresses Causes of Juvenile Delinquency*, SAN JOSE MERCURY NEWS, Nov. 23, 2002, at 1A.

119. KQED, *Juvenile Justice, Voices from the Trenches: Raymond Davilla* (2002), at <http://www.kqed.org/w/juvenilejustice/kqedorg/davilla.html>.

120. Arredondo et al., *supra* note 115, at 11.

121. *Id.* at 11-13.

122. *Id.* at 15.

123. Guido & Wronge, *supra* note 117.

124. Arredondo et al., *supra* note 115, at 16.

125. *Id.* at 17.

126. Cichon, *supra* note 116, at 60.

127. CAL. HEALTHCARE FOUND., *supra* note 117; Greg Krikorian, *Mental Health Court Offers*

based on several criteria, including the presence of a diagnosed mental disorder or developmental disability, the ability to communicate with an attorney, the seriousness of the offense at issue, and the degree of violence in the youth's delinquent record.¹²⁸ Once eligible youth have accepted the court's jurisdiction,¹²⁹ the court employs a team of mental health professionals, school administrators, and probation officers to determine appropriate individual service plans.¹³⁰ Judges order the implementation of these service plans to provide for "home, family, therapeutic, educational, and adult transition services."¹³¹ Following disposition, judges continue to monitor each youth's progress in the assigned treatment program with assistance from an interdisciplinary team of mental health professionals, education and service providers, and representatives from the public defender and district attorney's office.¹³² Probation officers and a school-court liaison oversee juveniles' educational and treatment progress, with probation officers making frequent visits to ensure that juveniles meet the conditions of the disposition.¹³³ In addition, clinical psychologists conduct site visits and participate in regular treatment meetings as long as treatment continues, while psychiatric social workers hold service providers accountable for providing agreed-upon assistance.¹³⁴ Upon successful completion of the treatment program, delinquent charges are dismissed.¹³⁵

While it is still too early for any comprehensive analysis of this approach to have been completed,¹³⁶ some data have indicated that the juvenile drug court concept may effectively facilitate recovery and lower participants' likelihood of

New Options (Jan. 4, 2002), at <http://www.namisc.org/newsletters/January02/MentalHealthCourt.htm>.

128. CAL. HEALTHCARE FOUND., *supra* note 117. If necessary the court may also opt to order a comprehensive psychological evaluation to aid in eligibility determinations. Agata DiGiovanni, *The Los Angeles County Juvenile Mental Health Court: An Innovative Approach to Crime, Violence, and Delinquency Among Our Youth*, 23 J. JUV. L. 1, 6 (2003).

129. Richard Kwon, L.A. Youth, New Court Program Helps Teen Offenders with Mental Health Problems (May-June 2002), at http://www.layouth.com/4_5_11.htm.

130. L.A. County Programs, Increasing Mental Health Services, at <http://www.cpoc.org/JJCPA/losangeles.htm> (last visited Apr. 28, 2003).

131. DiGiovanni, *supra* note 128, at 6.

132. Cheryl Romo, Pub. Counsel Law Ctr., Focus on Youth Demands Discipline—Training Session for Aiding Troubled Kids Falls on Tragic Day (Oct. 1, 2001), at <http://www.publiccounsel.org/news/2001/oct101.htm>.

133. DiGiovanni, *supra* note 128, at 7.

134. *Id.*

135. *Id.* at 6.

136. Given the Santa Clara and Los Angeles courts' openings in February and October 2001, respectively, there has simply not been time to comprehensively evaluate their success.

recidivism.¹³⁷ There is unfortunately even less extensive longitudinal data on the effectiveness of juvenile mental health court interventions. Nevertheless, the preliminary data from CITA do seem to offer some hope for this model's viability. For example, Santa Clara Juvenile Court Judge Davilla reports that internal assessments show a relatively substantial reduction in recidivism for those who participate in the specialized CITA program.¹³⁸

IV. A RETURN TO THE REHABILITATIVE IDEAL: USING THE JUVENILE MENTAL HEALTH COURT MODEL TO REDISCOVER THE THERAPEUTIC GOALS OF THE JUVENILE JUSTICE SYSTEM

While the juvenile mental health court movement may be laudable in its aspirations, the promise of juvenile mental health courts is incredibly limited when placed against the background of an overwhelmingly large population of children and adolescents in need of mental health services.¹³⁹ This is not a flaw in the model espoused by juvenile mental health court advocates; it is simply a reflection of the reality that anywhere from a large minority to a sweeping majority of minors who come before the juvenile justice system exhibit mental disabilities.¹⁴⁰ Creating a network of juvenile mental health courts large enough to serve such a large proportion of the juvenile offender population seems unwise and entirely unnecessary in light of the existing juvenile justice system's potential to do the same.

The support for therapeutic jurisprudence ideals and programs embodied in the juvenile mental health court movement—even in the face of punitive reforms—may signal a renewed opportunity for the juvenile justice system to return to its fundamental emphasis on treatment and rehabilitation for all offenders.¹⁴¹ Juvenile mental health courts have reintroduced the important goals

137. See Hora et al., *supra* note 51, at 502; see also Kessler, *supra* note 61, at 63.

138. KQED, *supra* note 119 (“[W]e have lowered the recidivism rate . . . to 7 percent [compared to the 25 percent recidivism rate for the general juvenile population].”).

139. Romo, *supra* note 132.

140. NAT'L MENTAL HEALTH ASS'N, *supra* note 37; Warboys & Wilber, *supra* note 11, at 506 (citing N.A. Brandenburg et al., *The Epidemiology of Childhood Psychiatric Disorders: Recent Prevalence Findings and Methodological Issues*, 29 J. AM. ACAD. CHILD & ADOLESCENT PSYCHIATRY 76 (1990)); see also Baerger et al., *supra* note 37, at 21 (finding that more than one-third of adolescents arrested and adjudicated within the juvenile justice system exhibit symptoms of major affective disorders); *supra* Part I.

141. Gilbert et al., *supra* note 5, at 1197-1201. Gilbert posits that “[a] more heightened and intensified emphasis on therapy and rehabilitation, accompanied by appropriate accountability and due process safeguards, does not represent a dramatic philosophical shift from past and current juvenile justice considerations and objectives.” *Id.* at 1200-01.

of “accountability, treatment, healing, and a long-range successful outcome for the child and family,” which may well be a “necessary step for meaningful reform.”¹⁴² The commentary that juvenile mental health courts have inspired from the press, public, and the courts’ own actors—noting, for example, the courts’ ability to look at ““why a kid got involved in the system and how we can prevent it from happening again,””¹⁴³ and ““the real, underlying issues . . . with these kids.””¹⁴⁴—is remarkably reminiscent of the traditional juvenile justice system’s goals. CITA has been touted as “a national model for its efforts to address delinquency’s causes,”¹⁴⁵ while the court’s judge has firmly suggested ““if we can get young people on track early on, get their parents on track early on . . . we can make an impact.””¹⁴⁶

The juvenile mental health court model has successfully readjusted its primary focus away from punishment and culpability and back toward the concepts of individualized treatment and rehabilitation. This focus on mental health in the juvenile justice system is not new; there was “considerable psychiatric involvement in the original juvenile courts” and a long-recognized linkage between psychiatry and juvenile delinquency.¹⁴⁷ Despite the current juvenile justice system’s failure to effectively address offenders’ mental health needs, there is continuing support for a “positive orientation toward fundamental issues related to mental health.”¹⁴⁸ Whatever punitive reforms have come to pass, juvenile court actors still essentially “believe offenders can be reformed, mental health services have value, and the youth’s mental status is significant for making case dispositions.”¹⁴⁹ If the juvenile court were to build on these beliefs, it might find a renewed ability to meet the mental health needs of those before it.

Nevertheless, the juvenile justice system’s emphasis on mental health concerns continues to wane as dispositions focus less and less on desirable

142. Cichon, *supra* note 116, at 61. Indeed, the title of the first juvenile mental health court alone—the “Court for the Individualized Treatment of Adolescents”—strongly suggests a return to the diagnostic, case-by-case approach of the early juvenile court.

143. Krikorian, *supra* note 127 (quoting Nancy Ramseyer, Deputy Public Defender, Los Angeles County).

144. Guido & Wronge, *supra* note 117 (quoting Judi Marshall, Deputy Probation Officer, Santa Clara County).

145. de Sa, *supra* note 118.

146. KQED, *supra* note 119 (quoting Judge Raymond Davilla, Santa Clara County Juvenile Court).

147. Am. Psychiatric Ass’n, *APA Official Actions: The Psychiatrist and the Juvenile Justice System*, 174 AM J. PSYCHIATRY 1584 (1990).

148. Carolyn S. Breda, *The Mental Health Orientation of Juvenile Courts*, 28 J. BEHAV. HEALTH SERVS. & RES. 89, 93 (2001).

149. *Id.*

treatment outcomes. But if this system were to draw upon the ideas and practical operations of the juvenile mental health courts, perhaps it could find its way back to an individualized, case-by-case approach to administering justice. The juvenile mental health court model can encourage juvenile courts to function as child-centered, family-focused, community-based, and culturally competent institutions.¹⁵⁰ Advocates have pushed for the development of specialized juvenile mental health courts in the hope that these courts would effectively identify, triage, and treat mentally-disabled youth with a comprehensive array of integrated and coordinated services.¹⁵¹ However, it is well within the power and purview of the larger juvenile court to address the concerns of these juvenile mental health court advocates without isolating mental health considerations in a specialty court.¹⁵² To do so, juvenile courts must begin to establish linkages with therapeutic treatment and social service providers at the organizational level. These courts should strive to institute more therapeutic procedures, roles, court rules, information systems, and sentencing options.¹⁵³ At a more general level, those guiding the system must adopt policies that foster therapeutic outcomes and awareness of mental health needs. Advocates should also attempt to convince state legislatures to enact and revise laws reflecting the principles of therapeutic jurisprudence.¹⁵⁴

V. POISED FOR REFORM: RECOMMENDATIONS FOR ADDRESSING THE MENTAL HEALTH ISSUES OF YOUTH IN THE JUVENILE JUSTICE SYSTEM

Even decades of punitive juvenile justice reform have not wholly eroded the rehabilitative ideal; virtually all juvenile courts retain some portion of their original mandate “to provide any and all necessary services to rehabilitate and treat youths.”¹⁵⁵ Acknowledging and meeting the mental health needs of youth in the juvenile justice system may not only enable states to address an important

150. Arredondo et al., *supra* note 115, at 14.

151. *Id.*

152. See Bernstein & Seltzer, *supra* note 83, at 149.

153. *Id.* at 147.

154. Gilbert et al., *supra* note 5, at 1201. For example, many therapeutic justice advocates favor a softening of the adversarial system to better obtain more just resolution of cases and the best available treatment options. Kondo, *supra* note 81, at 262.

155. See Mears et al., *supra* note 44; see also Theodore Fallon, Jr. & Dawn Dawson, *Juvenile Justice: Yesterday and Today*, in AM. ACAD. CHILD & ADOLESCENT PSYCHIATRY, *supra* note 61, at 14-15 (noting that “[a]lthough it may seem otherwise, even after a century of modifications, and broad variations from state to state, most juvenile justice laws and governmental structures specify that the juvenile justice system continues to act in the best interest of the youth.”).

factor contributing to antisocial behavior,¹⁵⁶ but may also lead the juvenile court back to its original aim to serve the “best interests of youths.”¹⁵⁷ However, despite the increasing attention that children’s mental health needs have received in recent years, there remains only minimal recognition of the importance of these needs.¹⁵⁸ While today’s juvenile courts may not be the ideal means for handling youth with mental health needs, states must realize that they are a necessary one: “[I]t is crucial that we deal not only with the specific behavior or circumstances that bring [juveniles] to our attention, but also with their underlying, often long-term mental health and substance abuse problems.”¹⁵⁹

The juvenile justice system remains uniquely equipped to address the backgrounds and characteristics of young offenders and to provide opportunities for rehabilitation through individualized assessments and treatment plans.¹⁶⁰ Though the juvenile court has grown increasingly similar to its retributive criminal counterpart, this shift has not completely detached the juvenile justice system from its focus on rehabilitation; both justice systems aim to achieve “[the] proscription of deviant behavior, social protection through supervision and incapacitation, and reform and rehabilitation of delinquents.”¹⁶¹ To this end, the American Psychiatric Association believes both that the importance of the involvement of mental health professionals in the juvenile justice system remains constant in the face of punitive reforms¹⁶² and that resources should be reallocated to adequately address the mental health needs of those young offenders amenable to treatment.¹⁶³

The sad reality is that the current juvenile justice system is simply not equipped to meet the mental health needs of large numbers of juveniles who

156. See Redding, *supra* note 31; GAIL WASSERMAN & JOLEEN BAGWELL, NAMI AND CENTER FOR THE PROMOTION OF MENTAL HEALTH IN JUVENILE JUSTICE (2001), <http://ocd.nami.org/youth/juvenilejusticescreening.html>.

157. Mears et al., *supra* note 44.

158. *Id.*; Teplin et al., *supra* note 41, at 1139; Michelle Wierson et al., *Epidemiology and Treatment of Mental Health Problems in Juvenile Delinquents*, 14 ADVANCES BEHAV. RES. & THERAPY 93 (1992).

159. SHAY BILCHIK, U.S. DEPT. OF JUST., OJJDP FACT SHEET: MENTAL HEALTH DISORDERS AND SUBSTANCE ABUSE PROBLEMS AMONG JUVENILES (1998), <http://www.childrensprogram.org/media/pdf/mentalhealthdisorders.pdf>; see also Thomas et al., *supra* note 1, at 624.

160. Wong, *supra* note 33, at 165.

161. Stephen Wizner, *On Youth Crime and the Juvenile Court*, 36 B.C. L. REV. 1025, 1029 (1995).

162. Am. Psychiatric Ass’n, *supra* note 147, at 1584.

163. Am. Psychiatric Ass’n, Anticrime Bills Would Punish Juveniles While Jeopardizing Psychiatric Assessment (Sept. 6, 1996), at <http://www.psych.org/pnews/96-09-06/crime.html>.

either have psychiatric disorders or are at risk of developing them.¹⁶⁴ Juvenile courts' access to services available in schools, welfare agencies, and community organizations¹⁶⁵ may make them exceptionally capable of tailoring integrated treatment plans to the mental health needs of children who come before them.¹⁶⁶ Juvenile justice officials must recognize the proper care of youth with diagnosable or emerging mental health problems as "among their greatest challenges."¹⁶⁷ While research to date has produced only limited evidence of how best to contend with obstacles to meeting the mental health needs of juvenile offenders, recommendations for mental health care reform in juvenile justice abound.¹⁶⁸ The recommendations address a wide range of issues, including the problems of screening and assessment, education and training, coordination across systems, treatment, and delivering mental health care during incarceration.¹⁶⁹

A piecemeal approach to meeting juvenile mental health needs—as has been adopted in far too many juvenile court jurisdictions—is inadequate.¹⁷⁰ Screening programs are fruitless if results do not come to the attention of juvenile court personnel, just as the court's awareness of detected mental health needs is of little use should there be no services available to meet those needs. An effective system for addressing juvenile mental health needs must incorporate strategies for dealing with these needs from the time they are identified through the completion of post-dispositional treatment.¹⁷¹ Youth must be adequately screened before they can be matched with appropriate mental health services. Court staff

164. HANDLE WITH CARE, *supra* note 30; Teplin et al., *supra* note 41, at 1139.

165. See Nat'l Mental Health Ass'n, *supra* note 36.

166. See NAT'L COUNCIL OF JUVENILE & FAMILY COURT JUDGES, *supra* note 3, at 3.

167. Cocozza & Skowrya, *supra* note 39, at 7.

168. Mears et al., *supra* note 44.

169. See Mark Soler, *Health Issues for Adolescents in the Justice System*, 31 J. ADOLESCENT HEALTH 321, 322-24 (2002).

170. Quite possibly, no integrated juvenile justice system has ever existed; instead there have been "an assortment of aggregate entities of varying quality that do not generally communicate with each other in meaningful ways." Charles Billikas, *The Ideal Juvenile Rehabilitation Program: An Integrated System*, 21 NEW ENG. J. ON CRIM. & CIV. CONFINEMENT 411, 418 (1995).

171. Griffin and Jenuwine argue that "in an ideal setting, a mentally ill youth who was arrested could move from an assessment center, to a detention center with treatment planning, to a mental health court, to a court order for community-based services . . . allow[ing] the juvenile courts to embrace the tenets of therapeutic jurisprudence." Griffin & Jenuwine, *supra* note 58, at 86. The philosophy behind such a system is very straightforward—as the Maryland Juvenile Justice Coalition summarizes, an effective treatment model must "identify the services and supports that a child and his/her family needs and provide them as long as they are needed." MD. JUVENILE JUSTICE COALITION, PRINCIPLES OF A MODEL JUVENILE JUSTICE SYSTEM 10 (2002), http://www.acy.org/web_data/Model%20Juvenile%20Justice%20System%202002.pdf.

must be trained to work with these youth and to arrange for these services while juveniles remain under the jurisdiction of the court. Interagency coordination must bridge the gaps between the juvenile justice, mental health, and educational systems to enable juveniles to obtain the necessary treatment in all areas of their lives.

A. Screening and Assessment

In order to meet youths' mental health needs, their mental health status must be evaluated at both the initial point of contact with the juvenile justice system and at every subsequent stage in the adjudication process.¹⁷² Indeed, each "referral to the juvenile justice system presents an opportunity to identify a child in need of mental health treatment."¹⁷³ This evaluation may come in two forms—screening and assessment. Screening is the relatively brief process used to identify youth at an increased risk for mental disorders or in need of immediate attention and more complete review.¹⁷⁴ Assessment offers this review and further examines a youth's psychological needs and problems.¹⁷⁵

Most juvenile courts do not adequately screen youth in contact with the juvenile justice system and also lack clear guidelines for identifying mental

172. See BILCHIK, *supra* note 159; Soler, *supra* note 169, at 322; Warboys & Wilber, *supra* note 11, at 507 ("At each stage of the juvenile court process, there are opportunities for . . . mental health professional[s] to play an extremely important role."). Nurcombe and Partlett provide a listing of typical times mental health professionals might become involved in the process and what issues courts might ask them to explore:

Prior to the disposition hearing. Amenability to treatment? Appropriate disposition? Recommended treatment?

Prior to a transfer hearing. Amenability to treatment? Dangerousness? Competence to waive due process rights? Competence to stand trial? Mental health at the time of offense?

Prior to adjudication. Competence to stand trial? Competence to waive due process rights? Mental state at the time of the offense? Recommended psychiatric treatment? Appropriate disposition?

Prior to diversion. Amenability to treatment? Appropriate diversion? Recommended treatment?

NURCOMBE & PARTLETT, *supra* note 7, at 306.

173. Cichon, *supra* note 116, at 59.

174. THOMAS GRISSO & LEE UNDERWOOD, NAT'L CTR. FOR MENTAL HEALTH & JUVENILE JUSTICE, SCREENING AND ASSESSING MENTAL HEALTH AND SUBSTANCE USE DISORDERS AMONG YOUTH IN THE JUVENILE JUSTICE SYSTEM 1 (2003), <http://www.ncjrs.org/pdffiles1/ojdp/204956.pdf>.

175. *Id.*

health problems in youth.¹⁷⁶ In many jurisdictions, juveniles simply do not receive mental health screening or assessment at all.¹⁷⁷ In those systems that do offer mental health evaluations, the instruments used often present numerous reliability, validity, and administrative problems.¹⁷⁸ The development of systematic intake procedures to determine and evaluate mental health needs is, however, essential to meeting those needs in the juvenile justice system.¹⁷⁹ Every minor in contact with the system should be screened and—if necessary—evaluated for the presence of mental health disorders.¹⁸⁰

The screening and assessment process involves more than the simple administration of a psychometric testing instrument. To gain a full picture of a juvenile's mental health needs, medical histories for both the youth and her family must be obtained and evaluated.¹⁸¹ Yet in the current system, intake personnel are rarely provided all relevant "reports, records, or background information pertinent to the child's behavior," leaving the juvenile court with

176. Soler, *supra* note 169, at 322

177. Fallon & Dawson, *supra* note 155, at 16 (finding that many juveniles are not screened for mental health problems either pre- or post-adjudication); Drew H. Barzman et al., *Attention-Deficit Disorder Diagnosis and Treatment*, 25 J. LEGAL MED. 23, 25 (2004); Soler, *supra* note 169, at 323 (noting that many systems do not offer mental health assessments at arrest, admission, disposition, or placement).

178. See Coccozza & Skowrya, *supra* note 39, at 9; Soler, *supra* note 169, at 323. Social scientists caution against the use of instruments that have not been adequately researched or tested on adolescents. GRISSE & UNDERWOOD, *supra* note 174, at 5. Many recommend using only instruments tailored to minimal reading levels that are "amenable to administration with youth of diverse ethnic, cultural, and linguistic backgrounds." *Id.* These and other considerations suggest that great care should be taken in selecting the most appropriate screening and assessment instruments for youth. *Id.* One promising example of a standardized screen may be the recently developed Massachusetts Youth Screening Instrument—a shorter, easily administered, well-normed inventory. Coccozza & Skowrya, *supra* note 39, at 9.

179. Redding, *supra* note 31 (noting that "mental illness and substance abuse are significant risk factors for delinquency"); see also Barzman *supra* note 177, at 26 ("[A] systematic method of identification of mental illness is the cornerstone to developing an appropriate approach to youths who may need further evaluation and treatment.").

180. Coccozza & Skowrya, *supra* note 39, at 9; see also Barzman, *supra* note 177, at 25 ("One recommended approach is to screen youths upon their entry into the juvenile justice system to evaluate for unknown mental health issues."). More specifically, Curtis Heaston recommends a three-tiered assessment approach, with the first level of assessment for juveniles first entering the justice system, the second level in the courtrooms for juveniles whose cases are filed, and the third level for juveniles held in the detention center. Curtis Heaston et al., *Mental Health Assessment of Minors in the Juvenile Justice System*, 11 WASH. U. J.L. & POL'Y 141, 149 (2003).

181. Claudette Brown, *Crossing Over: From Child Welfare to Juvenile Justice*, 36 MD. B.J. 18, 22 (2003).

little help in its efforts to make a well-informed decision.¹⁸² In addition to correcting this initial informational shortfall, court personnel should make reassessments and administer necessary interventions, psychopharmacological or otherwise, where appropriate.¹⁸³ Early screening and continuing mental health evaluation are essential to facilitate the expedient and appropriate placement of youth with mental disorders in the juvenile justice system into safe, appropriately suited treatment environments.¹⁸⁴

B. Educating and Training Juvenile Justice Personnel

From judges and defense attorneys to prosecutors and probation officers, most juvenile court personnel have received little to no formal education or training in handling youth with mental disorders. The large majority of them have limited background knowledge of child and adolescent development generally, let alone the subset of issues related to childhood mental disability.¹⁸⁵ As a result, many actors in the juvenile justice system may be unable to understand the results of mental health assessments, the mental needs of individual youth, or the promise of appropriate treatment options.¹⁸⁶ Juvenile justice personnel must have access to more opportunities for education and training to respond effectively to the mental health needs of juvenile offenders.¹⁸⁷

C. Coordinating Across Systems

Both the problems created by fragmented mental health services and the need for increased coordination have long been recognized.¹⁸⁸ Although many youth present coexisting mental disabilities, individual state entities often offer only limited services to individuals qualifying for treatment in multiple

182. Cichon, *supra* note 116, at 57.

183. Brown, *supra* note 181, at 22.

184. HANDLE WITH CARE, *supra* note 30, at 41-42.

185. Fallon & Dawson, *supra* note 155, at 17; Soler, *supra* note 169, at 323; *see also* David E. Arredondo, *Children, Crime, and Consequences: Juvenile Justice in America*, 14 STAN. L. & POL'Y REV. 13, 28 (2003). To make informed determinations, "decision-makers need familiarity with the general principles of child development and a reasonable knowledge of the risks and needs presented by each individual offender." *Id.*

186. Soler, *supra* note 169, at 323-24.

187. *See* Redding, *supra* note 31; *see also* Langemo, *supra* note 50, at 162 ("Juvenile court staff should have significant education, training, and experience in both the physical and mental attributes of juveniles, [which] should continue on a regular basis in order for the staff to maintain up-to-date knowledge.") In addition, staff "should assist in collecting and developing social and psychological information on juveniles." *Id.*

188. Soler, *supra* note 169, at 323.

systems.¹⁸⁹ Such youth clearly need the services of more than a single public system.¹⁹⁰ Care coordination involves accessing and assembling medical, psychiatric, social, educational, and other support services essential to meeting these youth's mental health needs.¹⁹¹ The Child and Adolescent Service System Program (CASSP)—an organization that links mentally-disabled children and adolescents with needed services—has consistently advocated for a continuum of care that provides youth with an array of child-centered, family-focused, community-based, multi-system, and culturally competent services.¹⁹² These recommendations have at last begun to make inroads into the juvenile justice system and cross-system collaboration is quickly emerging as indispensable to the effective provision of mental health treatment solutions for children and adolescents.¹⁹³

All agencies involved in the treatment and care of youths with mental disorders—including the criminal and juvenile justice systems, mental health systems, schools, family and social service organizations, law enforcement agencies, medical institutions, and substance service systems—must collaborate to develop and implement effective treatment strategies.¹⁹⁴ Comprehensive integrated services are more likely to attend to the underlying causes of delinquency and recidivism, thereby offering youths an opportunity for a smoother transition out of the juvenile justice system into productive adult lives.¹⁹⁵ An approach that brings each and every agency responsible for administering juvenile mental health treatment together for planning, cross-training, and service delivery is ideal.¹⁹⁶

D. Delivering Mental Health Care to Incarcerated Juvenile Offenders

Between 1923 and 1974, the rates of admission to juvenile correctional facilities increased nine-fold.¹⁹⁷ Congress' recognition that many of these placements were inappropriate prompted it to pass the Juvenile Justice and

189. Dennis E. Cichon, *Developing a Mental Health Code for Minors*, 13 T.M. COOLEY L. REV. 529, 570-71 (1996).

190. MD. JUVENILE JUSTICE COALITION, *supra* note 171, at 10.

191. Gilbert et al., *supra* note 5, at 1180.

192. See, e.g., N.Y. State Office of Mental Health, Child and Adolescent Service Program, at <http://www.omh.state.ny.us/omhweb/ebp/cassp.htm> (last visited, Apr. 15, 2005).

193. Cocozza & Skowrya, *supra* note 39, at 7.

194. HANDLE WITH CARE, *supra* note 30; Cocozza & Skowrya, *supra* note 39, at 7-8.

195. Warboys & Wilber, *supra* note 11, at 518.

196. Cocozza & Skowrya, *supra* note 39, at 11.

197. Weithorn, *supra* note 1, at 803.

Delinquency Prevention Act (JJDPa) of 1974.¹⁹⁸ Although the JJDPa's supporters had hoped to remove all but the most serious juvenile offenders from correctional facilities, the Act ultimately demanded the removal of only non-offenders (dependent and neglected youth) and status offenders (youth whose actions were considered delinquent only as a result of their status as minors).¹⁹⁹ Though rates of institutionalization quickly plummeted,²⁰⁰ the Act still left many children with mental and emotional disorders vulnerable to incarceration in the difficult and often severely overcrowded environments of detention centers and youth prisons.²⁰¹

Punitive reforms in juvenile justice, including state transfer laws and shrinking juvenile court jurisdiction, have begun to unravel the juvenile justice policy created under the JJDPa.²⁰² The escalating numbers of juveniles tried in adult criminal courts and incarcerated in adult jails and prisons has become a particularly alarming trend.²⁰³ The youth incarcerated in prisons more than tripled in the 1990s,²⁰⁴ despite evidence that juveniles incarcerated in adult institutions are "5 times more likely to be sexually assaulted, twice as likely to be beaten by staff, and 50% more likely to be attacked with a weapon than youth in juvenile facilities."²⁰⁵ These conditions may be especially damaging for youths with mental disorders, who are almost eight times more likely to commit suicide in adult jails than in juvenile institutions.²⁰⁶ The data strongly suggest that incarcerating juveniles, in particular those with mental health needs, in adult prisons is inappropriate and that states should work to develop separate juvenile facilities for transferred offenders that are better able to meet the special needs of incarcerated youth.²⁰⁷

Even in juvenile correctional facilities, however, mental health services are inadequate. Most juvenile facilities provide only crisis intervention and occasional group counseling; the vast majority do not administer one-on-one

198. Juvenile Justice and Delinquency Prevention Act of 1974, Pub. L. No. 93-415, 88 Stat. 1109 (codified as amended at 42 U.S.C. §§ 5601-5785 (2000)).

199. Weithorn, *supra* note 1, at 803.

200. *Id.*

201. Nat'l Mental Health Ass'n, *supra* note 36.

202. Soler, *supra* note 169, at 326.

203. *Id.*

204. JAMES AUSTIN ET AL., U.S. DEP'T OF JUSTICE, JUVENILES IN ADULT PRISONS AND JAILS: A NATIONAL ASSESSMENT 5 (2000), <http://www.ncjrs.org/pdffiles1/bja/182503.pdf>.

205. Soler, *supra* note 169, at 326.

206. MICHAEL G. FLAHERTY, U.S. DEP'T OF JUSTICE, AN ASSESSMENT OF THE NATIONAL INCIDENCE OF JUVENILE SUICIDE IN ADULT JAILS, LOCKUPS, AND JUVENILE DETENTION CENTERS (1980).

207. Soler, *supra* note 169, at 327.

therapy nor offer services in collaboration with providers outside the system.²⁰⁸ All professionals—from social workers and nurses to correctional officers and facility administrators—must advocate for the adequate staffing of mental health professionals to address the mental health needs of incarcerated youth.²⁰⁹

The National Mental Health Association recommends certain guidelines for reforming treatment during confinement, including round-the-clock mental health services and special treatment for children with histories of family abuse, violence, substance abuse, and educational difficulties.²¹⁰ Treatment should be individualized and provided in the least restrictive environment possible, and children should be transferred to appropriate medical or mental health facilities when conditions so warrant.²¹¹ Effective treatment plans cannot terminate upon release and discharge plans should facilitate the integration of incarcerated children back into their families and communities.²¹² Unfortunately, most facilities have failed to develop even weak links with community-based health programs and aftercare services to meet the specific needs of youth released from custody.²¹³ These links must be forged. If incarceration is unavoidable, juvenile offenders with mental disabilities should be placed in correctional or mental health institutions able to meet their needs and returned to their families and communities as swiftly as possible.

E. Effective Community-Based Treatment Options

The juvenile court, throughout much of its existence, has placed a premium on allowing children to grow up in community settings.²¹⁴ As noted above, the juvenile justice and mental health systems must work together to develop programs and implement services that meet the mental health needs of youth, preferably in their home environments.²¹⁵ The National Mental Health

208. *Id.* at 323.

209. Nat'l Mental Health Ass'n, *supra* note 36. In addition, the American Medical Association strongly supports both model legislation that addresses the mental health care needs of detained and incarcerated youth and further steps necessary to implement such legislation on state and federal levels. Louis J. Kraus, *Standards for Juvenile Detention and Confinement Facilities*, in AM. ACAD. OF CHILD & ADOLESCENT PSYCHIATRY, *supra* note 61, at 27.

210. Nat'l Mental Health Ass'n, *supra* note 36.

211. *Id.*

212. *Id.*

213. Soler, *supra* note 169, at 323.

214. Zimring, *supra* note 6, at 2481.

215. Nat'l Mental Health Ass'n, *supra* note 36; *see also* BILCHIK, *supra* note 159; Soler, *supra* note 169, at 324 (noting the importance of interagency collaboration in improving treatment services).

Association recommends that these services be “treatment-oriented, appropriate for the child’s age, gender, and culture, individualized, and family focused.”²¹⁶ Most jurisdictions, however, fail to provide adequate non-institutional public mental health services for children and families.²¹⁷ Rekindling the rehabilitative underpinnings of the juvenile justice system may intensify the public’s interest in providing mentally disabled youth with the services to which they are entitled.²¹⁸

Non-residential, community-based services would offer these jurisdictions cost-effective opportunities to intervene when a juvenile’s aggressive or delinquent behavior first arises.²¹⁹ These services are designed to keep youth active in their home, school, and community environments “while providing a comprehensive set of services that respond to their mental health needs and related problems.”²²⁰ They maintain the integrity of the juvenile’s family unit,²²¹ are less restrictive and invasive for emotionally and mentally disordered youth, and offer more effective treatment prospects than either institutional or residential placements.²²² Across the board, “[t]here is a growing, if not already established, consensus that community-based care is more effective than hospitalization in treating all but the most severe mental disorders,” with a growing body of research documenting the “superiority” of community-based treatment over institutionalization.²²³

While returning mentally disabled juvenile offenders to safe and stable homes is critical to effective treatment plans, youth in the juvenile justice system are often from highly dysfunctional family settings or have suffered from parental neglect or abuse.²²⁴ Successful community-based services do not merely return delinquent youths to their often confused or anxious families; they strive to treat the families of delinquent offenders in addition to the juveniles

216. Nat’l Mental Health Ass’n, *supra* note 41.

217. See Weithorn, *supra* note 1, at 829.

218. Anderson, *supra* note 9, at 78. (“It is still possible to imagine ways juvenile delinquents might be sanctioned and supervised effectively as juveniles, not adults, without removing them from the community. The drift away from historical juvenile justice remains premature.”).

219. Redding, *supra* note 31; Gary B. Sutnick, Note, “Reasonable Efforts” Revisited: *Reforming Federal Financing of Children’s Mental Health Services*, 68 N.Y.U. L. REV. 136, 146 (1993).

220. Coccozza & Skowrya, *supra* note 39, at 10.

221. Sutnick, *supra* note 219, at 145-46.

222. Weithorn, *supra* note 1, at 788-94.

223. Cichon, *supra* note 189, at 538.

224. Redding, *supra* note 31; TRINA Osher & Pat Hunt, INVOLVING FAMILIES OF YOUTH WHO ARE IN CONTACT WITH THE JUVENILE JUSTICE SYSTEM 1 (Nat’l Ctr. for Mental Health & Juvenile Justice, Research & Program Brief, Dec. 2002), <http://www.ncmhjj.com/pdfs/publications/Family.pdf>.

themselves.²²⁵ Supportive family involvement is crucial, and mental health treatment plans must provide the families of juveniles with psychiatric disorders with the requisite knowledge and tools necessary to effectively manage the mental health needs of their children.²²⁶ Focusing on families, as opposed to juvenile delinquents in isolation, can lead to “a fundamental change in the lifestyle of the youths and families that will, at minimum, substantially reduce the likelihood of their further involvement with the justice system, increase public safety, and significantly enhance the likelihood that the youths and their families will function as productive community members.”²²⁷

Finding effective treatment models for youth and families involved in the juvenile justice system and meeting their emotional, mental health, and behavioral needs can be quite difficult.²²⁸ However, the traditional “one-size-fits-all” model often used in juvenile justice and mental health systems does not appropriately address these needs.²²⁹ Non-residential community-based programs avoid depriving juveniles of the liberty necessary for productive development, are less expensive than institutionalization, and are more effective in treating all but a small minority of youth facing mental disabilities.²³⁰ Several available treatments—wraparound services,²³¹ multi-systemic therapy,²³² and functional family therapy²³³—offer juveniles and their families comprehensive and coordinated services from a variety of service systems.²³⁴ These same treatments

225. OSHER & HUNT, *supra* note 224, at 2.

226. Sutnick, *supra* note 219, at 145 (“[P]rograms that work with families as whole units generally achieve more long-term success than does [sic] hospitalization because they teach the families strategies for dealing with their children’s needs.”).

227. Gilbert et al., *supra* note 5, at 1187 (noting that “[t]he laws nationwide are becoming more and more reflective of the theory that intervention strategies of treatment must be provided to not only the juvenile at risk but also the juvenile’s family”).

228. Bruce Kamradt, *Wraparound Milwaukee: Aiding Youth with Mental Health Needs*, 7 JUV. JUST. 14, 14 (2000).

229. *Id.*

230. Cichon, *supra* note 189, at 530.

231. Mears et al., *supra* note 44 (“Wraparound service programs focus on providing treatment that is tailored to the needs of each youth. . . . [T]he Wraparound philosophy is specifically oriented toward placing youths in ‘small group homes with individualized care, flexible programming, and a ‘never give up’ philosophy.’” (citation omitted)).

232. See Coccozza & Skowrya, *supra* note 39, at 10 (“[Multisystemic Therapy] is a family- and community-based treatment model that provides services in the home and community settings and addresses a range of family, peer, school, and community factors.”).

233. See HANDLE WITH CARE, *supra* note 30, at 43 (“Functional Family Therapy . . . is an ‘outcome-driven prevention/intervention program for youth who have demonstrated the entire range of maladaptive, acting out behavior and related syndromes.’”).

234. Soler, *supra* note 169, at 323.

promise the juvenile justice system more successful therapeutic outcomes and dramatic drops in recidivism rates—according to some, decreases in recidivism rates range from twenty-five percent for “structured, meaningful, and sensitive treatment” to eighty percent for programs deemed to be the “most successful.”²³⁵

VI. IMPLEMENTING AN EFFECTIVE SYSTEM TO MEET JUVENILE MENTAL HEALTH NEEDS IN THE MODERN JUVENILE COURTS

Without adequate mental health treatment programs in place for juvenile offenders and their families, “there will be serious long term and financial consequences.”²³⁶ Unfortunately, both the juvenile justice and mental health systems are chronically under-funded.²³⁷ In a constrained budgetary environment,²³⁸ funding shortages severely limit the mental health services localities can offer juvenile offenders.²³⁹ Juvenile offenders are given relatively low priority within this population of children and adolescents with mental health needs, and their often forced reliance on costly emergency services and limited case management causes further strain on the limited funds made available to meet their needs.²⁴⁰ The lack of early and effective mental health intervention jeopardizes youthful offenders’ ability to remain at home in family care, spawning a “downward spiral” of deteriorating functioning that often results in expensive short and long-term institutional placements.²⁴¹ Not only is delayed mental health intervention more expensive and less effective than early intervention, it leads to other social costs as well, including “school failures, teen pregnancies, juvenile delinquency, welfare, community disintegration, violence and imprisonment.”²⁴²

Although prospects for many youthful offenders with mental health needs are bleaker than ever, there is still strong hope that the juvenile courts’ ability to serve these offenders could be “revitalized” through additional and reorganized funding.²⁴³ State and local governments seem particularly well-suited to fuel this

235. HANDLE WITH CARE, *supra* note 30, at 42.

236. Nayowith, *supra* note 4, at 367.

237. Griffin & Jenuwine, *supra* note 58, at 74.

238. Nayowith, *supra* note 4, at 366.

239. Redding, *supra* note 31.

240. *Id.*

241. Nayowith, *supra* note 4, at 367. The costs of this “downward spiral” are high. For example, while residential treatment facilities in New York State operate at a cost of roughly \$400 per day and community-based day services reach a lower daily cost of only \$150 to \$300, one day of acute inpatient care in a municipal hospital costs nearly \$1000. *Id.*

242. *Id.* at 383.

243. Langemo, *supra* note 50, at 161-62.

revitalization and are encouraged to follow in the footsteps of the model programs described below. The promising results from mental health courts' increased emphasis on treatment, including reduced recidivism and economic savings, may create strong incentives for these governments to increase funding allocations to meet the mental health needs of juvenile offenders.²⁴⁴ Well-funded initiatives geared toward early identification of youthful offenders with mental disabilities would not only provide for more successful and humane treatment; they would also enable the juvenile justice system to reproduce economic and social benefits created by adult mental health courts, including decreased recidivism, a reduction in unnecessary detentions, and a better use of expensive detention beds.²⁴⁵ Although there is still little federal and state funding available for outpatient and at-home services, the community-based mental health services on which these models are based have repeatedly been shown to be both therapeutically effective and more economically efficient than institutional or residential treatment.²⁴⁶ Many new and innovative program models are now "designed with appropriate treatment and cost-effectiveness in mind."²⁴⁷ Dollars allocated today to meet the mental health needs of youth with mental disabilities "will be repaid many times over through lower public costs" by way of "reduction[s] in expensive long term health care, diminished need for welfare benefits, and less costly judicial processes," as well as corresponding increases in "educational achievement, employment opportunities, improved development of communities and the enhancement of family life."²⁴⁸ Moreover, as the mental health resource needs for offenders with serious mental disabilities are more precisely identified, the system will be better able to match available resources with existing mental health treatment needs and future resource development priorities, in the end producing "more effective longitudinal coordination of care and rehabilitation services."²⁴⁹

Overall, juvenile justice funding must be made adequate to support a "comprehensive continuum of child and family treatment and support services in communities," and flexible enough to allow for "the most appropriate placements of children and the most efficient use of available dollars."²⁵⁰ The bureaucratic distribution of current funding streams reinforces interagency competition rather than encouraging integrated cooperation.²⁵¹ "Flexible-funding," however, could

244. Kondo, *supra* note 81, at 310-11.

245. Arredondo et al., *supra* note 115, at 3.

246. See Cichon, *supra* note 189, at 543.

247. Nayowith, *supra* note 4, at 387.

248. *Id.* at 383.

249. Arredondo et al., *supra* note 115, at 4.

250. Nayowith, *supra* note 4, at 370.

251. Billikas, *supra* note 170, at 418.

attach money to an individual youth and his or her mental health treatment needs, encouraging “multisystem treatment for complex, troubled youths.”²⁵² This funding may be redirected from institutions to community-based services offering case management and “adequately fund[ed] services that prevent out-of-home placement.”²⁵³ The bulk of resources currently spent on ineffective and costly “institutions and residential placements can be used instead to pay for non-residential intensive supervision, the wraparound intervention strategy, and family therapeutic programs that have proven results.”²⁵⁴

Political pressure must also be placed on state legislators to increase funding for integrated juvenile justice and mental health initiatives.²⁵⁵ Advocates have successfully exercised such pressure in many jurisdictions, and states across the country have begun to fund juvenile delinquency control and prevention efforts, largely as block grants to counties or other municipalities.²⁵⁶ Block grants provide communities with the necessary resources and flexibility for “local control in program development, implementation and design.”²⁵⁷ Typically, incentives are included to reduce delinquency and curb the use of residential placements by treating offenders effectively in the community.²⁵⁸ Such grant programs have been viewed as overwhelmingly successful and are a step toward meeting the mental health needs of youthful offenders. A few of the more exemplary legislative initiatives—the Reasoned and Equitable Community and Local Alternatives to the Incarceration of Minors (RECLAIM) program in Ohio,²⁵⁹ the Virginia Juvenile Community Crime Control Act (VJCCCA),²⁶⁰ and the Schiff-Cardenas Crime Prevention Act of 2000²⁶¹—are discussed below.

252. *Id.* To implement such a funding system, existing funding sources must be reassessed and reorganized to direct streams toward juvenile offenders and the localities and community mental health providers that serve them. *See* Redding, *supra* note 31.

253. MD. JUVENILE JUSTICE COALITION, *supra* note 171, at 12.

254. *Id.* at 18.

255. Langemo, *supra* note 50, at 162.

256. *See* MD. JUVENILE JUSTICE COALITION, *supra* note 171, at 6.

257. *Id.*

258. *Id.*

259. *See* Ohio Dep’t of Youth Servs., RECLAIM Ohio, Reasoned and Equitable Community and Local Alternatives to the Incarceration of Minors, at <http://www.dys.ohio.gov/RECLAIMOhio.html> (last visited Sept. 13, 2004).

260. Virginia Juvenile Community Crime Control Act, VA. CODE ANN. §§ 16.1-309.2 to -309.10, -322.1 to -322.2, -322.4 (Michie 2004).

261. *See* DiGiovanni, *supra* note 128, at 4-5.

A. RECLAIM Ohio

RECLAIM Ohio offers Ohio counties the opportunity “to develop or purchase a range of community-based options to meet the needs of each juvenile offender or youth at risk of offending.”²⁶² Piloted in 1994 and implemented statewide in 1995, RECLAIM apportions juvenile court funding “for the local treatment of youthful offenders and at-risk youth,” with allocations “based on a four-year average of felony adjudications, with deductions for [the Department of Youth Services] and community corrections facility bed day usage in the prior year.”²⁶³ Paired with the Youth Services Grant initiative, monies received are “used for a vast array of treatment, intervention, diversion and prevention programs” including community-based treatment, intensive probation, and residential treatment.²⁶⁴ RECLAIM is designed both to improve the state Division of Youth Services’ treatment and rehabilitation efforts and to increase localities’ autonomy by giving juvenile court judges expanded sentencing options and community-based disposition alternatives.²⁶⁵

Overall, RECLAIM has been a successful program—institutional populations have decreased since its enactment, while localities have achieved a greater ability to meet the treatment needs of the juvenile offender population.²⁶⁶ Moreover, the program encouraged collaboration among a fragmented network of juvenile courts, the Division of Youth Services, and various other state agencies. In all, funds retrained pursuant to RECLAIM surpassed \$25 million in 1999²⁶⁷ and the program’s achievements promise to ensure similar levels of funding in years to come.

B. The VJCCCA

In 1994, the Virginia state legislature responded to an acknowledged lack of comprehensive mental health services by enacting the VJCCCA.²⁶⁸ The Act offers localities an opportunity to establish continuums of care and “an array of

262. Ohio Dep’t of Youth Servs., *supra* note 259.

263. *Id.*

264. *Id.*

265. Office of Juvenile Justice & Delinquency Prevention, Ohio: Sharing Responsibility for Administration of Juvenile Justice, at http://ojjdp.ncjrs.org/pubs/reform/ch3_d.html (last visited Sept. 13, 2004).

266. *Id.*

267. Ohio Dep’t of Youth Servs., *supra* note 262.

268. See VA. COMM’N ON YOUTH, THE STUDY OF JUVENILE JUSTICE SYSTEM REFORM 3, 69-71 (1996), [http://leg2.state.va.us/dls/h&sdocs.nsf/By+Year/HD371996/\\$file/HD37_1996.pdf](http://leg2.state.va.us/dls/h&sdocs.nsf/By+Year/HD371996/$file/HD37_1996.pdf).

pre- and post-dispositional services”²⁶⁹ for juvenile offenders designed by agency teams of local personnel²⁷⁰ by “develop[ing], implement[ing], operat[ing] and evaluat[ing] programs and services responsive to their specific juvenile offender needs and juvenile crime trends.”²⁷¹ In 2000, the VJCCCA provided nearly \$30 million in block grants to localities across Virginia to “support locally-designed community-based programs for court-involved youth.”²⁷² Funding allocations are based on a number of factors, including the number and nature of arrests and the average daily cost of serving a child, and—although the program is voluntary—all 134 cities and counties currently participate.²⁷³ The VJCCCA offers judges additional alternative sentencing options, additional funding for new and existing programs, and increased operational flexibility.²⁷⁴

C. The Schiff-Cardenas Crime Prevention Act

A similar block grant funding initiative emerged in California just six years after the VJCCCA project. Legislators hoping to reduce juvenile crime and delinquency enacted the Schiff-Cardenas Crime Prevention Act in 2000.²⁷⁵ The Act allocated \$121.3 million to localities to implement juvenile justice plans.²⁷⁶ To be eligible for funding, such plans must include assessments of existing community resources that “specifically target at-risk juvenile offenders, and their families;” identify and prioritize communities “fac[ing] a significant public safety risk from juvenile crime;” and provide for “a continuum of responses to juvenile crime and delinquency” demonstrating “a collaborative and integrated approach for implementing a system of swift, certain, and graduated responses for at-risk youth and juvenile offenders.”²⁷⁷ Participating localities must file annual reports detailing certain designated “outcome measures,” including the rate of juvenile arrests; the rates of successful completion of probation, restitution, and court-ordered community service; the arrest, incarceration, and

269. Robert E. Shepherd, Jr., *Legal Issues Involving Children*, 29 U. RICH. L. REV. 1117, 1118 (1995).

270. VA. COMM’N ON YOUTH, *supra* note 268, at 7.

271. Virginia Juvenile Community Crime Control Act, VA. CODE ANN. § 16.1-309.2 (Michie 2004).

272. VA. CRIME COMM’N, ASSESSMENT OF THE VIRGINIA JUVENILE COMMUNITY CRIME CONTROL ACT FORMULA AND THE ROLE OF OFFICES ON YOUTH, H.R. DOC. NO. 42, at 1 (2000), at [http://leg2.state.va.us/dls/h&sdocs.nsf/By+Year/HD422000/\\$file/HD42_2000.pdf](http://leg2.state.va.us/dls/h&sdocs.nsf/By+Year/HD422000/$file/HD42_2000.pdf).

273. See Virginia Juvenile Community Crime Control Act, VA. CODE ANN. § 16.1-309.2.

274. See *id.* § 16.1-309.2.

275. CAL. GOV’T CODE § 30061 (West 2004); DiGiovanni, *supra* note 128, at 4.

276. DiGiovanni, *supra* note 128, at 4.

277. CAL. GOV’T CODE § 30061(b)(4)(A)(i)-(iii).

probation violation rates of program participants; and the annual per capita costs of the program.²⁷⁸ To date, the juvenile justice plans would appear to have successfully offered California localities increased funding for innovative programs to meet juvenile mental health needs, including the nation's first and only juvenile mental health courts.²⁷⁹

D. Suggested Future Initiatives

Given the present and future successes of RECLAIM Ohio, the VJCCCA, and the Schiff-Cardenas Crime Prevention Act, other jurisdictions should follow suit by enacting similar community block grant programs. As one report suggested to the Commonwealth of Maryland, "reform is possible in the immediate future" with only "relatively modest increases in state funds."²⁸⁰ To begin, states may be able to simply reallocate funds spent on institutions to much smaller programs and community-based intervention strategies.²⁸¹ Eliminating reliance on institutions promises to offer long-term savings as recidivism falls and fewer youth are ordered into expensive institutional or residential placements.²⁸² Finally, states need not rely solely on their own treasuries to find resources to meet young offender's mental health needs in the juvenile justice system; many may increase their access to federal funds by relying on certain federally funded services like case management or by taking advantage of federal funds available for community-based services that help curb the high costs of institutional and residential care.²⁸³ With so many avenues available to increase funding for mental health treatment and services in the juvenile justice system, jurisdictions across the country should demonstrate their understanding of the importance of juvenile mental health needs by increasing the funding available to address these needs.

CONCLUSION

Mental health advocates who abhor the current systemic breakdown and increasingly punitive nature of juvenile courts have called on the juvenile justice system to follow the lead of criminal justice reforms and forge separate,

278. *Id.* § 30061(C)(i)-(iv).

279. See DiGiovanni, *supra* note 128, at 4-5, 12.

280. MD. JUVENILE JUSTICE COALITION, *supra* note 171, at 38.

281. See *id.* at 38-39.

282. See *id.* at 39.

283. *Id.* In particular, substantial funds may be available through a number of federal funding sources, including Social Security, Temporary Aid to Needy Families (TANF), and the Medicaid Targeted Case Management Program. *Id.* at 38.

specialized courts to deal exclusively with mentally-ill youth. Ultimately, however, this solution is sorely incomplete. While the small number of youth served by juvenile mental health courts might finally receive adequate consideration of their respective mental health treatment needs, these courts all but abandon the much larger contingent of children who either have less serious needs or have committed more serious offenses. In the end, I believe that the therapeutic justice principles and systemic treatment model reforms of the juvenile mental health court movement would be better applied in an intact, mainstream juvenile justice system. With state block grant programs in place, the promises of improved mental health treatment within this system are great. By not segregating children with defined or diagnosed mental disorders, but instead calling for a true and committed return to the juvenile court's individualized treatment model and greater rehabilitative ideal, many more young offenders will finally be able to obtain the mental health services they need—indeed deserve—from the juvenile justice system.

SYMPOSIUM

Pharmaceutical Innovation and Cost: An American Dilemma

- 713 Introduction**
Mark Siegler, M.D., Alix Weisfeld, and Richard A. Epstein, LL.B.
- 717 The Problem of New Uses**
Rebecca S. Eisenberg, J.D.
- 741 Regulatory Paternalism in the Market for Drugs: Lessons from Vioxx and Celebrex**
Richard A. Epstein, LL.B.
- 771 The Patient's Role in Choice of Medications: Direct-to-Consumer Advertising and Patient Decision Aids**
Marshall H. Chin, M.D., M.P.H.
- 785 The Effects and Role of Direct-to-Physician Marketing in the Pharmaceutical Industry: An Integrative Review**
Puneet Manchanda, M.Phil., Ph.D. and Elisabeth Honka

Introduction—Pharmaceutical Innovation and Cost: An American Dilemma

Mark Siegler, M.D.,* Alix Weisfeld,[†] and Richard A. Epstein, LL.B.[‡]

The four papers which follow, presented during an interdisciplinary symposium at the University of Chicago, respond to an atmosphere of growing public dissatisfaction with the pharmaceutical industry. The industry's problems include the rising cost of drugs, the slowing rate of innovation, concerns about the FDA's ability to effectively regulate the safety and efficacy of drugs, and the impact of the recently passed prescription drug benefit legislation. The past year has seen a flood of new books cataloguing these problems, including Marcia Angell's *The Truth About the Drug Companies: How They Deceive Us and What To Do About It*,¹ Jerry Avorn's *Powerful Medicines: The Benefits, Risks, and Costs of Prescription Drugs*,² Jerome Kassirer's *On the Take: How America's Complicity with Big Business Can Endanger Your Health*,³ and John Abramson's *Overdosed America: The Broken Promise of American Medicine*.⁴ Thus, despite a century of progress in developing safe and effective drugs that improve the length and quality of life, we are left with a fundamental dissatisfaction over the costs of medications and the rate of new drug development and innovation.

In November 2004, the University of Chicago's MacLean Center for Clinical Medical Ethics, John M. Olin Program in Law and Economics at the Law School, Committee on Clinical Pharmacology and Pharmacogenomics, and

* Lindy Bergman Distinguished Service Professor of Medicine and Surgery and Director, MacLean Center for Clinical Medical Ethics, University of Chicago.

† University of Chicago Law School J.D. expected 2006, MacLean Center for Clinical Medical Ethics.

‡ James Parker Hall Distinguished Service Professor of Law, The University of Chicago; Peter and Kirsten Bedford Senior Fellow, The Hoover Institution.

1. MARCIA ANGELL, *THE TRUTH ABOUT THE DRUG COMPANIES: HOW THEY DECEIVE US AND WHAT TO DO ABOUT IT* (2004).

2. JERRY AVORN, *POWERFUL MEDICINES: THE BENEFITS, RISKS, AND COSTS OF PRESCRIPTION DRUGS* (2004).

3. JEROME P. KASSIRER, *ON THE TAKE: HOW MEDICINE'S COMPLICITY WITH BIG BUSINESS CAN ENDANGER YOUR HEALTH* (2004).

4. JOHN ABRAMSON, *OVERDOSED AMERICA: THE BROKEN PROMISE OF AMERICAN MEDICINE* (2004).

Harris Graduate School of Public Policy Studies convened scholars from different disciplines to discuss these and related issues. The conference, titled "Pharmaceutical Innovation and Cost: An American Dilemma," was held at the Law School. A series of papers on innovation and regulation were presented at the conference, some of which were selected for publication in this volume.

The first paper in the series, *The Problem of New Uses* by Professor Rebecca Eisenberg,⁵ outlines the challenges our legal system faces when balancing the social cost of data and product exclusivity against providing adequate incentives for further research. In particular, she examines the need to promote corporate research into new uses of existing drugs at a time when drug companies fear negative results that call their drugs' safety into question (e.g. Cox-2 inhibitors) and the prospect of generic manufacturers free-riding on data from new trials. Eisenberg concludes that the best system would combine public disclosure of data with extended product exclusivity, an intent she believes the FDA should read into existing legislation. Professor Richard A. Epstein's paper, *Regulatory Paternalism in the Market for Drugs: Lessons from Vioxx and Celebrex*,⁶ criticizes the FDA for a paternalistic regulatory approach that privileges the safety of all potential drug users over informed individual choice. Epstein claims that the FDA has strayed from its primary mission of protecting consumers against impure substances and fraud. Epstein further argues that patients, not regulatory agencies, are in the best position to assess what risks are acceptable, and that the FDA's attempt to police the drug market solely through upstream regulation shows indifference to the opportunity cost of denying treatment to patients whose individual cost-benefit calculation counsels use of moderately risky drugs.

Both Eisenberg and Epstein emphasize the economies of information at play in the market for prescription drugs. The two final pieces, which examine aspects of pharmaceutical advertising, provide additional perspectives on this theme. The paper by Marshall Chin analyzes the way patients acquire information about drugs in the burgeoning era of direct-to-consumer advertising (DTCA).⁷ The period when patients received almost all of their information about which drugs to take from their physicians has now passed. Although his paper, *The Patient's Role in Choice of Medications: Direct-to-Consumer Advertising and Patient Decision Aids*, concludes that DTCA is appropriate, Dr. Chin contends patients

5. Rebecca S. Eisenberg, *The Problem of New Uses*, 5 YALE J. HEALTH POL'Y L. & ETHICS 211 (2005).

6. Richard A. Epstein, *Regulatory Paternalism in the Market for Drugs: Lessons from Vioxx and Celebrex*, 5 YALE J. HEALTH POL'Y L. & ETHICS 235 (2005).

7. Marshall Chin, *The Patient's Role in Choice of Medications: Direct-to-Consumer Advertising and Patient Decision Aids*, 5 YALE J. HEALTH POL'Y L. & ETHICS 265 (2005).

must diversify into other sources of information, including decision aids and clinician guidance, to make optimal choices. In their paper *The Effects and Role of Direct-to-Physician Marketing in the Pharmaceutical Industry: An Integrative Review*, Puneet Manchanda and Elisabeth Honka look at the relationship between detailing (informational sales pitches by drug company sales representatives to physicians) and physician prescribing behavior.⁸ The authors aggregate data from numerous empirical studies of detailing's impact on physician behavior and conclude that detailing's impact is most significant early in a drug's life cycle, but then declines as physicians gain access to other sources of information about the drug.

At a time when the pharmaceutical industry finds itself the focus of tremendous public attention, we hope that these papers offer some insight into how the industry can best fulfill its promise of safe, innovative, and moderately-priced drugs.

8. Puneet Manchanda & Elisabeth Honka, *The Effects and Role of Direct-to-Physician Marketing in the Pharmaceutical Industry: An Integrative Review*, 5 YALE J. HEALTH POL'Y L. & ETHICS 279 (2005).

The Problem of New Uses

Rebecca S. Eisenberg, J.D.*

INTRODUCTION

Discovering new uses for drugs that are already on the market seems like it ought to be the low-lying fruit of biopharmaceutical research and development (R&D). Firms have already made significant investments in developing these drugs and bringing them to market, including testing them in clinical trials, shepherding them through the FDA regulatory approval process, building production facilities, and training sales staff to market them to physicians. By this point, the drugs have begun to enjoy goodwill among patients and physicians and casual observations in the course of clinical experience may point to potential new uses. One might expect that firms would be well-motivated to invest in the further clinical trials necessary to market their products for new uses. But in practice, the legal and economic environment for drug development complicates firms' incentives to pursue this research. Examining the problem of motivating firms to invest in rigorous testing of new uses for previously approved drugs provides an interesting window on this environment.

Drugs are information-rich chemicals that in many respects are more akin to other information products (such as databases) than they are to other chemicals (such as industrial solvents). Drugs are chemicals that have been tested extensively to determine their safety and efficacy in treating disease. It is the information derived from such testing that distinguishes the chemicals we call "drugs" from similar chemicals sold for other purposes, or even for the same purposes.¹ Creating new molecules has become relatively cheap, but determining which molecules are safe and effective for which therapeutic purposes has

* Robert & Barbara Luciano Professor of Law, University of Michigan Law School.

1. For example, dietary supplements sold in health food stores may be used for the same purposes as drugs, but with considerably less information to support their safety and efficacy. Current U.S. law nonetheless permits their sale, subject to very limited regulatory oversight. *See* 21 U.S.C.A. §§ 321, 331, 342, 343, 350 (West 1999 & Supp. 2004); 42 U.S.C. §§ 281, 287c-11 (2000). *See generally* Ctr. for Food Safety & Applied Nutrition, FDA, Regulatory Strategy for the Further Implementation and Enforcement of the Dietary Supplement Health and Education Act of 1994 (Nov. 2004), at <http://www.cfsan.fda.gov/~dms/ds3strat.html>.

remained stubbornly expensive, time-consuming, and risky.² Information about the effects of drugs has considerable social value as a resource for guiding doctors, patients, and insurers to make sound choices about which therapeutic products to use. But drug-developing firms capture only a fraction of this value. Drug companies make money by selling drugs, not by selling information about the effects of drugs. Information from clinical trials may enhance sales of drugs if it indicates that they are safe and effective, but it may also cause sales to plummet if it indicates that they are unsafe or ineffective. The social value of negative information about drugs is captured by consumers, payors, and sellers of substitute products rather than by the seller of the drug under study. From the perspective of a firm that has a lucrative pharmaceutical product on the market, rigorous clinical trials of new indications present a risk of generating results that could destroy the value of the product rather than enhance it.

A recent case in point is Vioxx, a product that was approved by the FDA for treatment of pain and inflammation associated with osteoarthritis, menstruation, and rheumatoid arthritis.³ Vioxx sales were generating \$2.5 billion per year when the drug was taken off the market by its sponsor, Merck, following the revelation of serious adverse cardiovascular effects in the course of a trial of Vioxx for the prevention of recurrent colon polyps.⁴ Early clinical trials had suggested adverse

2. A recent study from the industry-funded Tufts Center for the Study of Drug Development estimates average costs to develop a new drug at \$802 million, using self-reported data and applying a discount rate of eleven percent to capitalize average out-of-pocket costs of \$ 403 million to the point of marketing approval. See Joseph A. DiMasi et al., *The Price of Innovation: New Estimates of Drug Development Costs*, 22 J. HEALTH ECON. 151, 165, 166 (2003). Critics immediately challenged this estimate as inflating the true costs. See Ceci Connolly, *Price Tag for a New Drug: \$802 Million; Findings of Tufts University Study Are Disputed by Several Watchdog Groups*, WASH. POST, Dec. 1, 2001, at A10; Press Release, Pub. Citizen, Tufts Drug Study Sample Is Skewed; True Figure of R&D Costs Likely Is 75 Percent Lower (Dec. 4, 2001), <http://www.publiccitizen.org/pressroom/release.cfm?ID=954>; see also MERRILL GOOZNER, *THE \$800 MILLION PILL: THE TRUTH BEHIND THE COST OF NEW DRUGS* (2004).

More recently, a Bain & Co. study estimated the average costs of drug development at more than twice the number calculated in the Tufts study, citing declining R&D productivity, rising costs of commercialization, increasing payor influence, and shorter exclusivity periods. See Jim Gilbert et al., *Rebuilding Big Pharma's Business Model*, IN VIVO: BUS. & MED. REP., Nov. 2003, at 1, http://www.bain.com/bainweb/PDFs/cms/Marketing/rebuilding_big_pharma.pdf. These cost estimates, which include research and development (R&D) costs of failed products as well as those directly attributable to successful products, are highly sensitive to the success rate for candidate products, rising when the success rate declines. The recent dearth of successful new products for the pharmaceutical industry thus inevitably increases the calculated costs per product.

3. See Ctr. for Drug Evaluation & Research, FDA Public Health Advisory: Safety of Vioxx (Sept. 30, 2004), http://www.fda.gov/cder/drug/infopage/vioxx/PHA_vioxx.htm.

4. See Barbara Martinez et al., *Merck Pulls Vioxx From Market After Link to Heart Problems*,

cardiovascular effects for Vioxx, but Merck took the position that the results were inconclusive and hoped that ongoing trials of the product for additional indications, culminating in supplemental FDA approval, would set these concerns to rest.⁵ Instead, further trials indicated that Vioxx did indeed significantly increase the risk of serious cardiovascular events.⁶ This is life-saving information that has considerable value from a public health perspective. Indeed, in a much-publicized study, one FDA scientist has estimated that from 1999 through 2003, approximately 27,000 heart attacks and sudden cardiac deaths could have been avoided if physicians had prescribed alternative medications instead of Vioxx.⁷ But from the perspective of Merck and its shareholders, this information has triggered a catastrophic loss of value.⁸ The social value of better information about the effects of drugs in patients can thus depart dramatically from its private value to firms that invest in clinical trials, making it difficult to rely on private markets to generate credible information. Profit-seeking firms face powerful incentives to develop and disclose information selectively, and perhaps even to delude themselves, in order to maximize product sales.⁹ Motivating firms to provide high quality information about the effects of

WALL ST. J., Oct. 1, 2004, at A1.

5. See Anna Wilde Mathews & Barbara Martinez, *E-Mails Suggest Merck Knew Vioxx's Dangers at Early Stage*, WALL ST. J., Nov. 1, 2004, at A1. The earlier results are reported in Claire Bombardier et al., *Comparison of Upper Gastrointestinal Toxicity of Rofecoxib and Naproxen in Patients with Rheumatoid Arthritis*, 343 N. ENG. J. MED. 1520 (2000).

6. The study results are reported in Robert S. Bresalier et al., *Cardiovascular Events Associated with Rofecoxib in a Colorectal Adenoma Chemoprevention Trial*, 352 N. ENG. J. MED. 1092, 1098 (2005). Although publicity has focused on the demonstration of cardiovascular risks, the study also found a reduction in the recurrence of colon polyps. See Gina Kolata, *Good Pill, Bad Pill: Science Makes It Hard To Decipher*, N.Y. TIMES, Dec. 22, 2004, at A1.

7. Memorandum from David G. Graham, Associate Director for Science, Office of Drug Safety, to Paul Seligman, Acting Director, Office of Drug Safety, Risk of Acute Myocardial Infarction and Sudden Cardiac Death in Patients Treated with Cox-2 Selective and Non-Selective NSAIDs (Sept. 30, 2004), <http://www.fda.gov/cder/drug/infopage/vioxx/vioxxgraham.pdf>; see Anna Wilde Mathews, *New Vioxx Study Projects Cases of Heart Attacks*, WALL ST. J., Oct. 6, 2004, at A2.

8. The results caused Merck to pull Vioxx off the market, forgoing \$2.5 billion a year in sales and causing its market capitalization to drop by \$26.8 billion. See Martinez, *supra* note 4.

9. Indeed, Merck faces allegations that it suppressed early warnings about the hazards of Vioxx. See Mathews & Martinez, *supra* note 5. Mathews and Martinez cite an internal email message from a Merck scientist that notes that "the possibility of increased CV [cardiovascular] events is of great concern"; adds parenthetically, "I just can't wait to be the one to present those results to senior management!"; and suggests that patients with high risk of cardiovascular problems be kept out of the study so that the difference between patients receiving Vioxx and the others "would not be evident." *Id.*

drugs in patients is thus a major challenge for the legal system.

In this Article, I examine three forms of legal regulation that affect the incentives of firms to invest in clinical trials: patents, FDA regulation, and trade secrecy. Although each of these legal regimes offers firms some protection from free riders who might otherwise use the information from clinical trials in competition with them, each has significant shortcomings as a regulatory mechanism for promoting the development of information about the effects of drugs through rigorous clinical trials.

Patent protection on drugs typically begins and ends too early to permit firms to capture the full value of subsequently developed information about drug effects. It therefore does a better job of motivating the initial R&D that is necessary to bring new products to market than it does of motivating the development of new information about old drugs. The discovery of a new use for an old drug might support a patent on a method of treatment, but such a patent offers little effective protection against generic competition once the drug itself is off-patent and may lawfully be sold for an older, unpatented use.

FDA-administered exclusivities do not begin to run until a drug is on the market, but they typically end before the expiration of patent protection. Additional exclusivity may later be obtained for conducting clinical trials of new uses of previously approved products, but like patents on new uses, these FDA-administered exclusive rights are limited to the new use and thus provide little protection from generic competition once the term of protection has expired for an older use of the same product. The most effective way that the FDA motivates investment in clinical trials is simply by demanding it as a precondition for approval of a New Drug Application (NDA). But once a drug is approved for a first indication, the permissibility of off-label sales dampens the incentives of firms to conduct further trials of additional indications. Such trials are not only costly, but also pose a risk of exposing previously unrecognized toxicities, thereby reducing rather than expanding product demand.

Trade secrecy mitigates this risk by allowing firms to suppress data from clinical trials, withholding its value not only from competitors but also from consumers who might otherwise demand less of the product. But trade secrecy greatly compromises the social value of the information as a resource for improving public health and for promoting further R&D. It also exposes drug companies and regulators to charges of bad faith and incompetence, compromising the signaling function of regulatory approval as a marker of safety and efficacy.

I. PATENTS

Patent law traditionally takes the lion's share of credit for motivating

investments in drug development. The pharmaceutical industry is famously dependent upon patent protection to support its R&D costs and has consistently advocated for stronger patent protection throughout the world.¹⁰ But patent law is better suited to protecting tangible products and processes than it is to protecting information. Although patent applicants are required to make enabling disclosures of how to make and use their inventions,¹¹ and judicial decisions celebrate the value of these disclosures as the *quid pro quo* for the patent right, the informational content of patent applications is generally treated as a spillover for the benefit of the public rather than as an object of protection in its own right.¹² Even as recent judicial decisions have opened up the patent system to protecting information technology,¹³ patents have remained unavailable for data.¹⁴

Nonetheless, patents on tangible products (such as drugs) and processes (such as methods of treatment) might motivate firms to invest in data production in order to develop markets for their inventions. Data from clinical trials of new uses might expand the market for drugs, and patents on drugs and methods of use might be used to exclude free riders from competing for these sales during the patent term. This allows firms to capture much of the value of successful trials that show their products to be safe and effective for particular purposes, although it does not allow them to capture the value of trials that show their products to be unsafe or ineffective. The value of data from unsuccessful trials accrues to consumers and insurance payors who forego purchasing the drug and perhaps also to competitors who develop and manufacture substitute products,¹⁵ all

10. See Robert Weissman, *A Long, Strange Trips: The Pharmaceutical Industry Drive To Harmonize Global Intellectual Property Rules, and the Remaining WTO Legal Alternatives Available to Third World Countries*, 17 U. PA. J. INT'L ECON. L. 1069, 1075-85 (1996).

11. 35 U.S.C. § 112 (2000).

12. See generally John N. Adams & Gwen Averly, *The Patent Specification: The Role of Liardet v. Johnson*, 7 J. LEG. HIST. 156 (1986) (reviewing eighteenth century decisions that emphasized the role of a robust patent disclosure standard in promoting the introduction of technical knowledge rather than merely the introduction of finished products).

13. *AT&T Corp. v. Excel Communications, Inc.*, 172 F.3d 1352 (Fed. Cir. 1999); *State St. Bank & Trust v. Signature Fin. Group*, 149 F.3d 1368 (Fed. Cir. 1998).

14. See Examination Guidelines for Computer-Related Inventions, 61 Fed. Reg. 7478 (Feb. 28, 1996). Copyright also provides no protection for data. See *Feist Publ'g Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340 (1991).

15. For example, the withdrawal of Vioxx from the market initially increased sales of Celebrex. See Scott Hensley, *Pfizer Is Early Winner as Vioxx Users Switch Drugs*, WALL ST. J., Oct. 6, 2004, at D13. Soon thereafter, however, the National Institutes of Health (NIH) suspended the use of Celebrex in clinical trials on the basis of data suggesting that it presents similar cardiovascular risks. See Press Release, NIH, NIH Halts Use of COX-2 Inhibitor in Large Cancer

without infringing the patent rights of the firm that paid for the trial.

Even for successful clinical trials, the term of the patent may be poorly timed to permit holders of patents on drugs to capture the value of the data, particularly for trials of new uses. Drug development necessarily involves the discovery of new compositions of matter before their therapeutic value can be definitively established through clinical trials.¹⁶ Patent law promotes early filing of patent applications through novelty and statutory bar standards that put dilatory applicants at risk of losing patent protection entirely.¹⁷ This leads inventors to file patent applications on new molecules as soon as they can establish patentable utility for them, typically years before first commercial marketing of a drug.¹⁸ Under current law,¹⁹ patents expire twenty years after their filing dates,

Prevention Trial (Dec. 17, 2004), <http://www.nih.gov/news/pr/dec2004/od-17.htm>; Press Release, NIH, Use of Non-Steroidal Anti-Inflammatory Drugs Suspended in Large Alzheimer's Disease Prevention Trial (Dec. 20, 2004), <http://www.nih.gov/news/pr/dec2004/od-20.htm>. Prescriptions for Celebrex promptly plummeted. See Christopher Windham & Ron Winslow, *Prescriptions for Celebrex, Naproxen Drop Sharply*, WALL ST. J., Dec. 30, 2004, at A3.

16. For a description of the drug development process, see CTR. FOR DRUG EVALUATION & RESEARCH, FDA, FROM TEST TUBE TO PATIENTS: IMPROVING HEALTH THROUGH HUMAN DRUGS (1999), <http://www.fda.gov/cder/about/whatwedo/testtube-full.pdf>. See also *In re Brana*, 51 F.3d 1560, 1568 (Fed. Cir. 1995) (noting that drugs are eligible for patent protection before they have met the standards for FDA approval).

17. A patent application is barred under § 102(b) of the Patent Act if the inventor fails to file within one year of first publication or other public use of the invention. 35 U.S.C. § 102(b) (2000). Moreover, the dilatory applicant who keeps the invention secret risks losing priority to another applicant who subsequently claims the same molecule if he is deemed to have "abandoned, suppressed, or concealed" the invention. *Id.* § 102(g)(1).

18. An invention must be useful in order to be patented. *Id.* § 101; see also *Brenner v. Manson*, 383 U.S. 519 (1966) (holding unpatentable a new method of making a new steroid where the steroid had not yet been shown to have a practical utility). But modern cases clarify that the showing of utility necessary to satisfy this requirement of patent law is far less than the showing of safety and efficacy required by FDA to bring a new drug to market. *E.g.*, *In re Brana*, 51 F.3d at 1567-68 ("The Commissioner . . . confuses the requirements under the law for obtaining a patent with the requirements for obtaining government approval to market a particular drug for human consumption. . . . FDA approval, however is not a prerequisite for finding a compound useful within the meaning of the patent laws. Usefulness in patent law, and in particular in the context of pharmaceutical inventions, necessarily includes the expectation of further research and development. The stage at which an invention in this field becomes useful is well before it is ready to be administered to humans.") (citations omitted).

19. The term of U.S. patent protection was changed in 1995 to bring U.S. law into compliance with the Agreement on Trade-Related Aspects of Intellectual Properties (TRIPS). See 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

regardless of when they issue.²⁰ The Hatch-Waxman Act of 1984 provides for patent term extensions of up to five years to compensate for some of the time that the patent meter is ticking pending regulatory approval of a new drug, so long as the total remaining patent life after extensions does not exceed fourteen years from the date of approval.²¹ A study of drugs approved between 1990 and 1995 showed an average “effective patent life” between product launch and patent expiration of 11.7 years, with somewhat longer lives appearing toward the end of the period under study.²² But sometimes the effective patent life for new drugs is far shorter, diminishing the time in which the basic drug patent permits a firm to capture the value of information it has generated about the drug.²³

URUGUAY ROUND vol. 31, 33 I.L.M. 1197 (1994); Uruguay Round Agreements Act, Pub. L. No. 103-465, 108 Stat. 4809 (1994) (codified as amended in pertinent part at 35 U.S.C.A. § 154 (West 2001 & Supp. 2004)).

20. 35 U.S.C. § 154(a)(2) (2000). For U.S. patent applications filed prior to 1995, the applicant may elect instead a term that begins with issuance of the patent and ends seventeen years later. *Id.* § 154(c)(1). The seventeen-year term sometimes permitted patent applicants to prosecute their claims lethargically in order to defer issuance and prolong the period of patent protection after products got to market. Some patent applicants developed this strategy to a fine art, splitting patent applications into multiple patents prosecuted in series to obtain staggered patent terms. Recently, the Federal Circuit has become skeptical of this and other “evergreening” strategies for prolonging patent protection for drugs and has found ways to hold the later-issued patents invalid. *See, e.g., Geneva Pharm., Inc. v. GlaxoSmithKline PLC*, 349 F.3d 1373 (Fed. Cir. 2003) (holding invalid later-issued patents deriving from the same parent application as expired patents on the antibiotic Augmentin on grounds of “double patenting”).

21. 35 U.S.C.A. § 156 (West 2001 & Supp. 2004). The period of extension may include half of the time spent in clinical trials before the firm submits a New Drug Application (NDA) to the FDA and all of the time that the NDA is pending before the FDA prior to approval, with provision for adjustment if the applicant did not act with due diligence. *Id.* § 156(c), (g)(1)(B), (g)(6).

22. Henry G. Grabowski & John Vernon, *Effective Patent Life in Pharmaceuticals*, 19 INT’L J. TECH. MGMT. 98 (2000).

23. For example, the antidepressant drug Paxil did not get to market until after its basic patent had expired. *See SmithKline Beecham Corp. v. Apotex Corp.*, 365 F.3d 1306, 1309 (Fed. Cir. 2004). Term extensions are unavailable after patents expire, 35 U.S.C. § 156(a)(1) (2000), although interim extensions may be obtained if it appears that the regulatory review period will extend beyond the term of the patent. *Id.* § 156(d)(5). The basic patent on a class of compounds including the molecule that was ultimately brought to market under the brand name Paxil, U.S. Patent No. 4,007,196 (issued Feb. 8, 1977), had a terminal disclaimer causing it to expire on October 14, 1992. (A terminal disclaimer is a surrender by the patent applicant of a portion of the patent term, usually entered to avoid a “double patenting” rejection of a patent that claims an obvious variation on a previously patented invention. *See Geneva Pharm.*, 349 F.3d at 1377-78. The terminal disclaimer causes the second patent to expire on the same date as the first, thereby avoiding an extension of the patent term through patenting essentially the same invention twice. *See In re Longi*, 759 F.2d 887, 894 (Fed. Cir. 1985).) SmithKline Beecham brought a hemihydrate form of Paxil to market in

New information about the uses of a product will sometimes allow the developer to get a process patent. For example, clinical trials showing that a drug works for a new indication may support a process patent on a new method of treatment, even though the same drug has previously been used for another purpose.²⁴ But process patent claims that are limited to particular therapeutic uses are generally considered less valuable than product patent claims covering the drug itself because the process claims cannot be used to stop competitors from selling the same product for other uses.²⁵ In theory, the patent-holder could still enforce the process patent against patients who take the drug for the patented use, doctors who prescribe it for such use, pharmacists who fill the prescriptions, or competing manufacturers who urge any of these actors to substitute their bioequivalent generic versions of the product for the patent-holder's product in such prescriptions.²⁶ But these remedies are generally less satisfactory than an injunction that would stop a competitor from making the product entirely. It is more difficult to detect and prove infringing uses than it is to detect and prove infringing products, and it is less efficient to sue numerous patients and physicians than it is to sue a single manufacturer. Moreover, few industries prosper by suing customers,²⁷ and the marketing interests of the pharmaceutical

1993, following FDA approval of its NDA on December 29, 1992. Meanwhile, the firm had obtained a separate patent on the hemihydrate form of the molecule, U.S. Patent No. 4,721,723 (issued Jan. 26, 1988). This subsequent patent was still in effect on the FDA approval date and the firm selected this later patent for term extension. See 35 U.S.C. § 156(c)(4) ("[I]n no event shall more than one patent be extended . . . for the same regulatory review period for any product."). The Federal Circuit ultimately held this patent invalid, reasoning that clinical trials more than a year prior to the filing date of the patent application had placed the invention in public use, giving rise to a statutory bar under 35 U.S.C. § 102(b). *SmithKline Beecham*, 365 F.3d at 1321. Historical information on the approval history of Paxil (and other drugs) is provided at Ctr. for Drug Evaluation & Research, FDA, Drugs @ FDA, at <http://www.accessdata.fda.gov/scripts/cder/drugsatfda/> (last visited Mar. 28, 2005) [hereinafter *Drugs @ FDA*].

24. See, e.g., *In re Marshall*, 578 F.2d 301, 304 (C.C.P.A. 1978) (reversing rejection of claim to method of using old compound to control weight, where prior art had disclosed method of using same compound to treat esophagitis, gastritis, peptic ulcer, and irritable colon syndrome, noting that "[i]f anyone ever lost weight by following the [prior art] teachings it was an unrecognized accident").

25. See, e.g., *Allergan Inc. v. Alcon Labs., Inc.*, 324 F.3d 1322 (Fed. Cir. 2003) (holding patent on new use of drug does not provide infringement remedy under Hatch-Waxman Act against generic competitor who seeks FDA approval to market same drug for a different use not covered by the patent); *Warner-Lambert Co. v. Apotex Corp.*, 316 F.3d 1348 (Fed. Cir. 2003) (same).

26. In the examples in text, the doctors, pharmacists, and manufacturers would be liable for actively inducing direct infringements by the patients themselves. 35 U.S.C. § 271(b).

27. A rare example of an intellectual property owner seeking to enforce its rights by suing customers is the Recording Industry of America, which has brought infringement actions against

industry are probably better served by soliciting physicians to write prescriptions than by suing them for contributory infringement of their patents.²⁸ Although patent-holders would rather sue generic competitors, sale of an unpatented product that is suitable for substantial non-infringing use is not patent infringement²⁹ unless the seller actively promotes an infringing use.³⁰ If the competitor merely brings the generic product to market for the old use, the fact that the product may be prescribed and used off-label for a patented new use is not enough to make the seller liable as an indirect infringer.³¹

II. FDA REGULATION

Although FDA regulation is typically understood to be a burdensome cost of drug development and rarely gets any credit for promoting biopharmaceutical R&D, FDA regulation in fact has come to play an important role in motivating firms to study the effects of drugs. FDA regulation fortifies the incentives of firms to invest in generating this socially valuable information in two ways: first, by requiring the submission of information as a precondition to bringing new products to market and to making marketing claims about products; and second, by conferring exclusive rights in the use of data submitted to the FDA for regulatory purposes. Because of their resemblance to the rights conferred by patents, I begin by considering the effects of FDA-administered exclusive rights.

A. FDA-Administered Exclusivities

FDA regulation sometimes provides patent-like rights in data from clinical trials by deferring approval of the products of generic competitors for the periods of time specified by statute. Some of these statutory provisions essentially provide for data exclusivity, deferring the time when other firms may rely on the pioneer's data in seeking regulatory approval for their own generic versions of the same drug,³² while others provide product market exclusivity in a new

individuals who download and share copyright-protected music. See Press Release, Recording Indus. Ass'n of Am., Music Industry Commences New Wave of Legal Action Against Illegal File Sharers (Dec. 3, 2003), <http://www.riaa.com/news/newsletter/120303.asp>.

28. For an unsettling account of the relationship between the pharmaceutical industry and the medical profession, see JEROME P. KASSIRER, *ON THE TAKE: HOW MEDICINE'S COMPLICITY WITH BIG BUSINESS CAN ENDANGER YOUR HEALTH* (2005).

29. 35 U.S.C. § 271(c).

30. *Id.* § 271(b).

31. See *C.R. Bard v. Advanced Cardiovascular Sys. Inc.*, 911 F.2d 670 (Fed. Cir. 1990).

32. See, e.g., 21 U.S.C.A. § 355(j)(5)(F)(ii), (iii) (West Supp. 2004).

product.³³ Data exclusivity can be the functional equivalent of product market exclusivity if the submission of data to the FDA is a condition for market entry and if the cost of regulatory compliance is prohibitive for a generic competitor.³⁴

The first statutory provision for FDA-administered exclusive rights in approved drugs was enacted as part of the Orphan Drug Act of 1983,³⁵ a legislative package designed to fortify incentives to develop treatments for rare diseases. The exclusivity provision of the Orphan Drug Act directs the FDA to grant seven years of market exclusivity for products to treat rare diseases and conditions affecting small populations,³⁶ later defined as fewer than 200,000 patients in the United States.³⁷ This is not merely a data exclusivity provision, but a statutory prohibition against approving another application for the same drug for the same disease for a period of seven years.³⁸ Although one might expect that products qualifying for this protection would have markets that are too small to be lucrative, in fact many products that enjoy exclusivity under the Orphan Drug Act have had large and profitable markets for off-label use.³⁹ The effect of market exclusivity under the Orphan Drug Act is similar to seven years of patent protection, although it does not preclude approval of either (1) another drug for the same disease or condition,⁴⁰ or (2) the same drug for another disease or

33. See, e.g., *id.* § 360cc(a) (West 1999 & Supp. 2004).

34. Indeed, even before Congress enacted the statutory exclusivity periods discussed in this section, FDA regulation provided significant protection from generic competition even after drugs went off-patent just by treating data from clinical trials as proprietary information belonging to the sponsor. See Ellen J. Flannery & Peter Barton Hutt, *Balancing Competition and Patent Protection in the Drug Industry: The Drug Price Competition and Patent Term Restoration Act of 1984*, 40 FOOD DRUG & COSMETIC L.J. 269, 273-76 (1985).

35. Pub. Law No. 97-414, 96 Stat. 2049 (1983) (codified as amended in various sections of 21, 26, 35, & 42 U.S.C.).

36. 21 U.S.C.A. § 360ee(b)(2).

37. *Id.* § 360cc(a).

38. Specifically, the statute prohibits approval of “another application . . . for such drug for such disease or condition for a person who is not the holder of such approved application . . . until the expiration of seven years from the date of the approval of the approved application . . .” *Id.*

39. Examples of blockbuster products that have received orphan drug status include Taxol and AZT. The FDA provides cumulative lists of orphan drug designations and approvals. FDA, List of Orphan Designations and Approvals, at <http://www.fda.gov/orphan/designat/list.htm> (last visited Feb. 20, 2005).

40. FDA regulations define the statutory term “such drug” to mean a drug with the same “active moiety” and not “clinically superior.” 21 C.F.R. § 316.3(b)(13) (2004). This potentially provides a narrower range of exclusivity than a patent, which can sometimes define the invention quite broadly with claim language extending to cover a genus of structurally similar molecules. *Cf.* Berlex Labs. v. FDA, 942 F. Supp. 19 (D.D.C. 1996) (rejecting a challenge under Orphan Drug Act to the FDA’s approval of a competitor’s slightly different version of a biological product).

condition.⁴¹

In 1984, Congress added two more provisions for FDA-administered market exclusivity in the Drug Price Competition and Patent Term Restoration Act of 1984, commonly known as the “Hatch-Waxman Act.”⁴² As part of a complex legislative compromise between the interests of research pharmaceutical firms and generic competitors, the Hatch-Waxman Act provided five years of exclusivity for new chemical entities not previously approved by the FDA⁴³ and three years of exclusivity for supplemental NDAs on previously approved products, such as new indications or other changes in a previously approved product that require conducting new clinical trials to win FDA approval.⁴⁴ In contrast to the exclusive rights to sell “such product for such use” conferred by the Orphan Drug Act, these Hatch-Waxman Act provisions merely confer data exclusivity, preventing the FDA from allowing generic competitors to obtain streamlined review of their applications through use of an abbreviated new drug application (ANDA) without having to submit a full new drug application.⁴⁵ Prior to passage of the Hatch-Waxman Act, generic competitors had faced prohibitive regulatory entry barriers when they were required to either conduct their own clinical trials of generic versions of their products or obtain permission to rely on data previously submitted by the brand name product manufacturer in order to get their products approved by the FDA. Because generic firms could not hope to recover this cost through sales at competitive prices, brand name drugs often continued to dominate the market even after their patents expired. In order to promote generic entry, the Hatch-Waxman Act provided that for off-patent drugs, generic versions could be approved upon a showing of bioequivalence to the previously approved product through use of an ANDA.

The five-year period of exclusivity for new chemical entities defers FDA approval of generic entry through the less costly ANDA route even if the product

41. This can be a significant limitation. *E.g.*, *Sigma-Tau Pharm. v. Schwetz*, 288 F.3d 141 (4th Cir. 2002) (holding that orphan drug exclusivity for new indication for levocarnitine did not preclude FDA approval of generic versions of same product for older indications for which exclusivity had expired, notwithstanding that generic versions might be prescribed by physicians off-label for new indication that was still covered by exclusivity). The FDA’s lack of authority over off-label use of drugs is discussed further *infra* notes 59–60.

42. Pub. Law No. 98-417, 98 Stat. 1585 (codified as amended in scattered sections of 21, 35 & 42 U.S.C.).

43. 21 U.S.C.A. § 355(j)(5)(F)(ii) (West Supp. 2004).

44. *Id.* § 355(j)(5)(F)(iii). This latter source of exclusivity might be available, for example, to a manufacturer that makes a change in the dosage form for a product, or seeks approval of a drug for new indications, or conducts clinical trials to determine whether a drug may safely be switched from prescription to over-the-counter (OTC) status.

45. The more extensive requirements for a full NDA are set forth in § 355(b)(1).

is not protected by patent, but it does not prevent a competitor from obtaining approval of an unpatented product if it is willing to go to the trouble and expense of conducting its own clinical trials and to rely strictly on its own data for proof of safety and efficacy.⁴⁶ In effect, this amounts to FDA-administered proprietary rights in data from clinical trials. Because the five-year period of data exclusivity for a new chemical entity begins with first market approval, it typically runs concurrently with patent protection. However, in some cases it may last longer,⁴⁷ providing a minimum five-year period of exclusivity even for unpatented products or for products that are covered by invalid patents.

The three-year period of data exclusivity for supplemental NDAs that require clinical trials to gain approval begins with the approval date of the supplemental NDA,⁴⁸ making it potentially advantageous to defer the filing of a supplemental NDA until a product approaches the end of its patent life in the hope of prolonging exclusivity. At that point, the firm might, for example, seek

46. The statute sets up a complex system for tracking patents covering approved drugs and for staying regulatory proceedings pending litigation of patent infringement claims. *See id.* § 355(b), (c), (j) (West 1999 & Supp. 2004). Holders of approved NDAs are required to disclose all patents that they believe would be infringed by unauthorized sales of the approved drug, and the FDA publishes the list in a publication called the *Orange Book*. CTR. FOR DRUG EVALUATION & RESEARCH, APPROVED DRUG PRODUCTS WITH THERAPEUTIC EQUIVALENCE EVALUATIONS [THE ORANGE BOOK] (24th ed. 2004), <http://www.fda.gov/cder/ob/docs/preface/ectablec.htm>. A competitor wishing to file an ANDA for a drug that is bioequivalent to the approved drug must make a declaration with respect to each of the patents listed in the *Orange Book* stating either (1) that the drug is not patented; (2) that the patent has expired; (3) that the patent will expire on a specified date; or (4) that the patent is either invalid or will not be infringed by the ANDA product (known as a "Paragraph IV certification"). If an ANDA filer makes a Paragraph IV certification, it must provide notice to the patent-holder and NDA filer (typically the same firm), along with a detailed statement of the factual and legal basis for the assertion that the patent is invalid or not infringed. § 355(j)(2)(B) (West Supp. 2004). The patent-holder then has forty-five days within which to bring an infringement action against the ANDA filer in order to prevent the FDA from approving the ANDA effective immediately under § 355(j)(5)(B)(iii). The Hatch-Waxman Act added to the Patent Act a new section, 35 U.S.C. § 271(c)(2) (2000), which makes it a technical act of patent infringement to file an ANDA for a drug claimed in a patent or the use of which is claimed in a patent. This was necessary in order to permit litigation of the issue of patent infringement before the generic product got to market, because Congress declared in § 271(e)(1) that use in clinical trials was not an act of patent infringement. If the patent-holder brings an infringement action within forty-five days, that triggers a thirty-month stay of FDA approval for the ANDA under 21 U.S.C.A. § 355(j)(5)(B)(iii) (West Supp. 2004) while the parties litigate the infringement issue.

47. *See supra* note 23 and accompanying text (discussing Paxil).

48. § 355(j)(5)(F)(iii).

approval to switch a product from prescription to over-the-counter (OTC) sales,⁴⁹ after first testing the product in patients to determine if they may safely self-administer the drug without the supervision of a physician in order to qualify for the additional period of exclusivity. A supplemental NDA may also be used to get approval to market a previously approved drug for a new use. Either way, the data exclusivity thereby gained is limited to the terms of the new approval and will not prevent a competitor from using an ANDA to gain approval to sell the product as previously approved, or for previously approved indications.

This has proven to be a very significant limitation on the benefit of using a supplemental NDA to gain approval to market a drug for a new indication.⁵⁰ The three-year exclusivity does not preclude a generic competitor from using an ANDA to get approval to sell its version of the product for the original indication; further, once the generic version is available on the market, the FDA can do nothing to stop physicians from prescribing the generic product off-label for the new indication. Indeed, unless the new indication involves a different formulation of the product, state generic substitution laws may force the original innovator to lower its prices to meet the generic price to avoid substitution at the point of filling the prescription.⁵¹

The exclusivity that comes with a supplemental NDA is more effective in thwarting generic competition for a prescription to OTC switch. Gaining FDA approval to sell a drug in the OTC market will not preclude a generic competitor from filing an ANDA to sell the same product by prescription, but it may be difficult for the prescription generic to compete with the OTC branded product. Moreover, consumers may be more likely to select brand name products in the OTC market, while doctors and pharmacists, facing pressure from insurers to keep costs down, may be more likely to substitute cheaper generics in the prescription drug market.⁵²

The Food and Drug Administration Modernization Act of 1997⁵³ added a

49. The strategic considerations behind the timing of these moves are laid bare in studies by consulting firms that are posted on the internet. See, e.g., Kline & Co., *Impending Wave of Rx-to-OTC Switches Offers Significant Opportunities for Drug Companies* (Aug. 15, 2002), at http://www.klinegroup.com/6_2002815.htm.

50. See *Bristol-Myers Squibb v. Shalala*, 91 F.3d 1493 (D.C. Cir. 1996).

51. See Henry G. Grabowski & John M. Vernon, *Substitution Laws and Innovation in the Pharmaceutical Industry*, 43 J.L. & CONTEMP. PROBS. 43, 49-56 (1979).

52. Indeed, some insurers do not provide coverage of brand name products if generic equivalents are available. See, e.g., Univ. of Mo., University of Missouri Faculty & Staff Benefits: *Mandatory Generic Drug Substitution*, at <http://www.umsystem.edu/hrs/benefits/prescription/generic.htm> (last updated Oct. 15, 2004).

53. Pub. Law No. 105-115, 111 Stat. 2296 (codified as amended in scattered sections of 21, 26 & 42 U.S.C.). Although this provision was originally set to expire after five years, it has been

provision for six months of exclusivity as a reward for conducting pediatric trials of drugs.⁵⁴ This six-month period of exclusivity is not contingent upon approval of the drug as safe and effective in children and is not limited to pediatric use of the drug. It simply extends any existing market exclusivity held by the submitter, whether under a patent, the Orphan Drug Act, or Hatch-Waxman exclusivity provisions, further deferring the time when the FDA might approve a competing generic product.

Each of these provisions confers exclusionary rights under the auspices of the FDA rather than the U.S. Patent and Trade Office. The FDA-administered rights are linked to submission and consideration of data from clinical trials of drugs for safety and efficacy and have the effect of rewarding firms that invest in rigorous clinical trials by protecting them from competition. But there are gaps in the scope of exclusion, particularly in the context of clinical trials of new uses of previously approved products. The exclusive rights provided to firms that file supplemental NDAs for new uses do not preclude generic competitors from gaining FDA approval to sell the same products for previously approved uses; and once generic versions of these products are available, the FDA has no authority to prevent doctors and pharmacists from substituting the generic version off-label for the branded version sold by the holder of the supplemental NDA. Therefore, the three-year exclusivity provision for supplemental NDAs is likely to have little effect on incentives to conduct clinical trials of new uses of previously approved drugs.

B. FDA as Market Gatekeeper

A far more significant way that the FDA motivates firms to conduct rigorous clinical trials is by demanding data from clinical trials in its market gatekeeper role. The FDA is charged by statute with keeping new drugs off the market pending the submission of the results of “adequate and well-controlled investigations” indicating that they are safe and effective for their intended use.⁵⁵ FDA regulation gives firms powerful incentives to test their products thoroughly enough to satisfy rigorous scientific standards of safety and efficacy for at least one indication. In order to get an NDA approved, a firm must submit “full reports of investigations which have been made to show whether or not such drug is safe for use and whether such drug is effective in use.”⁵⁶ The statute repeatedly refers to the intended use of the drug in defining the standard for approval, indicating

extended. See Best Pharmaceuticals for Children Act of 2002, Pub. L. No. 107-109, 115 Stat. 1408 (codified as amended in scattered sections of 21 & 42 U.S.C.).

54. 28 U.S.C.A. § 355a (West 2001 & Supp. 2004).

55. 21 U.S.C. § 355(d) (2000).

56. *Id.* § 355(b)(1)(A).

that determinations of safety and efficacy are meaningful only with respect to a particular intended use. It thus directs the Secretary⁵⁷ to reject the NDA if the submitted reports “do not include adequate tests by all methods reasonably applicable to show whether or not such drug is safe for use *under the conditions prescribed, recommended, or suggested in the proposed labeling* thereof,”⁵⁸ if “the results of such tests show that such drug is unsafe for use *under such conditions* or do not show that such drug is safe for use *under such conditions*,”⁵⁹ or if “there is a lack of substantial evidence that the drug will have the effect it purports or is represented to have *under the conditions of use prescribed, recommended, or suggested in the proposed labeling* thereof.”⁶⁰ The central role of the particular indication that is being tested carries over into the statutory definition of “substantial evidence” as

evidence consisting of adequate and well-controlled investigations, including clinical investigations, by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, on the basis of which it could fairly and responsibly be concluded by such experts that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling or proposed labeling thereof.⁶¹

But once a new drug gets to market, the FDA does not prevent its off-label use for other indications that have never been tested. The FDA does not regulate the practice of medicine, and doctors are free to prescribe approved drugs as they see fit.⁶² This limits significantly the incentives of firms to continue testing their products for new uses once their NDAs have been approved, with a corresponding gap in the quality of data supporting the safety and efficacy of drugs for new uses. For many lucrative drugs, off-label sales account for a significant portion of sales.⁶³

57. The statute confers regulatory authority upon the Secretary of Health and Human Services, although the Secretary turns to FDA to make the necessary judgments.

58. 21 U.S.C. § 355(d)(1) (emphasis added).

59. *Id.* § 355(d)(2) (emphasis added).

60. *Id.* § 355(d)(5) (emphasis added).

61. *Id.* § 355(d) (emphasis added).

62. *See, e.g.,* FTC v. Simeon Mgmt. Corp., 391 F. Supp. 697, 706-07 (N.D. Cal. 1975), *aff'd*, 532 F.2d 708, 717 (9th Cir. 1976). Off-label prescribing of drugs is a significant part of medical practice in some specialties, including oncology. *See* Lars Noah, *Informed Consent and the Elusive Dichotomy Between Standard and Experimental Therapy*, 28 AM. J.L. & MED. 361, 397-98 (2002).

63. A much-cited example is the drug Gabapentin, approved by the FDA for adjunctive therapy in the treatment of partial seizures and postherpetic neuralgia and prescribed off-label for other indications representing as much as ninety-five percent of sales. *See* Alicia Mack,

Rigorous clinical trials of new uses of previously approved products are not only costly, but can also be extremely risky for a firm that has a lucrative product on the market. A conspicuous example of the risks that rigorous clinical trials pose to a drug manufacturer that is already enjoying brisk off-label sales can be found in the National Institutes of Health (NIH) Women's Health Initiative study on the effects of hormone replacement therapy (HRT) on the risk of heart disease in post-menopausal women.⁶⁴ Although the FDA had only approved the use of HRT for relief of menopause symptoms, prior observational studies had suggested that women who take HRT have a lower risk of heart disease. Even without further FDA approval, this evidence brought about widespread off-label prescription and use of HRT for the purpose of preventing heart disease. HRT manufacturers, although formally prohibited from actively promoting HRT for this purpose, nonetheless enjoyed significantly expanded sales from prescriptions in reliance on the results of the prior observational studies and stood to gain little from subjecting doctors' and patients' beliefs to more rigorous tests. When NIH (not the manufacturer) finally conducted a long-term, randomized, controlled study involving over 16,000 patients, the results indicated an *increased* risk of heart disease (as well as increased risks of other diseases) in women receiving HRT. This information is undoubtedly valuable to patients, physicians, health insurers, and policy makers, but it sharply reduced sales of Prempro.⁶⁵ In this case, government funding provided valuable and credible information that the product's manufacturer had little incentive to uncover on its own.⁶⁶

Examination of the Evidence for Off-Label Use of Gabapentin, 9 J. MANAGED CARE PHARMACY 559, 559 (2003), <http://www.amcp.org/data/jmcp/Contemporary%20Subject-559-568.pdf>. The manufacturer, Warner-Lambert (now owned by Pfizer) ultimately pled guilty to charges of illegal promotion of the drug for off-label uses and settled the case for \$430 million. See Bernadette Tansey, *Huge Penalty in Drug Fraud: Pfizer Settles Felony Case in Neurontin Off-Label Promotion*, S.F. CHRON., May 14, 2004, at C1.

64. Writing Group for the Women's Health Initiative Investigators, *Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results from the Women's Health Initiative Randomized Controlled Trial*, 288 JAMA 321 (2002).

65. According to a front page story in the *New York Times*, the manufacturer of Prempro (Wyeth) estimates that the number of women taking Prempro fell from 2.7 million to 1.5 million following the announcement of the study results. Gina Kolata et al., *Menopause Without Pills: Rethinking Hot Flashes*, N.Y. TIMES, Nov. 10, 2002, § 1, at 1.

66. One might imagine that health insurers or HMOs would be motivated to conduct clinical trials of drugs to determine their value and to decide whether to pay for them. Insurers presumably have access to patient populations and medical records that place them in a good position to observe the relative benefits and harms of different treatments and they sometimes make such data available for studies. See, e.g., David J. Graham et al., *Risk of Acute Myocardial Infarction and Sudden Cardiac Death in Patients Treated with Cyclo-Oxygenase 2 Selective and Non-Selective Non-Steroidal Anti-Inflammatory Drugs: Nested Case-Control Study*, 365 THE LANCET 475 (2005)

Although the FDA has no authority to prevent prescriptions of approved drugs for off-label uses, it has some statutory authority over the marketing claims that may be made on behalf of such drugs by the manufacturers and has sometimes sought to use this authority to prevent firms from promoting drugs for off-label uses. Firms have resisted this form of regulation, arguing with some success in the courts that it violates their First Amendment rights to disseminate information about their products to physicians.⁶⁷ For example, in *Washington Legal Foundation v. Friedman*,⁶⁸ an industry-supported nonprofit raised a First Amendment challenge to FDA guidance documents from the early 1990s that restricted manufacturer promotion of off-label uses for approved drugs and devices through distribution of reprints of publications and through manufacturer involvement in continuing medical education programs.⁶⁹ The FDA claimed that distribution of these materials by product manufacturers amounted to unapproved “labeling” that rendered these products “misbranded” in violation of the federal Food, Drug & Cosmetic Act (FDCA). The district court concluded that the regulated activities were commercial speech and put the burden on the FDA to show that the regulation was no more extensive than necessary to advance a substantial government interest.⁷⁰ The FDA advanced two interests in support of its regulation: (1) ensuring that physicians receive accurate and unbiased information so that they may make informed prescription choices; and (2) providing manufacturers with ample incentive to get previously unapproved uses “on label” by testing them and submitting them to the FDA for approval. The court concluded that the first interest was inadequate to justify the intrusion on speech, but that the second interest was substantial.⁷¹ Ultimately, the FDA revised its guidance documents to permit firms to distribute reprints of journal

(basing study on data from Kaiser Permanente). They might also be in a good bargaining position to require proof of safety and efficacy from drug manufacturers as a precondition to covering their products. By withholding coverage of off-label prescriptions, they sometimes play a role in demanding such information.

67. In recent years courts have invalidated regulatory and statutory restrictions on the promotion of pharmaceutical products on First Amendment grounds. *See, e.g.,* *Thompson v. W. States Med. Ctr.*, 535 U.S. 357 (2002); *Wash. Legal Found. v. Henney*, 202 F.3d 331 (D.C. Cir. 2000); *Wash. Legal Found. v. Friedman*, 13 F. Supp. 2d 51 (D.D.C. 1998).

68. 13 F. Supp. 2d 51.

69. *See* Draft Policy Statement on Industry-Supported Scientific and Educational Activities, 57 Fed. Reg. 56,412 (Nov. 27, 1992); David G. Adams, *FDA Policy on Industry-Supported Scientific and Educational Activities: Current Developments*, 47 FOOD & DRUG L.J. 629 (1992).

70. 13 F. Supp. 2d at 65, 69-74.

71. Although the regulations set forth in the FDA Guidance Documents directly advanced this interest, the court concluded that they were more extensive than necessary because this interest could be addressed in a less burdensome manner by simply requiring full disclosure. *Id.* at 72-74.

articles regarding off-label uses, effectively permitting some marketing of drugs for unapproved uses without the risk and expense of the sort of trials that are necessary to satisfy the FDA.⁷²

One would expect this change in regulations to diminish the incentives of firms to conduct rigorous clinical trials of previously approved products for new uses. Moreover, the current administration has shown notably less inclination to enforce restrictions on marketing claims against the pharmaceutical industry,⁷³ further minimizing the force of remaining restrictions. Nonetheless, some firms continue to conduct post-marketing studies of approved drugs in the hope of getting supplemental NDAs approving uses for new indications, despite the costs and risks.

Merck's trial of Vioxx for the supplemental indication of preventing recurrence of colon polyps is a striking recent example.⁷⁴ Why would Merck put its revenues from a successful product at risk by conducting such a trial? Extensive media attention to Vioxx in recent months offers a rare glimpse behind the scenes of such decisions.⁷⁵ Presumably Merck hoped to expand the market for Vioxx to include patients at risk of recurring colon polyps, rather than limiting sales to ulcer-prone patients with arthritis and menstrual cramps, and additionally hoped that the post-marketing study would show that the drug was safe and effective for this lucrative new indication. Of course, Merck might have attempted to generate off-label sales for this indication without going to the trouble of conducting the sort of trial that would meet with FDA approval of a supplemental NDA, perhaps by conducting more limited studies and circulating reprints.⁷⁶ However, a prophylactic indication against a relatively low risk might be a hard enough sell for an expensive drug to make the constraints on off-label marketing problematic.⁷⁷ A similar study was already underway for Pfizer's rival

72. Dissemination of Information on Unapproved/New Uses for Marketed Drugs, Biologics, and Devices, 63 Fed. Reg. 64,556 (Nov. 20, 1998). The revised guidance documents followed explicit Congressional authorization for manufactures to distribute publications that discuss off-label uses as part of the FDA Modernization Act of 1997, Pub. L. No. 105-155, § 401(a), 111 Stat. 2296, 2356-64 (1997) (codified at 21 U.S.C. §360aaa (2000)).

73. See MINORITY STAFF OF HOUSE COMM. ON GOV'T REFORM, 108TH CONG., FDA ENFORCEMENT ACTIONS AGAINST FALSE AND MISLEADING PRESCRIPTION DRUG ADVERTISEMENTS DECLINED IN 2003 (Comm. Print 2004).

74. See *supra* notes 3-9 and accompanying text.

75. See, e.g., Alex Berenson et al., *Despite Warnings, Drug Giant Took Long Path to Vioxx Recall*, N.Y. TIMES, Nov. 14, 2004, at A1; Martinez et al., *supra* note 4; Mathews & Martinez, *supra* note 5.

76. See *supra* note 72 and accompanying text.

77. *Id.*

product Celebrex,⁷⁸ threatening to put Merck at a marketing disadvantage if Celebrex were approved for an indication that remained off-label for Vioxx.

Recent newspaper accounts also suggest that early concerns about the safety of Vioxx may have fortified Merck's resolve to pursue studies of additional indications.⁷⁹ There were indications that Vioxx presented an increased risk of cardiovascular events in data from an early study comparing Vioxx to naproxen,⁸⁰ although Merck took the position at the time that the difference reflected a protective effect of naproxen rather than a toxic effect of Vioxx.⁸¹ Nonetheless, both Merck and the FDA thought the cardiovascular effects of Vioxx called for further study, although they agreed that it would be difficult and ethically problematic to design a clinical trial that would compare Vioxx and a placebo in at-risk patients solely for the purpose of observing side effects.⁸² According to *Wall Street Journal* reporters, Merck marketing executives also opposed a study of cardiovascular risks out of concern that it would signal a lack of confidence in Vioxx.⁸³ Instead, Merck scientists decided, in consultation with the FDA, to await further data on cardiovascular effects of Vioxx from ongoing studies of new indications, signaling optimism about future markets rather than concerns about side effects. Meanwhile, as more data came in, the FDA reached an agreement with Merck to disclose cardiovascular risks in the product labeling in 2002.⁸⁴ Perhaps Merck hoped that rigorous long-term studies, culminating in FDA approval of a supplemental NDA, would put these concerns to rest while expanding the market for its product. Ultimately, of course, that is not what happened. But although the trials were a failure from the perspective of Merck and its shareholders, this episode suggests that the current combination of regulatory carrots and sticks can sometimes motivate firms to undertake very risky investments in clinical trials of their products for new uses.

78. See Scott D. Solomon et al., *Cardiovascular Risk Associated with Celecoxib in a Clinical Trial for Colorectal Adenoma Prevention*, 352 NEW ENG. J. MED. 1071, 1072 (2005). The Celebrex trial began enrolling patients in November 1999 and stopped administering the study drug on December 16, 2004 after data analysis revealed increased cardiovascular risks to patients receiving the drug. The similar Vioxx trial began enrolling patients in February 2000 and was terminated on September 30, 2004. See Robert S. Bresalier et al., *supra* note 6.

79. Berenson et al., *supra* note 75.

80. See Claire Bombardier et al., *Comparison of Upper Gastrointestinal Toxicity of Forecoxib and Naproxen in Patients with Rheumatoid Arthritis*, 343 NEW ENG. J. MED. 1520 (2000).

81. *Id.* at 1526.

82. Berenson et al., *supra* note 75.

83. *Id.*; see also Mathews, *supra* note 7.

84. See Press Release, FDA, FDA Issues Public Health Advisory on Vioxx as its Manufacturer Voluntarily Withdraws the Product (Sept. 30, 2004), <http://www.fda.gov/bbs/topics/news/2004/NEW01122.html>; see also Mathews & Martinez, *supra* note 5.

III. TRADE SECRECY

From a public policy perspective, the most problematic form of legal protection for data from clinical trials is trade secrecy. Although the pharmaceutical industry has long taken the position that the data from clinical trials of drugs constitute proprietary trade secret information, trade secrecy severely restricts the social value of this information by giving patients and care providers access to only as much of the data as the trial's sponsor chooses to reveal. The FDA has consistently supported this position⁸⁵ and withheld the data from public disclosure as a matter of administrative practice,⁸⁶ although the statutory language invoked in support of this position is ambiguous.⁸⁷ Amendments to the FDCA as part of the Hatch-Waxman Act of 1984 appeared to require that safety and effectiveness data for a drug be made available to the public, "unless extraordinary circumstances are shown," as soon as the periods of data exclusivity have expired and an ANDA "could be made effective if such an

85. Although the FDA does not disclose the underlying data, it requires disclosure of certain information in the labeling of approved products. 21 C.F.R. pt. 201 (2004). Moreover, in recent years the FDA has begun putting more information about approved products up on its website, including analyses of the data from clinical trials by FDA staff. *See, e.g., Drugs @ FDA, supra* note 23.

86. *See, e.g., Anderson v. Dep't of Health & Human Servs.*, 907 F.2d 936 (10th Cir. 1990); *Pub. Citizen Health Research Group v. FDA*, 997 F. Supp. 56 (D.D.C. 1998); 42 Fed. Reg. 3094, 3106 (Jan. 14, 1977) (noting that the FDA has treated data from clinical trials as a trade secret since 1938); 39 Fed. Reg. 44,601, 44,612 (Dec. 24, 1974) ("The Food and Drug Administration has on numerous occasions testified before Congress that current statutory prohibitions prevent disclosure of useful information contained in the agency's files, and particularly, data relating to the safety and effectiveness of drugs. The Food and Drug Administration cannot change the law, and thus is bound by the present provisions until Congress acts.").

87. Proponents of trade secrecy have relied upon section 301(j) of the Federal Food, Drug, and Cosmetic Act (FDCA), 21 U.S.C.A. § 331(j) (West Supp. 2004), which prohibits:

The using by any person to his own advantage, or revealing, other than to the Secretary or officers or employees of the Department, or to the courts when relevant in any judicial proceeding under this Act, any information acquired under authority of section . . . 355 . . . concerning any method or process which as a trade secret is entitled to protection.

Id. It is by no means obvious from the statutory language that "any method or process which as a trade secret is entitled to protection" includes data from clinical trials, although by now longstanding administrative practice would make it difficult to adopt a narrower reading of the provision. *See James T. O'Reilly, Knowledge Is Power: Legislative Control of Drug Industry Trade Secrets*, 54 U. CIN. L. REV. 1 (1985); Richard S. Fortunato, Note, *FDA Disclosure of Safety and Efficacy Data: The Scope of Section 301(j)*, 52 FORDHAM L. REV. 1280 (1984).

application had been submitted.”⁸⁸ However, so far the industry has successfully resisted a plain meaning interpretation of this provision.⁸⁹

Trade secrecy and FDA regulation are intertwined at a number of levels. At least as a historical matter, an important component of the value of safety and effectiveness data from the perspective of drug manufacturers lay in its utility in overcoming regulatory entry barriers.⁹⁰ The FDCA requires the submission of “full reports”⁹¹ of clinical trials to comply with the requirements for an NDA, which has long been understood to require submission of the underlying data rather than just published summaries. If competitors could gain access to the data, they could use it to submit their own NDAs to the FDA to bring generic versions of previously approved products to market without having to incur the cost and risk of doing their own trials.

This concern about free riders using publicly available data to get approval to sell a generic product in competition with a pioneer was arguably more substantial prior to the Hatch-Waxman Act than it is today. Under current law, pioneers are substantially protected from generic entry during the statutory periods of data exclusivity by the inability of competitors to use an ANDA during that time.⁹² Moreover, current law directs the FDA to stay the approval of competing products that are covered by patents listed in the *Orange Book* for at least thirty months following a challenge by the patent owner, or until the expiration or successful challenge to the validity of the listed patents.⁹³ It is possible that a generic competitor might use publicly available data to submit its own NDA prior to the end of the data exclusivity period if all listed patents have expired or are invalid, but the Hatch-Waxman Act does not require public disclosure until the time when an ANDA could become effective.⁹⁴ The FDA will not approve a generic product on the basis of an ANDA until applicable data exclusivity periods and patents have expired. At that point, with or without disclosure of the underlying data, current law permits free riding on prior studies through use of an ANDA. The generic firm need only show that its product is bioequivalent to a previously approved product and has no regulatory need to

88. Hatch-Waxman Act, Pub. L. No. 98-417, § 104, 98 Stat. 1597 (1984) (codified as amended at 21 U.S.C. § 355(l) (2000)).

89. See Jane A. Fisher, *Disclosure of Safety and Effectiveness Data Under the Drug Price Competition and Patent Term Restoration Act*, 41 FOOD DRUG & COSM. L.J. 268, at 277-86 (1986).

90. O'Reilly, *supra* note 87.

91. See *supra* note 53 and accompanying text.

92. 21 U.S.C.A. § 355(j)(5)(F)(ii) (West Supp. 2004).

93. *Id.* §§ 355(c)(3), (j)(5)(B). A court before which the patent litigation is pending has some latitude to modify the period of the stay under the terms of the statute. See *supra* notes 39-43 and accompanying text.

94. § 355(l)(5) (West 1999 & Supp. 2004).

replicate the data previously submitted by the holder of the original NDA. By permitting substantial free riding even without access to the underlying data, the Hatch-Waxman Act has thus taken the wind out of the sails of an argument against data disclosure that rests upon protection from free riders.⁹⁵

Apart from this much-reduced value to drug manufacturers in overcoming regulatory barriers, data from clinical trials may be valuable to competitors in guiding their own R&D. The data may, for example, alert firms to hazards associated with a class of products, highlight the relative virtues of competing products, or point to potential new uses that merit further investigation, thereby allowing them to deploy their own R&D resources more efficiently. Trade secrecy permits firms to withhold this value from competitors while exploiting it themselves; however, it does so at considerable social cost. Public availability of data from clinical trials would allow firms to learn from each other's experience so that they could design better products and conduct better trials in the future. It would spare firms from having to continuously reinvent the wheel and steer them away from carrying out costly trials of products that are likely to fail, thereby perhaps bringing down the staggering average costs of new drug development.⁹⁶ It would also permit reanalysis of data by skeptical competitors in ways that might challenge the spin selected by the product's sponsor and facilitate meta-analysis of aggregated data from multiple studies of related products. The foregone social value of undisclosed data from clinical trials is likely to be a growing loss, as information technology improves and as growing understanding of the genetic basis of disease and drug response makes it possible to direct queries to data from multiple studies of different drugs in different patients. The FDA is sitting on a treasure trove of data for such purposes.

Public availability of data from clinical trials would also be valuable for patients, doctors, and insurers, permitting them to make better choices of drugs. To the extent that data disclosure is valuable to these customers, one might expect firms to have some motivation to provide it. Indeed, trade secrecy is a tricky strategy for information-rich products like drugs, because firms need to make some disclosure of product information in order to capture its value. On the other hand, firms might be reluctant to disclose negative data that would diminish sales of their products. Trade secrecy allows firms to pursue a strategy of selective disclosure of favorable information from clinical trials, although presumably with some loss of credibility for their claims.

FDA regulation has so far enabled firms to sustain trade secrecy for competitively valuable information while still capturing some of its value to

95. It is possible that the data could be used to secure regulatory approval to sell generic products in foreign markets.

96. See *supra* note 2.

customers. FDA approval, in consultation with panels of outside experts, serves a certification function that enhances the credibility of informational claims about products while preserving the substantial secrecy of the underlying data. FDA regulation combines the bureaucratization of study design and data analysis with a system of scientific peer review and certification of undisclosed data. In the process, it tends to standardize the data that is collected and the format in which summary information is disclosed to the public, clarifying and simplifying the information signals given to a public that is unable to evaluate the data for itself. But the combination of trade secrecy and FDA regulation inevitably leads to suspicion of a regulatory process that is not transparent, especially when previously undisclosed product risks ultimately emerge. Moreover, sequestering valuable data within the FDA limits its social value by constraining access on the part of health care providers who might use it to make better therapeutic choices and by competitors who might use it to develop better products at lower cost.

CONCLUSION

Clinical trials to assess the effects of drugs in patients constitute a valuable form of R&D that offers the prospect of improving decisions about how best to use drugs to prolong and improve human life. But because the results of rigorous trials could potentially reduce product sales rather than increase them, drug-developing firms may not reliably capture the value of this R&D. How to motivate firms to make socially efficient investments in studying the effects of drugs in patients is thus a major challenge for the legal system.

Patent protection, FDA regulation, and trade secrecy each offer firms some protection against the use of data from clinical trials by free riders in competition with them, but each has its limitations, particularly as a mechanism for appropriating the value of information about new uses of old drugs. Trade secrecy offers firms the prospect of suppressing unfavorable information, thereby minimizing the risk to firms that trials of new uses will diminish sales revenues. On the other hand, trade secrecy truncates the social value of the resulting information by sequestering it from the people who stand to benefit from its disclosure and is therefore the most problematic of these legal regimes. More value could be realized overall by combining exclusive rights in product markets with public disclosure of data from clinical trials. This balance of public disclosure with private exclusionary rights is familiar to students of the patent system. Congress appears to have attempted to achieve a similar balance for data from clinical trials in the Hatch-Waxman Act, although regrettably that is not how the FDA has interpreted the legislation. The resulting secrecy limits the information base for making current health care choices and for developing future products, and calls into question the good faith of drug developing firms and the judgment of the FDA. Perhaps it is time to try a more open approach.

Regulatory Paternalism in the Market for Drugs: Lessons from Vioxx and Celebrex

Richard A. Epstein, LL.B.*

INTRODUCTION: RUMBLINGS OF DISCONTENT

The trials and tribulations of the pharmaceutical industry made front-page news in the Fall of 2004. On September 30, 2004, Merck & Co. announced that it would voluntarily pull its Cox-2 inhibitor, Vioxx, from the market.¹ To say the least, the decision to take the drug off the market caused no little stir. Vioxx, which had entered the market with great fanfare in 1999, had become an instant blockbuster drug with over one hundred million prescriptions,² twenty million users,³ and about \$2.5 billion in annual sales.⁴ The success of the drug paralleled that of two Pfizer Cox-2 inhibitors, Celebrex and Bextra.⁵ The success of all

* James Parker Hall Distinguished Service Professor of Law, The University of Chicago; Peter and Kirsten Bedford Senior Fellow, The Hoover Institution. My thanks to Anup Malani, Henry I. Miller, M.D. and Cass R. Sunstein for comments on an earlier draft, to the participants at the pharmacoeconomics workshop at UCLA, and the public policy workshop at Stanford Law School for their helpful comments, and to Eric E. Murphy, University of Chicago Law School, class of 2005 for his excellent research assistance. For the record, I should like to state that I have worked on a wide variety of issues that pertain to pharmaceutical regulation and liability as a consultant for PhRMA, and as a consultant to Pfizer Company on a number of matters, including the Rezulin litigation. The views in this paper are entirely my own.

1. Barnaby J. Feder, *Criticism of Drug May Leave Pfizer Awash in Lawsuits*, N.Y. TIMES, Dec. 18, 2004, at C1; see also Anahad O'Connor & Denise Grady, *Pfizer and Celebrex: The Patients' Problems May Send Many Patients Back to Age-Old Aspirin*, N.Y. TIMES, Dec. 18, 2004, at C1.

2. Andrew Leckey, *Prognosis Is Cautious for Merck Shares, Analysts Say*, CHI. TRIB., Nov. 7, 2004, at C8.

3. Theresa Agovino, *Lawsuits Threaten Health of Merck; Vioxx Litigation May Cost Billions*, CHI. TRIB., Nov. 8, 2004, at A1. The article estimated that potential tort liability could amount to \$17.6 billion over the next decade.

4. Leckey, *supra* note 2.

5. For the FDA's cautious position on the decision to take Vioxx off the market, see FDA, Vioxx (rofecoxib) Questions and Answers, Question 12, at <http://www.fda.gov/cder/drug/infopage/vioxx/vioxxQA.htm> (Sept. 30, 2004) (noting that "[t]he results of clinical studies with one drug in a

three drugs is (or at least, was) attributable to their apparent ability to satisfy the best of both possible worlds by relieving pain without provoking the risk of stomach or intestinal bleeding inherent to ibuprofen and similar drugs. The number “2” appended to the term Cox, with respect to drugs such as Vioxx, signified a welcome measure of specificity.⁶ Drugs in this family could work effectively where needed without causing disruption where they were not wanted. Indeed, Merck had such confidence in the ability of Vioxx to specifically target its effects that the company was seeking to expand the portfolio of permissible uses by raising the dosage to determine the effectiveness of Vioxx in treating polyps—intestinal growths that could become cancerous. However, during these trials, Merck discovered in its own clinical data an apparent increase in the number of negative cardiovascular occurrences, which, if extrapolated, “may” suggest that as many as 27,000 persons had died from the use of the product.⁷

Merck’s decision to withdraw the drug from the market took place before the FDA made any such demand,⁸ which of course leaves open the possibility that the drug could be returned to the market without a new round of FDA approvals. The common folk wisdom in the litigation industry suggests that a voluntary removal plays much better before a jury in subsequent litigation than a forced removal after a prolonged FDA hearing, which is closer to the situation with the diabetes drug, Rezulin.⁹ Yet in this instance, Merck’s action seems to have had the opposite effect. The decision to take Vioxx off the market was widely read as a fatal admission of dangerous conduct by a firm that should never have made the launch in the first place. The veritable firestorm of reactions included the anticipated onslaught of ordinary tort actions for personal injuries buttressed by congressional investigations, inquiries by the Securities and Exchange Commission, derivative actions, suits for refunds, internal inquiries, and so

given class do not necessarily apply to other drugs in the same class. All of the nonsteroidal anti-inflammatory drugs (NSAIDs) have risks when taken chronically, especially of gastrointestinal (stomach) bleeding, but also liver and kidney toxicity.”).

6. See Andrew Pollack, *New Scrutiny of Drugs in Vioxx’s Family*, N.Y. TIMES, Oct. 4, 2004, at C1 (“There are two forms of COX, and one of them, COX-1, helps protect the stomach lining from acids. The older drugs block both forms, which is why they cause ulcers and gastrointestinal complications that have been estimated to result in 7,500 to 16,500 deaths a year in the United States. The COX-2 inhibitors, as their name implies, block COX-2 much more than the stomach-protecting COX-1.”).

7. Leckey, *supra* note 2; see also Bruce Japsen, *Merck Withdraws Arthritis Drug; Vioxx Increased Danger to Heart*, CHI. TRIB., Oct. 1, 2004, at C1.

8. See FDA, Vioxx (rofecoxib) Questions and Answers, *supra* note 5.

9. See discussion *infra* notes 67-69 and accompanying text.

forth.¹⁰ Merck shares lost \$12 from \$45.07 to \$33 the day it announced that it would take Vioxx off the market, only to stabilize in the \$29-\$33 range thereafter.¹¹ Merck is now thought to be a potential merger target, and its board of directors has offered some 230 of its most senior managers special bonuses that will be triggered if either the company is taken over by another firm or if some other firm acquires twenty percent of its outstanding shares.¹² This episode prompted sharp criticisms by independent corporate watchdogs who treated it as yet another blunder by a weak Merck board.¹³

When Vioxx was taken off the market, the attention quickly turned to Celebrex and Bextra. Celebrex, which enjoyed a somewhat larger market share than Vioxx with about twenty-six million users generating at present some \$3.3 billion in annual sales, had not been linked to any elevated risk of heart exposure. It was not spared from the criticism, however, that it too would be shown to possess the same or similar risks as Vioxx.¹⁴ On December 17, 2004, the other shoe dropped when Pfizer announced that one of two clinical studies on Celebrex revealed that it presented an elevated risk of heart attacks.¹⁵ The clear implication

10. See, e.g., Alex Berenson, *Merck's Board Appoints Panel To Investigate Handling of Vioxx*, N.Y. TIMES, Dec. 8, 2004, at C6. The inquiry does not have a termination date, nor is it clear that the special committee will publish its results. *Id.*

11. Barnaby Feder, *Merck's Actions on Vioxx Face New Scrutiny*, N.Y. TIMES, Feb. 15, 2005, at C1.

12. See Alex Berenson, *Merck Offering Top Executives Rich Way Out*, N.Y. TIMES, Nov. 30, 2004, at A1.

13. *Id.* One recent story lamenting the decline sums up the fiasco with the title "Not Everybody Loves Raymond," referring to the controversy surrounding Raymond V. Gilmartin, Merck's Chairman since 1994. Alex Berenson, *Not Everybody Loves Raymond*, N.Y. TIMES, Dec. 15, 2004, at C1.

14. See, e.g., *Arthritis Drug Worries*, CHI. TRIB., Oct. 17, 2004, at C7; Feder, *supra* note 11 (noting slumping sales of Celebrex and Bextra, even in the absence of clear proof of cardiac risks from normal dosages).

15. See FDA, Statement on the Halting of a Clinical Trial of the Cox-2 Inhibitor Celebrex (Dec. 17, 2004), at <http://www.fda.gov/bbs/topics/news/2004/new01144.html>. The statement noted the following:

The Food and Drug Administration (FDA) learned last night from the National Cancer Institute (NCI) and Pfizer, Inc., that NCI has stopped drug administration in an ongoing clinical trial investigating a new use of Celebrex (celecoxib) to prevent colon polyps because of an increased risk of cardiovascular (CV) events in patients taking Celebrex versus those taking a placebo.

Patients in the clinical trial taking 400 mg. of Celebrex twice daily had a 3.4 times greater risk of CV events compared to placebo. For patients in the trial taking 200 mg. of Celebrex twice daily, the risk was 2.5 times greater. The average duration of treatment in the trial was 33 months.

of the second finding is that the higher incidence of adverse cardiovascular events might be found in all Cox-2 inhibitors, not just Vioxx. As of this writing, Pfizer has not taken Celebrex off the market in light of the unresolved issues surrounding the inconsistent results from the various clinical trials, but it has stopped consumer advertising of the drug. The reports of lawsuits and the renewed popularity of aspirin (notwithstanding its high incidence of gastrointestinal side effects, particularly in people taking large amounts for extended periods) sent the value of Pfizer stock plunging¹⁶ and raised editorial calls for both the FDA and Pfizer to think hard about whether it is best to “yank” Celebrex from the market now that Vioxx has been pulled.¹⁷

Vioxx is not the only high-profile drug to have been withdrawn from the market. A similar fate awaited the Warner-Lambert (now Pfizer) drug Rezulin, which was voluntarily withdrawn, albeit with severe FDA pressure, from the market in 2000, three years after its launch.¹⁸ Rezulin’s strength lay in its ability to attack diabetes differently from drugs previously in use, but post-marketing data detected an increase in liver complications, including sixty-three cases of liver failure. The actual record is filled with various factual disputes about the frequency and severity of side effects. At the conclusion of the 1999 FDA hearings on the matter, the FDA recommended that the drug not be used as an initial therapy for diabetes but only as a second line treatment, and then in conjunction with other drugs.¹⁹ A year later, Pfizer withdrew Rezulin from the market after David Willman of the *Los Angeles Times* published an exposé in which he denounced Rezulin as a “killer drug.”²⁰ Rezulin’s withdrawal precipitated a number of lawsuits, including not only actions for personal injury or death attributable to the drug, but also actions by individual consumers and

A similar ongoing study comparing Celebrex 400 mg. once a day versus placebo, in patients followed for a similar period of time, has not shown increased risk.

Id. For front page stories on the fast-breaking events, see Feder, *supra* note 1; Gardiner Harris, *Pfizer and Celebrex: The Overview; Drug Trial Finds Big Health Risks in 2nd Painkiller*, N.Y. TIMES, Dec. 18, 2004, at A1; Bruce Japsen, *Heart Risks Found From Celebrex*, CHI. TRIB., Dec. 18, 2004, at C1; and O’Connor & Grady, *supra* note 1.

16. See Feder, *supra* note 1; O’Connor & Grady, *supra* note 1.

17. Editorial, *New Doubts About Celebrex*, N.Y. TIMES, Dec. 18, 2004, at A18.

18. For a fuller summary of the relevant events discussed in this paragraph, see *In re Rezulin Prods. Liab. Litig.*, 210 F.R.D. 61, 62-64 (S.D.N.Y. 2002) (denying the motion to certify a class).

19. *Id.* at 63.

20. David Willman, *The Rise and Fall of the Killer Drug Rezulin*, L.A. TIMES, June 4, 2000, at A1; see also David Willman, *Researcher’s Fees Point to Other Potential Conflicts at NIH*, L.A. TIMES, Jan. 28, 1999, at A5; David Willman, *Scientists Who Judged Pill Safety Received Fees*, L.A. TIMES, Oct. 29, 1999, at A22.

third party payers to recover the sums paid to acquire the drug in the first place.²¹

The stakes involved in these withdrawals of drugs from the market, whether voluntary or mandated, are enormous.²² From the point of view of the patient, if the recalls are correctly executed much needless suffering may be avoided. But if useful drugs are withdrawn with no substitutes, needy patients are deprived of another weapon in their arsenal against disease and misfortune. The stakes are every bit as large institutionally. Institutional actors are not only affected by the litigation that withdrawals spawn. The drug manufacturer also suffers the reputational losses of withdrawals; the medical profession and the pharmaceutical industry face added scrutiny;²³ and the FDA, the tort system, and the securities markets bear the reverberations of the decisions.

21. I see little benefit to the use of the tort system. For a discussion of some of the liability issues, see *Wakefield v. Warner-Lambert Co.*, No. 99,086 (Okla. Civ. App. July 20, 2004) (upholding a wrongful death verdict of \$1,500,000 in compensatory damages and \$10,000,000 in punitive damages). It was notable that the decedent had died from hemolytic anemia; the claim was that the decedent could not fight off the condition because his liver function had been impaired by Rezulin. The court upheld the decision not to allow the "comment k" defense, see RESTATEMENT (SECOND) OF TORTS § 402(A) (1964), on the ground that adequate warnings were only a defense in cases of "exceptional products," see, e.g., *Hill v. Searle Labs.*, 884 F.2d 1064, 1069 (8th Cir. 1989). The court did not address the admitted fact that the decedent had had five similar incidents before taking Rezulin. The punitive damages were said to relate to the general distribution of the drugs and the profits it generated. Other similar cases are scheduled for trial. The hostile, chilly reception to out-of-state defendants in state court should be evident. Note that I have worked with Pfizer on some Rezulin cases, but only became involved in this one in connection with a petition for certiorari to the Supreme Court.

It is worth noting that the cost-internalization arguments of product liability cut both ways. False attributions of liability lead to the unwillingness to introduce new drugs into the market at the same time that so may drug industry critics deplore the emphasis on so-called me-too drugs. See Arnold Relman & Marcia Angell, *America's Other Drug Problem: How the Drug Industry Distorts Medicine and Politics*, THE NEW REPUBLIC, Dec. 16, 2002, at 27. For a defense of me-too drugs, see Thomas H. Lee, "Me-Too" Products—Friend or Foe?, 350 NEW ENG. J. MED. 211 (2004). For my critique of Relman and Angel, see RICHARD A. EPSTEIN, DOES AMERICA HAVE A PRESCRIPTION DRUG PROBLEM?: THE PERILS OF IGNORING THE ECONOMICS OF PHARMACEUTICALS 1 (Inst. for Pol'y Innovation, Issue Brief, 2004).

22. Withdrawal from the market refers to the ability to sell new drugs. In addition, once a drug has been withdrawn those units already in the marketplace may also be recalled. Withdrawal and recall are clearly complementary strategies.

23. For two denunciations, see MARCIA ANGELL, THE TRUTH ABOUT THE DRUG COMPANIES: HOW THEY DECEIVE US AND WHAT TO DO ABOUT IT (2004); and JEROME P. KASSIRER, ON THE TAKE: HOW MEDICINE'S COMPLICITY WITH BIG BUSINESS CAN ENDANGER YOUR HEALTH (2004). For my critical review of both books, see Richard A. Epstein, *Pharma Furor*, LEGAL AFF., Jan./Feb. 2005, at 60.

In sorting out the various consequences of asserted drug failure, two interrelated questions are decisive: Which drugs should be let on the market in the first place? And which ones should be taken off? The ferocious public attacks in the Vioxx, Celebrex, and Rezulin cases are well encapsulated in this gloomy assessment offered by *The Lancet* after Vioxx was taken off the market:

[D]rug regulators must now reassess the safety and efficacy thresholds required for the licensing of a new pharmaceutical product. Clearly, this is an immensely complicated equation involving, among other factors, the nature of the condition being treated, the therapeutic strategies already available, and the perceived benefit-to-hazard ratio of the new treatment. The Vioxx story is one of blindly aggressive marketing by Merck mixed with repeated episodes of complacency by drug regulators. We need clear statements from all parties in this sorry tale about the lessons to be learned. Without more vigilant drug regulation in the future, doctors will continue to be misled and patients' lives will continue to be endangered.²⁴

The controversy over the usage of dangerous drugs has now reached a fever-pitch, which is all the more reason to step back for a moment from the dramatic incidents of these and similar cases to develop a coherent framework to decide whether the critics of both the pharmaceutical industry and the FDA are right. That question in turn requires that we consider *two* alternatives to the status quo. The first is that we tighten up the system of regulation, both before drugs are released into the market place and after they are in common use. The second is that we *relax* the use of state regulation in *both* the prior approval and recall scenarios. The latter position has received little support in polite company, but, on balance, it has much to commend it.

This Article addresses two interlocking issues. Part I develops a simple model to determine which drugs should be released into the marketplace and why. Its central point is that the inherent *heterogeneity* in all populations cuts strongly in favor of a relaxation in the standard of pre-market approvals, as is urged in a recent paper by Malani and Hu.²⁵ The regulator who works upstream of the physician and patient lacks any knowledge of individuated circumstances that should rationally influence the decision of which drug, if any, to take, and in what dosage. So long as physicians and patients have some skill in locating the patient's position in the distribution, there is no reason to rely on the upstream

24. Editorial, *Vioxx: An Unequal Partnership Between Safety and Efficacy*, 364 THE LANCET, 1287, 1288 (2004) [hereinafter *Vioxx: An Unequal Partnership*]; see also Eric J. Topol, *Good Riddance to a Bad Drug*, N.Y. TIMES, Oct. 2, 2004, at A15 (criticizing the use of Rezulin because of its increased risk of heart attack or stroke, from 1.9% to 3.5%).

25. Anup Malani & Feifang Hu, *The Option Value of New Therapeutics* 1 (Oct. 25, 2004) (unpublished manuscript, on file with the author).

averages that the FDA uses. Patients and physicians should be allowed to incorporate downstream knowledge into their decisions. As far as I can tell, there are no substantive provisions in the current legislation, with its mandates that drugs be both safe and effective, that prevent the FDA from considering the variation in responses across individuals in setting the appropriate standards for decision. In light of this basic situation, Part II then argues that this model should carry over to questions of withdrawal and recall of drugs from the marketplace, either by government mandate or firm decision. So long as individual users have acquired knowledge of their personal benefits and side effects of particular drugs, companies should be reluctant to pull drugs from the marketplace, and the government should be cautious in ordering them off.

Accordingly, something is sadly amiss in dealing with the regulatory framework on prescription drugs. On this critical issue, the FDA should use its power to keep drugs from the market or to withdraw them from it with far greater caution than it does today. Often, it relies on cost-benefit analyses that can only be termed, at best, tentative and, at worst, primitive. Its entire effort to make better judgments on what treatments should be used and why smacks of an unthinking paternalism that reveals its own institutional shortcomings, as well as those of its critics who plump for stricter regulation.

Looked at in the broad scheme of things, the entire regulatory apparatus today suffers from an excess of ambition. The FDA has a critical role to fulfill in keeping counterfeit and bogus drugs off the market. It should deal harshly and effectively with fraud. But when the question turns to whether individual physicians and consumers have sufficient information to make appropriate choices, it enters into a vast swamp through which it cannot find a consistent path. In its effort to protect ordinary patients from error, it probably makes more errors than it guards against because it lacks both the particularized knowledge and the strong incentives to get matters right that ordinary people bring to their own affairs. In this area, the cure is frequently worse than the disease. The problems of error and bias that have been so frequently identified are real, but they are not avoided by the FDA or the tort system, which have additional difficulties of their own.²⁶ Here is yet another case where the administrative agency should do the unproblematic task well—deal with purity and fraud—while showing a bit of caution in making judgments for others on matters of

26. See generally HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT (Thomas Gilovich et al. eds., 2002); JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES (Daniel Kahneman et al. eds., 1982). With these, contrast GERD GIGERENZER ET AL., SIMPLE HEURISTICS THAT MAKE US SMART (1999); and Gerd Gigerenzer, *Bounded Rationality: Models of Fast and Frugal Inference*, 133 SWISS J. ECON. STAT. 201 (1997). The titles reveal everything about the differences in approach.

safety and effectiveness. Protection against fraud is one thing; paternalism, whether or not intended, is quite another.

All this uneasiness about drug safety still leaves open the question of who should assemble all the information that surrounds the use of any standard drug even when fraud and misbranding are not at issue. At this point, however, there is no reason to place trust in a government monopoly, especially one that has shown itself to rate false positives (letting drugs that should be kept off the market onto the market) more highly than false negatives (keeping drugs off the markets that should be allowed). Since warnings are not coercive but informative, there is no need for a government monopoly. Private organizations can issue their own findings, and, if need be, other rating organizations can rate the various organizations that supply the data. The obvious point that individual patients, and often their physicians, are unable to assemble the needed data explains why we need third-party involvement. It does not, however, justify another government monopoly.

I. INITIAL DRUG APPROVAL

A. Downstream, Not Upstream

Although both the Rezulin and Vioxx cases focused on withdrawal and recall, those business and FDA decisions were clearly dependent on the judgments made in the initial approval process. The less risk-averse the FDA runs that initial process, the more likely it is that poor drugs will slip through the net and the greater the likelihood that dramatic business, regulatory, and litigation responses will come with the first signs of adverse events. The more stringent the initial regulatory process, the more likely it is that fewer poor drugs will slip through the net and hence the pressure for drug withdrawals and recalls will be reduced. The casual analyst might conclude, therefore, that more caution is in order at the first stage. But the error of that position is seen quickly enough by putting forward this simple proposal at the extreme: Avoid all problems with withdrawals and recalls by allowing *no* new drugs to reach the market. That conclusion would not even make perfect sense in a world in which all drugs had a negative expected value, so long as individual users who would benefit from the use of the drug can self-select. That conclusion, however, makes even less sense in the current world, where many drugs that enter the market perform as well, or even better,²⁷ than was expected on their launch. Stated otherwise, if all

27. Such was the situation with Norvir, an Abbott anti-AIDS drug, which was found to be more effective if used in combination with other pharmaceuticals when it could be taken in lower dosages. Abbott raised its prices for the drug and was faced with threats of losing its patent rights

drugs had a positive expected value in use on launch, then none should be kept off the market. But the obvious concern that spurs FDA involvement is that we cannot live in that Nirvana either. Drugs can kill, and if they do, no amount of damages will restore the victims and their family to the status quo ante. Nor will the deterrent effect of damages work well if the suppliers of new drugs are fly-by-night operations that are able to liquidate or go bankrupt in the short term so as to be unavailable, perhaps years later, to answer for their original defaults. Criminal sanctions are available, but are subject to high standards of proof that are unlikely to play a role in most cases.²⁸

A system of prior restraint, then, is in principle permissible to deal with this problem, but it is not one for which private law enforcement provides much traction. Private injunctions work tolerably well, for example, in land use cases in which one party pollutes the land of his or her neighbor, but they falter when pollution from multiple sources damages many separate individuals. At this point the sensible approach has the state intervene as the agent for the aggrieved parties. But the hard question still remains: How do we know with any particular drug application whether the exercise of that permit power benefits the individual members of the public whom it is supposed to protect? Once it is recognized that there are two kinds of error—letting drugs on the market too quickly and keeping them off for too long—then someone has to decide which error is larger for which application. This task is by no means simple, even if we ignored the standard litany of public choice concerns about how individual interest groups can capture public agencies and turn them to private advantage—a risk that is as great with a consumer advocacy group, such as Public Citizen, as it is with any pharmaceutical company.²⁹

under the Bayh-Dole Act, which allows for march-in rights in limited circumstances. See 35 U.S.C. § 203 (2000). But the statutory claim proved weak and the flap was short-lived. See *Abbott Laboratories Comments at NIH Public Meeting Regarding Norvir and Bayh-Dole March-in Provisions*, PR NEWswire ASS'N, May 25, 2004; Bruce Japsen, *Abbott AIDS Drug Pricing Leads to Review of Patent*, CHI. TRIB., May 21, 2004, at C1; Bruce Japsen, *Abbott Defends Price Boost on AIDS Drug at U.S. Hearing*, CHI. TRIB., May 26, 2004, at C1. Note that it is critical to allow for these price increases lest the manufacturer start with a high price that retards usage unnecessarily. To take the contrary position is no better than arguing that a landlord cannot raise rentals after the original rents are set.

28. See 21 U.S.C. §§ 331, 333 (2000) (misbranding).

29. The calculus of influence is very complicated, but size and resource base are surely not the sole determinants. An individual firm with huge assets is vulnerable to legislative threats of regulation and taxation and its credibility is always suspect relative to that of independent public interest organizations that have no direct financial interest in a particular issue, but strong ideological commitments. No private firm could get away with the assertion that it is proper to ignore present value calculations in determining the costs of new products, but Public Citizen has

B. Enter Heterogeneity

One key step in this complex process of social control is to develop a sound set of norms that indicate which form of government action is appropriate at which stage and why. In dealing with this question, Malani and Hu start off by noting that the “[FDA] employs a simple decision-rule when deciding whether to approve a new drug for use by physicians: The average treatment effect of the new drug must be superior to the average effect of a placebo.”³⁰ Both safety and efficacy fit into this equation. Yet, as Malani and Hu immediately point out, this rule is flatly incorrect the moment that one takes into account the ability of physicians and patients to exploit the heterogeneous responses to the drug in question.³¹ Every natural population has a variance whose essential features frequently can be captured in a normal distribution, that is, a bell-shaped curve, with a peak in the middle and a symmetrical distribution around it. Whether we think of height, eye color, lactose tolerance, or any of a million human traits, it is now indisputable that small genetic variations can lead to very large differences in observed behaviors or physical types.³² We know that human responses to drugs also conform to this pattern. Some people will do better with given drugs than others. The explanation could stem from a thousand causes: age, sex, race,³³ and the like. Assembling a comprehensive list of the relevant factors and assigning weights to them is probably beyond the capacity of modern epidemiological science. Finding the width of the variation is also no easy task, for there is no reason to suppose that it is uniform for all medications and all populations. The question is what to make of this indisputable but incomplete

made just that claim. See PUB. CITIZEN, AMERICA’S OTHER DRUG PROBLEM: A BRIEFING BOOK ON THE RX DRUG DEBATE 46 (2002), <http://www.citizen.org/rxfacts>. For a contrast, see the more careful empirical study, Joseph A. DiMasi et al., *The Price of Innovation: New Estimates of Drug Development Costs*, 22 J. HEALTH ECON. 151 (2003) (using an eleven percent discount rate).

30. Malani & Hu, *supra* note 25, at 1.

31. *Id.*

32. The most dramatic illustration comes from the observation that human beings and mice share virtually all their genes. It is a losing proposition to argue that the differences between them must be small, when we know otherwise, because their gene pools are the same. What is needed is some explanation as to how differences in gene expression are powerful enough to explain the observed differences. For a popular account, see Matt Ridley, *The DNA Behind Human Nature: Gene Expression and the Role of Experience*, 133 DAEDALUS 89 (2004).

33. See, e.g., Sally Satel, *Race and Medicine Can Mix Without Prejudice: How the Story of BiDil Illuminates the Future of Medicine*, MED. PROGRESS TODAY, Dec. 10, 2004, at http://www.medicalprogresstoday.com/spotlight/spotlight_indarchive.php?id=449. The original study was published by Anne L. Taylor et al., *Combination of Isosorbide Dinitrate and Hydralazine in Blacks with Heart Failure*, 351 NEW ENG. J. MED. 2049 (2004). See also Sally Satel, *I Am a Racially Profiling Doctor*, N.Y. TIMES, May 5, 2002, § 6, at 56.

proposition?

In approaching this question, here is one initial benchmark: The variance in outcomes should be the dominant determinant in dealing with this decision. The higher the variance, the greater the gain that comes from learning where any particular individual is located on the curve in question. The FDA rule that looks only at the means of the placebo and the drug population systematically suppresses all reference to those variations, and it is modestly worrisome that the question of variance was *not* one of the factors referred to in *The Lancet* editorial, nor as far as I can tell, in any other editorial.³⁴ In practice, this effect is softened somewhat because of the ability to seek marketing approval for some discrete subpopulation that becomes the subject of a Phase II or Phase III clinical trial. But even that concession does not fully meet the problem. Even if the drug company can target in advance the group for which the drug is appropriate, which is a large question, any variation within that designated class is ignored. Consequently, a drug that does not meet the overall standard will be excluded from the market even if it works for some subpopulation. In addition, the drug will not be available for individuals outside the test population for whom it may work. Ideally, if there are some people for whom the new drug works better than a placebo and some for whom it does not, then by all means those who profit from the drug should take it while those who do not should avoid its use. But that desirable result can be reached only by allowing a drug that flunks the FDA standard to enter the market. Indeed, the greater the variation in response, the higher the return to the individual valuations, both within and across subpopulations. There will, in fact, be many cases where the average person does worse on the drug than on the placebo, but so what? All that proves is that a smaller fraction of the population can profit from the drug. It does not prove that the drug has no social value. The ban then becomes a blunt instrument because it does not separate out good from bad applications of the drug in question. In principle it is better to start dosages at the low end of the range and to increase them in light of the full range of individual responses.

The argument, however, is still more complicated than this simple version implies, for there is no guarantee that anyone will be able to determine with assurance *which* individuals could profit from the drug, *which* are not affected much one way or the other, and *which* are hurt by its use. If it turned out that no one had any indication in advance as to the effect that the drug had on him or her, the expected value of the drug declines because of the inability to route it to the right people. But even here, it does not necessarily follow that it should be kept from the market. People with very serious conditions might wish to throw the

34. It was not mentioned, for example, in *Vioxx: An Unequal Partnership*, *supra* note 24, nor in Topol, *supra* note 24.

“Hail Mary” pass because they have nothing to lose from an adverse reaction to the drug. But, they have much to gain from the random possibility that the new treatment will work where all other treatments have failed, which explains why some drugs are approved only as “second-line” or “third-line”—for use when other interventions have failed.

This last question is but one way of saying how difficult it is to introduce standardized measures of the expected benefits that flow from a positive drug use. Indeed, this same problem recurs in multiple forms. For example, suppose the adverse consequence of a given drug treatment is instant death and the positive effect is a modest reduction in some allergic reaction; then no one wants to take the risk that he falls on the wrong side of the distribution. For example, the antibiotic chloramphenicol is seldom used because of a low incidence of fatal aplastic anemia, but it remains on the market for situations in which it may be life-saving.³⁵ But the very starkness of that illustration may, paradoxically, remove the need for a ban so long as the firm is required to reveal publicly both the probability and magnitude of all effects to the extent that these statistics have been acquired through either animal or clinical tests. In the extreme case just mentioned, the disclosure requirement will be a death knell to any efforts to market the drug at all. Even without a ban, there is little risk that any drug company will want to put forward hydrogen cyanide as a new wonder cure. Nor would liability be much of an issue either. No one would use the drug because its negative payoffs would dominate everyone’s decision, no matter what their place in the overall distribution.

Yet most cases do not have that stark profile. Oftentimes, the drugs in question are given to very sick or debilitated individuals whose prognosis for palliation or cure is dim. In some cases, there will be people who can tolerate a drug well, while others cannot. The point here is that it is routinely possible in most cases to develop some signs—indications or contraindications—which supply people with at least a rough idea of where they stand with respect to a particular drug use. Every warning label contains a list of that sort and the FDA itself maintains a drug information site that provides “information about the products we regulate.”³⁶ To the extent that individuals or their physicians have reliable information, the case for keeping the drug off the market is far weaker

35. See, e.g., *Salmon v. Parke, Davis & Co.*, 520 F.2d 1359, 1361 (4th Cir. 1975) (“Chloromycetin, Parke, Davis’ trade name for chloramphenicol, is a potent, broad-spectrum antibiotic. Properly administered, it is a valuable, life-saving drug that can effectively treat stubborn infections. But it can be injurious—even fatal—if its use is not carefully monitored. According to the Food and Drug Administration, its most common, serious toxic effect is the development of anemia.”).

36. See Ctr. for Drug Evaluation & Research, FDA, Information About the Products We Regulate, at <http://www.fda.gov/cder/drug> (last visited Feb. 14, 2005).

than it is if no such knowledge is available. Information of this sort is certainly available in most cases. Most drugs, like the Cox-2 inhibitors or the statins, which are used to control cholesterol, fall into distinct classes that offer some advance warnings as to which individuals are likely to gain or suffer most from a treatment. In addition, in most cases it should be possible to start individuals on low dosages of products and observe whether the beneficial effects outweigh the unpleasant side effects. Where the drug has some positive effects then the dosage might be cautiously increased, keeping a watchful eye for dose-sensitive side effects. Yet even when the drug does not cure the condition, it hardly follows that its use should be abandoned. The drug may still produce some beneficial effects at a lower dosage or in combination with other drugs whose general properties are well understood. The entire process is one of incremental adjustments in which individual feedback is immediately available and highly reliable. The one confounding problem comes from the placebo effect, which can be quite profound on people who have not received any active medications. But even in this context, the best approach may be to disclose the existence of the effect to patients and then let it operate to help them.

The possible permutations for drug use are quite varied, and it is for just that reason that the FDA should be reluctant to apply a bright line rule to keep drugs off the market. When any drug is kept from the market, regulation necessarily forecloses all the possible downstream adjustments that can be made by individual patients and their physicians in the use of particular drugs. Finding the right niche and level is standard business for countless drugs sold in the market today. Notwithstanding constant debates over its use,³⁷ Prozac remains on the market because individual physicians have had success in treating many depressive patients who have proved unresponsive to other treatments.³⁸ Steroids remain on the market even though they have a long list of adverse side effects, from weight increase to mood swings, which should daunt the most hardened potential users.³⁹ Accutane, an acne medication, remains on the market even though its potency can take the starch out of anyone, especially pregnant women for whom its use is manifestly and graphically counterindicated.⁴⁰ Today even thalidomide—rechristened Thalomid—is back on the market, and is extremely

37. See, e.g., Rick Giombetti, Prozac, Suicide and Dr. Healy (Mar. 20, 2002), at <http://www.whale.to/a/prozac.html>.

38. The warnings are found at Eli Lilly & Co., Prozac.com, Safety Facts, at http://www.prozac.com/how_prozac/safety_facts.jsp?reqNavId=2.5 (last visited Apr. 5, 2005).

39. Harm Kuipers, *Anabolic Steroids: Side Effects*, in *ENCYCLOPEDIA OF SPORTS MEDICINE AND SCIENCE* (T.D. Fahey ed., 1998), <http://sportsci.org/encyc/anabstereff/anabstereff.html>.

40. Evidence of such can be easily found on the web. See, e.g., ROCHE, ACCUTANE (ISOTRETINOIN) CAPSULES, <http://www.rocheusa.com/products/accutane/pi.pdf> (revised June 2002).

useful and profitable, among other things, for the treatment of leprosy.⁴¹

Playing with fire, then, is part of the overall picture—but the logic is inescapable. So long as downstream information is better than the generalized information in the possession of the FDA, the drug in question should be left on the market. Warnings galore can be printed on the packets and inserted in the Physicians Desk Reference. Informed consent could be required at the patient level. But the fundamental asymmetry remains. In some cases, tort liability should be added into the mix to deter the marketing of drugs without adequate warnings. The case for allowing a drug on the market is even stronger than allowing certain activities to go ahead, for example putting smoke stacks in operation, though they pose some environmental risk because of the benefits created from the activities. With drugs the self-help remedy is fully available, patients could simply not take the drug, which is not the case when pollution comes roaring through the front door. On the other hand, if the FDA bans a drug, that action allows for no second chance to correct any error in its judgment. If the FDA allows the drug on the market, there are all sorts of additional ways and opportunities to direct its use to that subset of the population that has the greatest use value.

C. How Safe, How Effective?

Part of the difficulty with the FDA's approval process stems from the definition of its mission. The FDA's position was summarized in these words after the Vioxx incident:

Modern drugs provide unmistakable and significant health benefits. It is well recognized that FDA's drug review is a gold standard. Indeed, we believe that FDA maintains the highest worldwide standards for drug approval. FDA grants approval to drugs after a sponsor demonstrates that they are safe and effective. Experience has shown that the full magnitude of some potential risks do not always emerge during the mandatory clinical trials conducted before approval to evaluate these products for safety and effectiveness.⁴²

Yet the articulation of this proposition conceals all relevant difficulties about the application of this standard. It is one thing to ask a party to illustrate that he drove on the right side of the road at the time of a collision. The line in the middle of the road is a conscientious effort to create a dichotomous universe in

41. For data, see CELGENE PHARMACEUTICALS, THALOMID (THALIDOMIDE), <http://www.celgene.com/PDF/thalomidPI.pdf> (last visited Mar. 23, 2005).

42. *Merck and Vioxx: Putting Patient Safety First?: Hearings Before the Senate Comm. on Finance*, 108th Cong. (2004) [hereinafter *Merck Hearings*] (statement of Sandra L. Kweder, Deputy Director, Office of New Drugs, FDA).

which actions do or do not comply with law.⁴³ The decision to allow a drug or keep it off the market might be termed “imperfectly dichotomous.” First, the decision to keep it off means that it is not used, but the decision to let it on leaves it for subsequent actors to decide. Second, no matter how hard one tries, there is no bright-line equivalent to the midline on a public highway to guide this decision. There are no drugs that are uniformly safe, and there are none that are uniformly effective.⁴⁴ All judgments about whether to let the drug on the market require a comprehensive kind of trade-off, which ultimately rests on questions of degree and extent. Once the true task of the mission is revealed, it becomes idle to attack the FDA whenever it lets the wrong drug on the market: It has made a calculated risk that proved, perhaps, wrong in the equation. However, once the inquiry is understood to be about trade-offs at the margin, making collective decisions to block drugs that will have use in some cases but not in others is a far larger sin. Quite simply, the misstatement of the criteria for drug permits leads to

43. This system of clear property rights is also congruent with the strict disjunction between liability and no-liability on which a tort system works. For this reason, it is superior to the kinds of Hand formula balancing tests that require a conscious comparison of the burden of precautions with the expected benefit that they would yield. For the Hand formula’s original enunciation, see *United States v. Carroll Towing, Co.*, 159 F.2d 169, 173 (2d Cir. 1947) (Hand, J.). The most celebrated defense of this formula as a universal solvent for the tort law is Richard A. Posner, *A Theory of Negligence*, 1 J. LEGAL STUD. 29 (1972). My defense of strict liability dates back to Richard A. Epstein, *A Theory of Strict Liability*, 2 J. LEGAL STUD. 151 (1973). For a comparison of the two systems, see RICHARD A. EPSTEIN, *SIMPLE RULES FOR A COMPLEX WORLD* 92-97 (1995).

44. One exception to the basic rule involves the administration of those substances that are found naturally in the body, such as thyroxin. These substances replicate natural processes and thus are virtually foolproof, at least if added in the right dosages. See Mary J. Shomon, *All About Thyroid Drugs* (Dec. 14, 2003), at <http://thyroid.about.com/cs/thyroiddrugs/a/overview.htm> (“The conventional treatment for hypothyroidism is thyroid hormone replacement—basically, taking a prescription drug that acts similarly in the body to the human hormone thyroxine that the thyroid would normally produce.”). One possible exception to this rule has to deal with hormone replacement therapy for postmenopausal women, which has been under extensive scrutiny as of late. But there are three complications here. First, the treatment in question need not replicate the levels that the body normally produces at a particular stage in life. Rather, it may seek to increase the hormone levels above what they are in normal individuals. Second, the identified risk factors did not address the risks for women who take less than standard dosages. Third, the initial studies did not distinguish between women who started therapy before menopause from those who started later. For the original study that recommended stopping hormone replacement therapy, see Writing Group for the Women’s Health Initiative Investigators, *Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results from the Women’s Health Initiative Randomized Control Trial*, 288 JAMA 321 (2002). For the inevitable qualifications and complications, see Tara Parker-Pope, *Rethinking Hormones Again: Heart Risk May Be Lower in Women Who Start Early*, WALL ST. J., Oct. 12, 2004, at D1.

a fundamental misconception of the FDA mission.

This standard litany of FDA responsibilities as bracketing both safety and effectiveness also helps conceal the fundamental difference between these two statutory tests. Clinical trials are certainly part of the standard protocol, and these are customarily divided into Phase I, Phase II, and Phase III trials.⁴⁵ The initial concern of the Phase I trial is with basic safety: How much can patients tolerate of a new drug. Hence, the question to be answered is whether a small number of individuals can tolerate various levels of exposure so as to make it worthwhile to continue the experiment. But that said, the determination at Phase II trials (larger affairs, with different dosage levels, intended to measure safety and effectiveness in patients of the type for whom the drug is ultimately intended), and Phase III trials (often extended operations at multiple sites, dealing with both safety and effectiveness) works in different ways.

In principle, drugs should, at least in new classes, be able to demonstrate their effectiveness from relatively small groups. To be sure, in the statistical sense, a “significant” result in close cases requires a large population, which allows the investigator to determine that the two groups are not drawn from the same urn. Make that population large enough and a response difference of one percent will be statistically significant. But the *social* significance of that smallish statistical significance is another affair altogether. The additional return from that one percent increment in overall effectiveness is sufficiently small that leery patients will not willingly pay heavily for this modest improvement, either in cash or in alternative medical risks that any treatment brings in its wake. (A large effect for one percent of patients is another matter altogether.) The only significant outcomes that are worth pursuing, therefore, are those that offer eye-popping results on small populations that don’t require any refined statistical analysis for verification. In one sense, therefore, the most promising drugs in pioneer classes should be regarded as “effective” with relatively little information.

Safety, however, raises a very different set of concerns. Recall that the increased rate of heart attack and stroke in Vioxx increased from 1.9 to 3.5% in an undifferentiated user population.⁴⁶ Because the base rates and increments are both low, it takes very large populations, often over prolonged periods of use, to make a sensible judgment on safety issues. Yet each front-end clinical cost that is added to the mix delays the use of successful drugs as well as that of unsuccessful ones. On balance, therefore, there is a lot to be said for allowing

45. See 21 C.F.R. § 312.21 (2004) (describing what occurs at each phase of the testing).

46. Peter Gornick & Ronald Kotulak, *Patients Calm After Merck Pulls Vioxx*, CHI. TRIB., Oct. 2, 2004, at C1. Most sources simply note that the risk of adverse consequences was about “double” without giving the numbers. See, e.g., Agovino, *supra* note 3.

marketing after some effectiveness is established with, of course, the use of warnings to highlight the unresolved nature of the risk. The superiority of downstream individuation should not be ignored in setting the basic parameters.

D. Using—and Stopping—Clinical Trials

The difficulties in setting the appropriate criteria for drug marketing plays itself out most vividly in the story of an Amgen drug, glial cell line-derived neurotrophic factor (GDNF), which had been made available in clinical trials as a potential treatment for Parkinson's disease. Some individual patients had reported marked personal improvements from use of the drug, which allowed them to redo kitchens when previously they could not hold a nail stapler.⁴⁷ But when Amgen ran its clinical trial, it first found that the drug worked no better than a placebo on average. It then discovered that the drug carried with it serious safety risks and hazards. After reporting the information to the FDA and consulting with outside ethicists, Amgen stopped GDNF clinical trials, leaving its previous users in a lurch.⁴⁸ The howls of protest from unhappy patients are confirmed by the desperate measures they took before Amgen's decision was made final. GDNF is administered by a pump that injects the compound into the brain through a catheter. Many patients refused to shut down their pumps because they feared that they could not be reopened if the clinical trials had continued. The nagging suspicion is, therefore, that one reason why Amgen made this decision is the risk of liability and regulatory grief that would follow if it took any other path.

It does not take an expert to realize the genuine difficulties in interpreting the data. Some, perhaps all, of the improvement might be properly attributable to the "placebo effect" in starting any form of treatment.⁴⁹ But, alternatively, design flaws in the study could have reduced the effectiveness of the drug relative to its full potential. The potential side effects could be quite severe, but, then again, they were observed only in monkeys in dosages several times higher than those used in people on a brain only one-twelfth the size. The adverse effects were not found in human beings, at least to date, so the time of onset, frequency, and

47. Andrew Pollack, *Many See Hope in Drug Pulled During Testing*, N.Y. TIMES, Nov. 26, 2004, at A1.

48. Andrew Pollack, *Patients in Test Won't Get Drug, Amgen Decides*, N.Y. TIMES, Feb. 12, 2005, at C1.

49. The placebo effect is difficult for anyone to confront because it says that people's own subjective evaluations might not supply them with the best decision. But here the solution seems to be more disclosure, not a discount of patient preferences. Let people know that they may be taking placebos. If they improve with that knowledge, then let them continue. If there is a risk of an undisclosed danger of adverse side effects from the experimental drug, then disclose that as well.

seriousness of the effects are hard to assess. Some doctors supported the use of the drug; others were against it. In light of the murky medical situation, no doubt, Amgen could be concerned about the size of the potential market for this highly controversial product. And there is, in my view, no duty for them to invest further in a drug that may promise them the unhappy trifecta of small markets, lagging profitability, and high liability exposure. But still the situation is unsettling. Suppose that the drug has a placebo effect and high risk; with full disclosure, why prevent people from taking it if they report pronounced improvements that are both undeniable and easily verifiable?

There is, of course, a now abundant line of literature that purports to supply that reason by demonstrating that individuals suffer from a myriad of cognitive biases and defects that cause them to ignore base rates and miscalculate the odds in making decisions.⁵⁰ No doubt all this is true, as is evidenced by the way in which the FDA structures its basic standard so as to systematically understate the benefits from earlier drug approval by ignoring heterogeneity in the user population.⁵¹ But whatever the source and strength of these cognitive biases, no person should have any deep-seated emotional resistance against correcting his decisions once he obtains better information about what course of action will improve his own welfare. And these decisions are made by individuals whose feedback mechanism gives them instant information as to whether their individual condition moved to either the plus or the minus side. The implicit paternalism of allowing FDA supremacy assumes that a distant bureaucracy, which has its own institutional biases, will be a better guardian of all potential users than the people themselves.⁵² It is often said that the ability to take risks and bear their consequences is one of the marks of a self-reliant population. The presumptions here should be set strongly in favor of allowing individuals to continue to take those drugs of choice even as other individuals, quite properly, decide to follow the opposite course of action. The decision to ingest a given

50. See generally Daniel Kahneman & Amos Tversky, *On the Reality of Cognitive Illusions*, 103 PSYCHOL. REV. 582 (1996); Daniel Kahneman & Amos Tversky, *Prospect Theory: An Analysis of Decision Under Risk*, 47 ECONOMETRICA 263 (1979). For criticism, see Gerd Gigerenzer, *On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky*, 103 PSYCHOL. REV. 592 (1996). The field is dominated by two schools, which differ in the importance attached to these biases. It is easy to figure out which is which from the titles of their works. Compare, e.g., GERD GIGERENZER ET AL., *SIMPLE HEURISTICS THAT MAKE US SMART* (1999), with JUDGMENT UNDER UNCERTAINTY: *HEURISTICS AND BIASES* (Daniel Kahneman et al. eds., 1982).

51. For discussions of failures in both ordinary life and administrative agencies, see Cass R. Sunstein, *Probability Neglect: Emotions, Worst Cases, and Law*, 112 YALE L.J. 61 (2002).

52. See HENRY I. MILLER, *TO AMERICA'S HEALTH: A PROPOSAL TO REFORM THE FOOD AND DRUG ADMINISTRATION* (2000) (critiquing the FDA and urging competitive drug reviews to expedite the approval process).

drug is the polar opposite of any public goods or collective action problem that might call for state intervention.⁵³

E. Upping the Baseline

In light of these considerations, it is quite disturbing to see that the Vioxx dispute has strengthened the hand of those who think that a more restrictive set of tests should be required to let, and keep, drugs on the market. One of the worst proposals of this sort is to keep the traditional FDA protocol that stresses means to the exclusion of variance, but against a different baseline. Marcia Angell puts the proposal in the following words in speaking about Congress and the FDA:

[P]erhaps most important is what Congress has *not* done. It has not authorized the FDA to require that new drugs be tested against older ones as a condition of approval. The fact that drug companies get away with comparing drugs only with placebos is what makes it possible for the industry to live on me-too drugs. If not for that, drug companies would have no choice but to work on truly innovative drugs.⁵⁴

Under this proposal the new drug will be required to beat a baseline that is established by the first entrant of its class into the market. The clear subtext to this position is twofold: These markets are not competitive in any event, and all these me-too drugs are just look-alikes anyhow, so one should just go for the lowest price.⁵⁵ Both of these short-term assumptions seem to be wrong. After a close examination of the market in surgical stents, Dr. Thomas Lee concluded that me-too products “reflect and create competition among drug and device manufacturers, and that competition is also a powerful driver of better quality and lower cost.”⁵⁶ In similar work, DiMasi and Paquette noted that the elapsed time between the first arrival of a new drug within a class and its competitors has dropped from a median of 10.2 years to 1.2 years, with greater consumer choice.⁵⁷

In a deeper sense, this Angell proposal is fatally flawed because it replicates off a different baseline the same error that Malani and Hu identified in the current standard. It gives no weight to the potential variation within the subject

53. See, e.g., MANCUR OLSON JR., *THE LOGIC OF COLLECTIVE ACTION* (1965) (detailing the under-production of public goods).

54. ANGELL, *supra* note 23, at 204.

55. *Id.* at 89-90.

56. Lee, *supra* note 21, at 211.

57. Joseph A. DiMasi & Cherie Paquette, *The Economics of Follow-on Drug Research and Development Trends in Entry Rates and the Timing of Development*, 22 PHARMACOECONOMICS 11 (Supp. No. 2, 2004).

population in cases in which there is some individuation at the user level. To be sure, the movement from zero to one drug in the marketplace may have a more positive welfare effect than the movement from one to two drugs and so on down the line. Such is a consequence of the law of diminishing returns that is applicable in all cases. But even with diminishing returns, the gain will not fall to zero and the new product in question could provide a back-up insurance if the initial product, such as Vioxx or Rezulin, is pulled off the market (often unwisely). The newer and higher standard could easily delay the introduction of any follow-on drug, even if it is in general superior in safety and/or efficacy, to any drug that was first in class, as is commonly the case.⁵⁸ After all, the closer the means between the two compounds, the larger the statistical sample that is needed to establish the significance. That barrier will grow with each additional entrant who must make its way over a successively higher bar.

The problems here are still more acute because the sequencing of drugs into the marketplace is by no means as clear as this model suggests. Medical research, such as that on Cox-2 inhibitors, builds on basic science research that is publicly available. It follows that multiple companies will be pursuing the same leads simultaneously. Surely in some situations it could well be that one firm gets its patent first, but the second firm is able for a variety of reasons to get its drug through the FDA more rapidly. Is it really wise social policy to require a race to have only a single winner when its consequence (since research programs are often secret) could be to force firms to play in an all-or-nothing world where tiny advantages in the laborious approval process receive huge awards for no reason?

In dealing with these issues, it is often asserted that cutting down on me-too drugs makes sense. They are said to be a social waste because they only duplicate the kinds of expenditures made by others. But this is not an argument that is distinctive to new drugs in the marketplace. Rather, it applies to all cases in which competition requires the second entrant to duplicate some expenses of the first. But that wasteful expense argument hardly justifies this conclusion. Assume, for example, that we had no FDA to check on product quality and relied exclusively on damage remedies and the rare private injunction to guard against drug failure. In that setting, no one would claim that the first entrant should be able to block all subsequent entrants from seeking to take away its market share simply because the subsequent entrant will have to incur some costs of its own. The complete response here is that the only way to eliminate all duplication in costs is to give the first entrant into every market a legal monopoly that allows it to exclude its competitors. No way.⁵⁹

58. *Id.*

59. For a defense of the competitive solution, see Mark A. Lemley, *Ex Ante Versus Ex Post Justifications for Intellectual Property*, 71 U. CHI. L. REV. 129, 135-41 (2004) (criticizing ex post

The problem here is not new. The scope of a legal monopoly has been one of the dominant issues in the entire patent law, in which parties always fight and fret about the scope of a patent grant. The famous Supreme Court decision in *O'Reilly v. Morse*⁶⁰ is on point. Morse sought to claim the use of the entire electromagnetic spectrum for communicating at a distance. As the Court held, the preclusive effect would be far too great relative to any incentive needed to spur the inventive impulse.⁶¹ A similar issue arises in a modern context over the question of how many different variations of a given basic molecule can be subsumed under a single patent. It is one of the perennial questions of patent law, which constantly seeks to balance the need for incentive on the one hand against the exclusionary features of the patent on the other.⁶²

The insertion of FDA regulation does not change that issue in the slightest simply because the agency is in a position to check for safety and effectiveness. The FDA is by no means the only government agency to discharge health and safety functions. The long constitutional history of safety and health regulation has been marked by the fear that safety regulation will be used as a cloak to create a monopoly position for one of the regulated parties. To give one simple example, a rule that requires that all milk sold within a given state be pasteurized has never been construed to keep out the second entrant to the market once the initial entrant has complied with all health standards. Quite the opposite, when this issue has been squarely presented under the dormant commerce clause, any insistence that local facilities be used to meet an objective standard has been brushed aside because of the well-justified fear that the rules in question will perpetuate local monopoly power.⁶³ The key task here is to find some way in which the question of competitive balance is consistent with the overall system of regulation. And if other communities are willing to trust the health and safety of their citizens to their own regulators, there better be some strong safety reason to dispel the obvious inference that the local regulation is intended to prop up a

justifications for not using the market system that call for centralized control).

60. 56 U.S. (15 How.) 62, 85-86 (1853).

61. *Id.*

62. See *In re Harnisch*, 631 F.2d 716, 718 (C.C.P.A. 1980) (discussing the scope of so-called Markush claims whereby different radicals are added to a standard chemical backbone). One irony is that Searle (later taken over by Pharmacia, and then Pfizer) sought to block the Merck patent on Vioxx by claiming that it was covered by an earlier Markush claim that Pfizer had filed. That claim was rejected in an exhaustive opinion in the ensuing interference action before the Patent and Trademark Office.

63. See, e.g., *C & A Carbone, Inc. v. Town of Clarkstown*, 511 U.S. 383, 394-95 (1994) (invalidating the requirement that out-of-state disposal operations ship waste through local facility); *Dean Milk Co. v. City of Madison*, 340 U.S. 349, 355-56 (1951) (striking down ostensible health justification for requiring Illinois milk producers to bottle within five miles of Madison).

local monopoly.

In dealing with the pharmaceutical industry, it is important to recognize that obtaining a patent is only the first hurdle to marketing a drug. The legal situation would become quite untenable if patent law refused to allow the patentee of an initial molecule or process to assert *Morse*-like rights over adjacent products, only to have its economic objectives undermined by the FDA's insistence that the new drug climb a higher hurdle than the previous one. The Supreme Court's dormant commerce clause jurisdiction makes *nondiscrimination* in the application of health and safety rules the touchstone of legality,⁶⁴ absent some very powerful showing of harm that an antidiscrimination norm cannot touch.⁶⁵ To be sure, the dormant commerce clause has no direct application in thinking about the proper reach of federal regulation. But the issues before the FDA are identical to those raised by state regulation. The use of a higher standard for the second and all subsequent entrants creates an indefensible form of discrimination that should not be tolerated on grounds of public policy, even if it were not the subject of constitutional challenge—which should be the case.⁶⁶ Using the FDA as an agent of industrial policy to exclude latecomers in the race represents an irresponsible use of public policy. The simple fact is that latecomers will always suffer a disadvantage, whether they are in a regulated or unregulated market. The newcomer will always have the first mover advantage. The private entrant that knows the effectiveness (both means and variance, across relevant subgroups) does not need to have the FDA warn it that new entry is likely to produce a meager rate of return, if such be the case. That firm can run the calculations itself to decide whether its new product could pry away enough of the market to make a difference. The only considerations that are relevant are those that turn on safety and effectiveness. Nothing about the economics of the situation suggests that the FDA should assert a strong role in keeping drugs off the market, given the other forms of regulation that are available on safety matters.

64. See, e.g., *City of Philadelphia v. New Jersey*, 437 U.S. 617, 626-27 (1978) ("But whatever New Jersey's ultimate purpose, it may not be accomplished by discriminating against articles of commerce coming from outside the State unless there is some reason, apart from their origin, to treat them differently.").

65. See, e.g., *Maine v. Taylor*, 477 U.S. 131, 151-52 (1986) (upholding Maine's restriction on the importation of baitfish because Maine had sufficient justification, including that it threatened Maine's fisheries).

66. For my defense of the doctrine of unconstitutional conditions, see Richard A. Epstein, *The Constitutional Protection of Trade Secrets Under the Takings Clause*, 71 U. CHI. L. REV. 57, 68 (2004) (noting that the doctrine of unconstitutional conditions "places limits on the ability of the government to require individuals to waive their constitutional rights, including those to property under the Takings Clause, in order to escape the burden of some regulatory exaction") (footnote omitted).

II. DRUG WITHDRAWALS AND RECALLS

A. What's Different About Withdrawals and Recalls?

The discussion of the initial approval process leads to the next question, which asks how the analysis of risk and reward changes once a drug that has made its way onto the market proves to have some unwanted, and perhaps fatal, side effects. How should the FDA proceed on matters of withdrawal and recall? The most evident difference between approval and withdrawal should be in the amount of information available with which to make any considered judgment about a drug's efficacy and safety. That is, the longer the period that a drug is on the market, the more information that can be acquired about its use. Large numbers of patients using a drug for a long period of time should also lead to more reliable judgments about the individual responses to the dosages that have been supplied in particular cases. It is commonplace for independent parties to run studies that compare the effectiveness of different drugs on the market.

In light of that information, the same considerations that govern the initial permission to use drugs should apply with greater force in the withdrawal stage. If the results of a drug turn out to be disappointing, we could expect prescriptions to dwindle and the drug to be pulled, without FDA interference. But in the more common situation, the results of drug usage are likely to be varied, whereby some people benefit enormously while others do not tolerate the drug well at all. At this point the situation differs from that on original launch in only one particular: There is direct experiential evidence on whether a drug works or does not work in individual cases. The question is how that new particular alters the balance between upstream and downstream control. The better information does not, I believe, reverse the balance of convenience that was in favor of downstream control. The same difficulties with heterogeneous responses counsel against making a collective decision that precludes individual choice that is based on superior, localized information, even in the presence of serious side effects.

The political dimensions of this choice are more difficult because of the following general relationship: The more potent drugs are likely to do more good and cause more harm than less potent drugs in the same class of treatment. Given the higher variance, the action at the tails of the bell-curved distribution becomes ever more vivid. The low end of the distribution cries out for a ban, but that comes at the high cost of blocking use at the other end, where the perceived benefits will receive less attention precisely because they are less dramatic than real cases of failure—a political attention bias, as it were. That said, even with high variance drugs, with the expanded levels of drug usage it should be more, not less, possible to figure out protocols that allow the separation of patients into those who do and do not benefit from drug treatment. Here are some of the

alternatives that are open for control, but only if the withdrawal is not ordered: The drug could be made available, as is typically the case, only by prescription. The physician in turn can limit the dosage, shorten the periods of time of use, mix it with other drugs in the same or different classes, and so on. Stopping the drug always remains an option, as does starting it again after a change in diet, physical condition, or other medications. In this way, the hope is that it will be possible to preserve the use of the drug for those for whom it supplies the greatest benefit while limiting or avoiding use altogether for others.

B. A Tale of Two (or Three) Withdrawals

Confirmation of these basic considerations is found by a closer look at the controversies surrounding the withdrawal of Rezulin and Vioxx from the marketplace.

1. Rezulin

In March, 1999, the FDA conducted extensive hearings over whether Rezulin should have been removed from the market.⁶⁷ During the course of these hearings, the persons calling for Rezulin's continued sale were not solely workers for Warner-Lambert and its affiliates. Many independent physicians and patients were quite insistent that the drug had done an immense amount of good and were adamant in their desire to continue to use it for themselves or to prescribe it to their patients.⁶⁸ Now it is possible to say that all these people are wrong and

67. FDA, Endocrinologic and Metabolic Drugs Advisory Committee Meeting No. 72 (Mar. 26, 1999), <http://www.fda.gov/ohrms/dockets/ac/99/transcript/3499t1a.pdf> [hereinafter FDA Committee Meeting].

68. *See id.* Here is one statement from Dr. Robert Busch: "[W]e could fill this room with patients who have benefited from troglitazone [Rezulin]." *Id.* at 23. Also on point is the more detailed statement of Dr. Steven V. Edelman, himself a diabetic:

We know the consequences of poorly controlled diabetes: blindness, dialysis, amputations, heart attacks, strokes, depression, and unfortunately much, much more. Every day in America over 400 people die directly due to the effects of diabetes, and it's so important to look at the risk of Rezulin versus the benefits of improved glucose control when you're looking at a very serious disorder that affects the quality of life of millions of Americans on a day-to-day basis.

If one death is too many, then, yes, take Rezulin off the market, but then you must also take off glucosfos insulin, sulfonylureas, Motrin, aspirin, Tylenol, and many other medications used to treat patients with cancer and HIV.

I follow over 500 people at the Veterans' Affairs Medical Center in UCSD who are

indeed wrong-headed in their convictions, or, worse, that they did not know that they were being harmed when they felt better, even though they knew the risks. And it is certainly possible to identify people who were hurt by the use of the drug.⁶⁹ But to worry about long-term consequences for persons who have to struggle day-by-day involves the supreme paternal confidence that we know far better what is good for people than they know for themselves. It is easy to make that assumption about drug use when people who are fortunate enough not to need any drug treatment at a particular time contemplate the problem in the abstract. The data always seem daunting, the medical evidence incomprehensible. But the level of comprehension radically changes when the choices to be made move out of the hypothetical realm and into choices that involve life-or-death decisions or matters of chronic pain. At this point the incentives alter. People will learn a great deal under stress and will have a very reliable feedback loop as to whether their choices are right or wrong: Do they feel better or worse? In the face of that evidence, why make the collective decision to force withdrawal of a drug when that decision makes a substantial portion of the population worse off?

2. *Vioxx*

The situation with *Vioxx* is of course different in that there was no FDA withdrawal order.⁷⁰ The response of the FDA was not to challenge the soundness of the Merck decision but to reassure an anxious public that its own vigilance does not end when products reach the marketplace.⁷¹ But there is little reason to tarry on the question of whether the FDA should have ordered the withdrawal before Merck acted. The real question is whether the withdrawal should have been ordered at all. What follows is the FDA summary of the explanation for Merck's decision to remove the drug, which earned Merck's implicit endorsement:

6. What are the likely long-term health effects, if any, of taking this product?

taking Rezulin therapy. You can't buy this drug back from these individuals because it has helped them to achieve and maintain control over their diabetes where previously it was not possible despite intensive efforts.

Id. at 24-25.

69. *See id.* The testimony of Dr. Sydney Wolfe, Director of Public Citizens Health Research Group, notes such: "[O]ur estimates of liver deaths from Rezulin up through the beginning of February of '99 [are] 43 deaths, including American and Japanese cases, from liver toxicity from this drug." *Id.* at 66-67.

70. *See supra* note 6 and accompanying text.

71. FDA, Statement on *Vioxx* and Recent Allegations and the Agency's Continued Commitment to Sound Science and Peer Review (Nov. 17, 2004), <http://www.fda.gov/bbs/topics/news/2004/NEW01136.html>.

The new study shows that Vioxx may cause an increased risk in cardiovascular events such as heart attack and strokes during chronic use.

7. What evidence supports the Public Health Advisory?

Merck's decision to withdraw Vioxx from the market is based on new data from a trial called the APPROVe [Adenomatous Polyp Prevention on VIOXX] trial. In the APPROVe trial, Vioxx was compared to placebo (sugar-pill). The purpose of the trial was to see if Vioxx 25 mg was effective in preventing the recurrence of colon polyps. This trial was stopped early because there was an increased risk for serious cardiovascular events, such as heart attacks and strokes, first observed after 18 months of continuous treatment with Vioxx compared with placebo.

8. Why wasn't the APPROVe trial stopped earlier?

The APPROVe trial began enrollment in 2000. The trial was being monitored by an independent data safety monitoring board (DSMB). It was not stopped earlier because the results for the first 18 months of the trial did not show any increased risk of confirmed cardiovascular events on Vioxx.

9. What did FDA know about the risk of heart attack and stroke when it approved Vioxx?

FDA originally approved Vioxx in May 1999. The original safety database included approximately 5000 patients on Vioxx and did not show an increased risk of heart attack or stroke. A later study, VIGOR (VIOXX GI Outcomes Research), [in patients with rheumatoid arthritis] was primarily designed to look at the effects of Vioxx on side effects such as stomach ulcers and bleeding and was submitted to the FDA in June 2000. The study showed that patients taking Vioxx had fewer stomach ulcers and bleeding than patients taking naproxen, another NSAID, however, the study also showed a greater number of heart attacks in patients taking Vioxx. The VIGOR study was discussed at a February 2001 Arthritis Advisory Committee and the new safety information from this study was added to the labeling for Vioxx in April 2002. Merck then began to conduct longer-term trials to obtain more data on the risk for heart attack and stroke with chronic use of Vioxx.⁷²

It is useful to follow the argument paragraph by paragraph. The first point in paragraph six is of course the reason for the concern. No one should make light of the risks of heart attack and stroke and no one will; these risks are vivid and well understood by professionals and patients alike. Anyone who is convinced of the truth of this data will inquire further and discover that Vioxx does not have a

72. See FDA, Vioxx (rofecoxib) Questions and Answers, *supra* note 5.

clean bill of health.

The difficulties begin with paragraph seven, where the use of Vioxx 25 mg created the increased risk of cardiovascular use “first observed after 18 months of continuous treatment with Vioxx compared with placebo.” The increased risk level was from 1.9% to 3.5% in populations that are at risk generally because of age and health difficulties. But even if these numbers are dead accurate, they cut *against* withdrawal from the market, not for it. There are many people who could benefit by some combination of lower dosage and shorter usage, or possibly lower dosage and longer usage. As is so often the case with clinical studies, it is not possible to do work that plots an explicit dosage-response level so that one could compile a table that says “with an X mg pill the increased risk of a cardiovascular injury is Y.” But anyone who is armed with specific knowledge of his or her own cardiac condition can combine this background information with that personal knowledge. When appropriate, they can experiment with altering dosage patterns, switching off between Vioxx and Celebrex, or switching to other forms of painkillers. Whether it is worthwhile to take a chance on limited and altered use depends in part on the other benefits and costs of the proposed regimen. The one point that can be made for sure is that a uniform decision to stop all Vioxx on a dime need not be the best course of action for all, or even most, of Vioxx users.

Paragraph eight is defensive in tone about the decision to allow the trial to progress as long as it did, but fails to explain why the use of Vioxx is not safe for eighteen months. Nor does it address the question of how long one must remain off Vioxx or other NSAIDs before it is safe to go on them again. Lots of sensible questions, very few conclusive answers.

Paragraph nine is part of the FDA defense of its own internal processes.⁷³ But its one concrete bit of information again counsels against the removal of the drug from the market. Vioxx seems to work better for persons with serious intestinal issues. No evidence exists as to whether its use could have perhaps

73. These processes had been subject to a scathing attack by Dr. David Graham (who was active in the Rezulin removal). *See, e.g.*, FDA Committee Meeting, *supra* notes 67, <http://www.fda.gov/ohrms/dockets/ac/99/transcript/349941b.pdf>; Merck Hearings, *supra* note 42 (statement of David Graham, Associate Director for Science, Office of Drug Safety, Ctr. for Drug Evaluation and Research, FDA). For the effective FDA rebuttal, see FDA, Statement by Dr. Steven Galson, Acting Director, Center for Drug Evaluation and Research (CDER), Regarding November 18, 2004, Committee on Finance of the U.S. Senate Hearing on Drug Safety and the Worldwide Withdrawal by Merck & Co., Inc., of Vioxx (Nov. 18, 2004), <http://www.fda.gov/bbs/topics/news/2004/NEW01138.html> (disclaiming Dr. Graham’s congressional testimony as not reflective of the FDA’s views). Nor is there any reason to set a presumption that the persons who take the most critical view of a current drug are likely to be correct. It is too risky to encourage endless escalation of judgments by giving the greatest credit to the most vocal critics.

saved the lives of the 16,000 people annually who would otherwise die from complications associated with ulcers and intestinal bleeding, but the polyps trials showed positive results for various intestinal disorders before they were halted.⁷⁴ Nor does the brief finding suggest that it is difficult for people to find out whether they are at greater risk for heart attacks or ulcers. But private downstream information, coupled with a good medical history, should be able to shed some real light on that question. No one doubts that there is a trade-off between Vioxx and naproxen, but this tradeoff does *not* play out in the same way in all cases. The additional warning should be able to counter the risk, given the stakes involved. There is no evidence that similar effort was used to explain the advantages of Vioxx on the label.

In short, the landscape reveals a picture in which Vioxx is better in some circumstances and worse in others. The only case in which the FDA should urge the ban is when some other drug *dominates* Vioxx on all relevant dimensions. Otherwise, downstream judgments, which seem to follow easily from the presented data, seem preferable. Yet it is quite striking that the denunciations of both the FDA and Merck do not refer to the benefit side, but simply reiterate the position that the FDA continues to operate as the “gold standard” of review, more stringent than that found anywhere else in the world.⁷⁵ Yet it is just that inflated view of its mission, and the unthinking assertion that higher standards for marketing approval lead to better health outcomes, that lies behind the entire misconceived mission of the FDA. There is, in practice, a massive difference between the sensible effort to prevent fraud and adulteration and the constant desire to make omnibus cost/benefit analyses, which all too often miscarry in the individual cases. In retrospect, it seems unwise to have withdrawn Vioxx given the problems that have come to light with both Celebrex and Bextra.

3. Celebrex

There is little reason to offer the details on the Celebrex situation (at least today) for the arguments are parallel to those with respect to Vioxx. So long as the risk is disclosed and known, any ban looks to be strongly overinclusive. Shorter periods and lower dosages of the drug may be appropriate. Indeed, if Vioxx were still on the market, some alternation between these two drugs might

74. See Gina Kolata, *Good Pill, Bad Pill: Science Makes It Hard To Decipher*, N.Y. TIMES, Dec. 22, 2004, at A1 (“In one of the great examples of the mixed messages of science, the same study that killed the blockbuster arthritis drug Vioxx after showing that it had heart risks also found that the drug had a significant benefit: it prevented precancerous colon polyps in some patients, one of the study’s principal researchers said.”).

75. Merck Hearings, *supra* note 42 (statement of Sandra L. Kweder, Deputy Director, Office of New Drugs, FDA).

have been a viable strategy. The calls for the return to aspirin or other NSAIDs as the painkillers of choice should not be dismissed out of hand, for it might be the appropriate response for some people. But the bleeding risks associated with its use do raise this irony: Were it not for its grandfathered status, could aspirin pass the new standards for getting on the markets if launched today? So long as downstream controls are available, Celebrex should remain on the market. Its sales may well fall in response to the new information, which is just fine; however, the total ban is not.

CONCLUSION: VARIATION AND BENEFITS

This analysis of FDA practices should give rise to multiple sources of concern for what is, and is not, taken into account. As is evident, all adverse effects receive maximum attention and lead to a chorus that calls for caution above all. The entry of new drugs should be slowed, greater supervision should be given to drugs that are already on the market, and strong products liability, fraud and breach of warranty suits should be pressed into service to back up the regulatory apparatus. This evident social consensus helps explain the reactions to both Vioxx and Celebrex, and seems in many cases to be supported by that oldest of medical maxims, *primum non nocere*, first do not harm. But unfortunately, the relevant considerations make it clear that this maxim—or any akin to it, such as “better safe than sorry”—does not capture the full set of relevant considerations in any cost/benefit analysis applicable to pharmaceutical products. Gains in these cases matter as much as losses, and members of the public are not “safe” if public policy causes the failure to get some new, albeit risky, therapy, and this failure results in serious impairments followed by death. In dealing with serious medical questions, there is no risk-free alternative that acts as the baseline from which these time-honored maxims can take place. It is dreaming to think that any upstream federal drug policy can eliminate risk. Necessarily, there is harm in *not* giving risky drugs that are beneficial just as there is harm in giving potent drugs with devastating side effects. Both kinds of error are always in the mix.

In light of this simple but sober truth, this nation should rethink its basic drug policy on all three matters discussed herein. So long as benefits count and so long as individual responses to standard treatments vary, individualized downstream determinations should trump standardized government calculations. The current call for reform finds an easy target when it takes the stance that whatever is good for the drug houses is bad for the American people. But that statement makes no more sense today than Engine Charlie Wilson’s famous remark of fifty years ago: “What’s good for the country is good for General Motors and vice versa.” Unfortunately, the world is a messier place than either of these bromides suggest. Sometimes social welfare aligns with the release and use of new drugs, sometimes not. More often, it is the former, not the latter, so long

as individual choice is available. Our pharmaceutical paternalism comes at a very high price, and we make a major mistake when our regulatory system sets its face against the introduction of new drug therapies. As the old song says, you always hurt the ones you love.

EPILOGUE

It is always dangerous business to write a scholarly article about an issue that is in full flux. That proposition has proved itself time and again during the revisions of this Article on the proper role for the FDA. The final revisions of this Article took place just after an advisory panel to the FDA recommended that Bextra, Celebrex, and Vioxx be left on the market. On April 7, 2005, the FDA confounded most observers by going beyond the recommendation of its panels. It requested that Pfizer remove Bextra from the marketplace and that black box warnings be put on Celebrex and a long list of NSAIDs.⁷⁶

In the FDA's brief advisory, it gave this explanation for its decision:

In reaching these decisions, FDA has carefully considered the available data on all of the NSAIDs. The Agency has also considered presentations, discussions, and votes from the joint public meeting of the FDA Arthritis Advisory Committee and the Drug Safety and Risk Management Advisory Committee held on February 16, 17, and 18, 2005 to discuss the CV safety concerns for these drugs along with their overall risk-benefit.⁷⁷

Which is to say, it gave no explanation at all. Nothing in its actions leads me to change my views. Unfortunately, I fear that no reasoned argument will lead the FDA to reconsider its views. My Article stands as is.

76. FDA, Public Health Advisory, FDA Announces Important Changes and Additional Warnings for COX-2 Selective and Non-Selective Non-Steroidal Anti-Inflammatory Drugs (NSAIDs) (Apr. 7, 2005), <http://www.fda.gov/cder/drug/advisory/COX2.htm>.

77. *Id.*

The Patient's Role in Choice of Medications: Direct-to-Consumer Advertising and Patient Decision Aids

Marshall H. Chin, M.D., M.P.H.*

INTRODUCTION

Explicit patient involvement in the selection of medications has become more frequent. Pharmaceutical companies have targeted lay persons for direct-to-consumer advertising (DTCA), and the rise of the patient empowerment movement has helped lead to more egalitarian models of shared decision making between patient and physician. This Article explores the challenge of involving patients more actively in medication choice through DTCA and patient decision aids. I will outline the optimal conditions for shared decision making between patients and physicians in drug selection and then discuss some of the key evolving issues surrounding increased patient involvement in the drug selection process. In particular, I will explore the flow of information to patients, with a specific emphasis on issues involved with DTCA, and also cover some of the challenges and promise of patient decision aids for the choice of medication. The former will cover some of the difficult macro health policy issues related to free speech and consumer protectionism while the latter will address some of the practical challenges of trying to improve patient decision making at the level of the individual clinical encounter. Current regulatory and enforcement practices have been insufficient to prevent the dissemination of some inaccurate or misleading advertisement. Informed, empowered patients can make decisions with their physicians that are more likely to be consistent with their values and preferences.

I. TRENDS IN PATIENT INVOLVEMENT

In the debate over the creation and diffusion of pharmaceutical products, active involvement of the patient has been an afterthought until recently. During the twentieth century, the pharmaceutical industry developed drugs, the FDA

* Section of General Internal Medicine, Department of Medicine; Diabetes Research and Training Center, The University of Chicago.

regulated drugs,¹ and the physician was the primary target for marketing of drugs. However, the sociopolitical environment has evolved. Over the past fifteen years or so, patient empowerment has become an increasingly valued goal within the health care field.² Patient self-management is crucial for chronic disease care, which now comprises a large percentage of health care in the United States.³ As a result, substantial attention has been devoted to finding innovative ways to get patients more actively involved in their care.⁴

This trend toward increasing patient empowerment has provided a fertile context for involving patients more directly in the medication selection process. From a marketing standpoint, pharmaceutical companies have realized the value of DTCA.⁵ Between 1997 and 2001, DTCA spending increased from \$1.1 billion to \$2.7 billion per year.⁶ DTCA is likely to remain common, as the advertising

1. See John P. Swann, FDA, History of the FDA, at <http://www.fda.gov/oc/history/historyoffda/default.htm> (last visited Feb. 16, 2005).

2. Patient empowerment refers generally to patients playing a more active role in their care whereas patient self-management denotes the actual tasks that patients must do to manage their illnesses such as taking medications, following a diet, and exercising. See Martha Funnell et al., *Implementing an Empowerment-Based Diabetes Self-Management Education Program*, 31 DIABETES EDUCATOR 53-56 (2005). The move toward patient empowerment reflects broader societal trends, traceable to the Civil Rights Movement, Vietnam War protests, and the rise of feminism in the 1960s and 1970s, in which paternalism and authority have been challenged, and individual autonomy has become increasingly treasured. See FRANK FREIDEL & ALAN BRINKLEY, *AMERICA IN THE TWENTIETH CENTURY* 449-92 (5th ed. 1982).

3. Chronic disease is prevalent and costly in the United States. See Catherine Hoffman et al., *Persons with Chronic Conditions: Their Prevalence and Costs*, 276 JAMA 1473, 1473 (1996). For example, between 1999 and 2002, 30.1% of Americans over the age of twenty had hypertension. See NAT'L CTR. FOR HEALTH STATISTICS, *HEALTH, UNITED STATES, 2004: WITH CHARTBOOK ON TRENDS IN THE HEALTH OF AMERICANS* 238 (2004). The quality of chronic care management is frequently inadequate, and the Institute of Medicine has identified patient-centered care as one of the foundations of quality care. See INST. OF MED., *CROSSING THE QUALITY CHASM: A NEW HEALTH SYSTEM FOR THE 21ST CENTURY* (2001).

4. See Sheldon Greenfield et al., *Patients' Participation in Medical Care: Effects on Blood Sugar Control and Quality of Life in Diabetes*, 3 J. GEN. INTERNAL MED. 448, 448 (1988); Edward H. Wagner et al., *Organizing Care for Patients with Chronic Illness*, 74 MILBANK Q. 511, 512 (1996).

5. Pharmaceutical companies traditionally advertised drugs to physicians, hospitals, and other providers through print advertisements, marketing at medical meetings, distribution of free samples of medications, and direct visits from sales representatives. Direct-to-consumer advertising (DTCA) bypasses these intermediaries and markets drugs to patients directly through a variety of media including television, newspapers, magazines, direct mail, and the internet. Patients still need a prescription for drugs that require one, but the marketing of the drug is directly to the consumer.

6. U.S. GEN. ACCOUNTING OFFICE, *PRESCRIPTION DRUGS: FDA OVERSIGHT OF DIRECT-TO-*

has been effective.⁷ In addition, novel ways to improve the shared decision making process between patient and physician, such as patient decision aids,⁸ have shown promise. Yet, significant concerns with regard to DTCA and decision aids persist.

II. SHARED DECISION MAKING

During a clinical encounter, the patient may be influenced by a variety of factors, including DTCA, personal beliefs, and the experiences of family and friends. Similarly, many factors influence physicians, including their medical education and drug advertising. Collectively, scientific evidence, physician clinical judgment, and patient preferences become incorporated into the decision making process and ultimately lead to a variety of outcomes in the drug selection process. Each of these three elements should be weighed differently depending upon the individual circumstance. For example, in some cases, clear scientific evidence shows the benefit of particular medications, such as beta blockers and angiotensin-converting enzyme inhibitors in many patients with heart failure.⁹ Patients with this indication should generally receive these medications. In other situations, clinical judgment and patient preferences are essential. For example, little scientific evidence exists to guide the management of older persons with diabetes.¹⁰ Whereas most younger patients with diabetes are likely to benefit from tight glucose control and intensive treatment of their cardiovascular risk factors, the situation is more variable among older persons. Some older persons

CONSUMER ADVERTISING HAS LIMITATIONS 10 (2002), <http://www.gao.gov/new.items/d03177.pdf>.

7. Advertisements for prescription drugs are regularly displayed on television and radio and in print for the lay public. The General Accounting Office reported DTCA increases both prescription drug spending and utilization. Between 1999 and 2000, prescriptions for the fifty most heavily advertised drugs increased thirty-two percent compared to fourteen percent for all other drugs. Most of the increase in expenditures resulted from increased utilization rather than increased prices. *Id.* at 11-12.

8. Patient decision aids are evidence-based tools, such as pamphlets, workbooks, interactive CD-ROMS or videodiscs, that are designed to help inform patients, clarify their values, provide communication skills to facilitate interactions with physicians, and allow them to make decisions that more accurately reflect their true wishes. See Annette M. O'Connor et al., *Modifying Unwarranted Variations in Health Care: Shared Decision Making Using Patient Decision Aids: A Review of the Evidence Base for Shared Decision Making*, HEALTH AFF., Oct. 7, 2004, at VAR-64.

9. See Sharon A. Hunt et al., *ACC/AHA Guidelines for the Evaluation and Management of Chronic Heart Failure in the Adult: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines*, 38 J. AM. COLL. CARDIOLOGY 2101, 2109 (2001).

10. See Elbert S. Huang et al., *Practical Challenges of Individualizing Diabetes Care in Older Patients*, 30 DIABETES EDUCATOR 558, 558 (2004).

with diabetes are relatively healthy and likely to accrue the benefits of aggressive treatment, and others may have other life-limiting conditions that attenuate the benefit of intensive control.¹¹ Patient preferences are particularly critical in what John Wennberg has called “preference-sensitive” conditions.¹² These are conditions in which the relative costs and benefits of treatment options are unclear and subject to the relative values the individual patient places on them. For example, older men with benign prostatic hyperplasia may have difficulty urinating. These men may be treated by either drug or surgical treatment and the choice partly depends upon patient preferences.¹³

The shared decision making process emphasizes that neither the patient nor the physician can be viewed in isolation. While interventions to influence behavior such as DTCA may be directed at a single party, the ultimate decision and whether or not the patient adheres to the treatment over time are influenced by the interaction between the patient and physician. The ideal outcome of the shared decision making process is variable depending upon one’s perspective and goal. Possibilities include clinical outcomes, a cost-effective outcome, a better decision or decision process, concordance of patient values and the choice made, patient autonomy, and equity. The challenge is that these outcomes frequently conflict, particularly when weighing individual patient values versus societal values. For example, patients with moderate asthma benefit from inhaled corticosteroids.¹⁴ They have decreased hospitalizations and improved quality of life. Yet some patients have an aversion to drugs and would prefer a more natural non-drug treatment approach. Though this approach may not be as effective as conventional medicines, the choice these patients make is more consistent with their values of a natural healing philosophy. Whether or not this is an optimal outcome or not depends upon how individual and societal values are prioritized.

Individual and societal cost-effectiveness tradeoffs can be particularly difficult to reconcile. For example, in patients with prior myocardial infarction, prior stroke, or peripheral vascular disease, clopidogrel is more effective than aspirin in preventing a recurrent vascular event.¹⁵ The estimated increased cost

11. Cal. Healthcare Found. & Am. Geriatrics Soc’y Panel on Improving Care for Elders with Diabetes, *Guidelines for Improving the Care of the Older Person with Diabetes Mellitus*, 51 J. AM GERIATRICS SOC’Y S265, S266 (2003).

12. John E. Wennberg et al., *Geography and the Debate over Medicare Reform*, HEALTH AFF., Feb. 13, 2003, at W96, W100.

13. See Carl Gjertson et al., *Benign Prostatic Hyperplasia: Now We Can Begin to Tailor Treatment*, 71 CLEV. CLIN. J MED. 857, 857 (2004).

14. See NAT’L ASTHMA EDUC. & PREVENTION PROJECT, NAT’L INSTS. OF HEALTH, PUB. NO. 97-4051, EXPERT PANEL REPORT 2: GUIDELINES FOR THE DIAGNOSIS AND MANAGEMENT OF ASTHMA 60 (1997).

15. CAPRIE Steering Comm., *A Randomised, Blinded, Trial of Clopidogrel Versus Aspirin in*

for patients with a prior stroke who receive clopidogrel rather than aspirin is \$31,200 per quality adjusted life year.¹⁶ For the individual patient, clopidogrel would seem to be the best choice. In addition, the \$31,200 incremental cost-effectiveness figure is less than the \$50,000 threshold traditionally used in medical cost-effectiveness analysis for defining a "cost-effective" treatment.¹⁷ However, if there is a fixed health care budget such as within a capitated health care plan or conceivably a government budget such as a state's appropriation to the Medicaid program, whether or not the use of clopidogrel rather than aspirin is the most cost-effective way to allocate resources is less clear.

III. DIRECT-TO-CONSUMER ADVERTISING

Direct-to-consumer advertising leads to more drug information for patients. The regulatory challenge is ensuring that accurate, understandable data are provided in the advertisements.

A. Flow of Information to Patients

Traditionally, patients have received most of their information about prescription drugs from their physicians.¹⁸ This mechanism assumes that physicians know the relevant drug information and have the inclination, time, and communication skills necessary to convey the data to their patients. Recently, however, DTCA has been used as a marketing technique and way to reach patients directly in order to reduce the information and power imbalance between patient and physician.

DTCA holds promise for improving patient education and patient empowerment, making patients more effective partners in their care, and encouraging patients to seek treatment for conditions that may be underdiagnosed

Patients at Risk of Ischemic Events (CAPRIE), 348 THE LANCET 1329, 1333 (1996).

16. Mark D. Schleinitz et al., *Clopidogrel Versus Aspirin for Secondary Prophylaxis of Vascular Events: A Cost-Effectiveness Analysis*, 116 AM. J. MED. 797 (2004).

17. Cost-effectiveness thresholds are controversial and highly dependent upon the societal and decisional context. MARTHE GOLD ET AL., COST-EFFECTIVENESS IN HEALTH AND MEDICINE 295 (1996). However, the \$50,000 cost-effectiveness threshold has been a commonly cited benchmark for discussion purposes. See Peter A. Ubel et al., *What Is the Price of Life and Why Doesn't It Increase at the Rate of Inflation?* 163 ARCHIVE INTERNAL MED. 1637, 1637 (2003).

18. Between 1997 and 2001, the pharmaceutical industry directed more than eighty percent of its promotional expenditures towards physicians. These efforts included giving drug samples to physicians and sending sales representatives to speak to physicians. See U.S. GEN. ACCOUNTING OFFICE, *supra* note 6, at 10. In addition, physicians have retained control over the diffusion of most medical knowledge to patients. See PAUL STARR, THE SOCIAL TRANSFORMATION OF AMERICAN MEDICINE 3-29, 79-144 (1982).

or stigmatized. In particular, DTCA can legitimize patients' valid but untreated health concerns such as depression or erectile dysfunction, and other stigmatized conditions.¹⁹ However, if the advertisements do not fairly convey the risks and benefits of the drug,²⁰ patient misperception of the drug's effectiveness could lead to patient pressure to prescribe inappropriate drugs. Patient misperception of DTCA could also lead to patient pressure to prescribe new drugs that have no significant benefit over similar, cheaper, older medications.²¹

In a recent national survey of 643 physicians regarding their perception of DTCA,²² overall perceptions of DTCA were mixed. Forty percent of respondents thought that DTCA had a positive effect, 30% thought it had a negative effect, and 30% thought it had no effect.²³ About 70% of respondents reported that DTCA helped educate and inform patients about treatments and led to better discussions. However, about 80% perceived that DTCA had unbalanced information and led to unnecessary treatments.²⁴ Only 32% of physicians reported that patients had less confidence in physician judgment,²⁵ and 39% stated that they had prescribed a drug because of DTCA.²⁶ The most common conditions that patients have discussed with their physicians as a result of DTCA included traditionally underdiagnosed ones such as impotence (11%) and depression (6%),²⁷ conditions which were also common among the new diagnoses that were made based upon a visit spurred on by DTCA.²⁸ Diagnosis is the first step on the path to treatment and improved quality of life for conditions such as depression, since diseases must be recognized before they are treated.

B. Regulatory Challenges

The challenge for the FDA, Centers for Medicare and Medicaid Services (CMS), and other regulatory and financing agencies is how to best balance

19. See C.E. Hoesl et al., *Erectile Dysfunction (ED) is Prevalent, Bothersome and Underdiagnosed in Patients Consulting Urologists for Benign Prostatic Syndrome (BPS)*, 47 EUR. UROLOGY 511 (2005); Andrew A. Nierenberg, *Current Perspectives on the Diagnosis and Treatment of Major Depressive Disorder*, 7 AM. J. MANAGED CARE S353 (2001).

20. See Joel S. Weissman et al., *Physicians Report on Patient Encounters Involving Direct-to-Consumer Advertising*, HEALTH AFF., Apr. 28, 2004, at W4-219, W4-224 (2004).

21. U.S. GEN. ACCOUNTING OFFICE, *supra* note 6, at 1.

22. Weissman et al., *supra* note 20.

23. *Id.* at W4-224.

24. *Id.*

25. *Id.*

26. *Id.* at W4-225.

27. *Id.* at W4-224.

28. *Id.* at W4-225.

innovation, consumer protection, cost containment, and free speech. The political and economic environments are complex. With the re-election of President George W. Bush, one would initially think that the ideology of deregulation would reign supreme and that government intrusion into the drug market would be minimized.²⁹ However, several factors make increased government regulation of drugs a real possibility over time. We are in an era of continually rising health care costs, and the federal government is about to become the largest purchaser of drugs in 2006 when the Medicare prescription drug benefit takes effect. Many analysts believe that the cost of this benefit has been grossly underestimated.³⁰ Severe economic pressures are likely to cause increased demands for justification of the value of prescription drugs. Even today, CMS has been enacting policies that will reimburse certain pharmaceutical products only if studies demonstrate that they have a beneficial effect.³¹

Currently the FDA has several requirements for prescription drug advertisements, which include that advertisements must be neither false nor misleading, and must provide a "fair balance" of information about risks and benefits, "facts" that are "material" to advertised use of the drug, and a "brief summary" that discloses every risk from the product's approved labeling or else "adequate provision" for disseminating the labeling.³² These preceding terms are

29. While the overall picture is mixed, President George W. Bush has favored deregulation in several industries such as energy. See NAT'L ENERGY POLICY DEV. GROUP, RELIABLE, AFFORDABLE, AND ENVIRONMENTALLY SOUND ENERGY FOR AMERICA'S FUTURE (2001), <http://www.whitehouse.gov/energy>.

30. Many "Baby Boomers" are within four years of cashing their first social security checks. Dr. Mark B. McClellan, the Administrator of CMS, estimates that the Medicare prescription drug benefit program will cost \$720 billion over ten years rather than the \$400 billion originally estimated by the Congressional Budget Office. Robert Pear, *New White House Estimate Lifts Drug Benefit Cost to \$720 Billion*, N.Y. TIMES, Feb. 9, 2005, at A1.

31. Gina Kolata, *Medicare Covering New Treatments, but with a Catch*, N.Y. TIMES, Nov. 5, 2004, at A1. Tiered formularies in the private sector represent another example of how cost pressures are likely to create incentives to restrict choice of medications for patients. See Steven A. Pearson et al., *Changing Medication Use in Managed Care: A Critical Review of the Available Evidence*, 9 AM. J. MANAGED CARE 715, 715 (2003); Thomas S. Rector et al., *Effect of Tiered Prescription Copayments on the Use of Preferred Brand Medications*, 41 MED. CARE 398, 399 (2003).

32. 21 C.F.R. § 202.1 (2004); KATHRYN AIKIN ET AL., PATIENT AND PHYSICIAN ATTITUDES AND BEHAVIORS ASSOCIATED WITH DTC PROMOTION OF PRESCRIPTION DRUGS—SUMMARY OF FDA SURVEY RESEARCH RESULTS: FINAL REPORT 9-10 (2004), <http://www.fda.gov/cder/ddmac/Final%20Report/FRfinal111904.pdf>; James M. Jeffords, *Direct-to-Consumer Drug Advertising: You Get What You Pay For*, HEALTH AFF., Apr. 28, 2004, at W4-253, W4-254.

frequently difficult to define precisely, leaving much latitude for how aggressively the FDA pursues consumer protection.³³ In addition, challenges exist in meeting the real intent of the law rather than merely the letter of the law. For example, many DTCA advertisements reproduce the product insert's lengthy list of side effects in a corner of the page, leading one FDA official to note the need to "help pharmaceutical companies design brief summaries that are potent public health tools rather than Mensa tests or eye exams."³⁴ Presented effectively, these summary boxes hold promise as a way to fulfill the legal requirement that information is comprehensible to consumers.³⁵

In 2002, the General Accounting Office (GAO) raised concerns over misleading DTCA and delays in FDA enforcement actions against them.³⁶ Representative Henry Waxman noted that delays remained in 2003, with as long as six months passing between complaint and action.³⁷ Moreover, the choice of enforcement sanction was frequently weak, such as a warning letter to the pharmaceutical company, as opposed to a stronger sanction such as one that would result in a financial penalty. Rep. Waxman summarized the situation by stating, "There simply is no incentive for drug manufacturers to tell the whole truth to consumers, and there is no real penalty for them if they do not."³⁸

In February 2004, the FDA released three guidance documents for manufacturers of pharmaceutical products and restricted devices.³⁹ The recommendations call for more consumer-friendly language for the risk and side effect information in print advertisements. They also recommend increasing help-seeking and disease-awareness advertisements that do not specifically market a

33. The FDA can send regulatory letters to companies when it finds a violation of DTCA rules. For examples, see U.S. GEN. ACCOUNTING OFFICE, *supra* note 6, at 18-20.

34. Peter J. Pitts, *Turning Point or Tipping Point: New FDA Draft Guidances and the Future of DTC Advertising*, HEALTH AFF., Apr. 28, 2004, at W4-259, W4-261.

35. Steven Woloshin et al., *The Value of Benefit Data in Direct-to-Consumer Drug Ads*, HEALTH AFF., Apr. 28, 2004, at W4-234, W4-235.

36. U.S. GEN. ACCOUNTING OFFICE, *supra* note 6, at 21-23.

37. Henry A. Waxman, *Ensuring that Consumers Receive Appropriate Information from Drug Ads: What Is the FDA's Role?*, HEALTH AFF., Apr. 28, 2004, at W4-256, W4-257.

38. *Id.*

39. FDA, GUIDANCE FOR INDUSTRY: BRIEF SUMMARY: DISCLOSING RISK INFORMATION IN CONSUMER-DIRECTED PRINT ADVERTISEMENTS (2004), <http://www.fda.gov/cber/gdlns/consumad.pdf>; FDA, DRAFT GUIDANCE FOR INDUSTRY AND FDA, CONSUMER-DIRECTED BROADCAST ADVERTISING OF RESTRICTED DEVICES (2004), <http://www.fda.gov/cdrh/comp/guidance/1513.pdf>; FDA, GUIDANCE FOR INDUSTRY: "HELP-SEEKING" AND OTHER DISEASE AWARENESS COMMUNICATION BY OR ON BEHALF OF DRUG AND DEVICE FIRMS (2004), <http://www.fda.gov/cber/gdlns/helpcomm.pdf>; Press Release, FDA, New FDA Draft Guidances Aim To Improve Health Information (Feb. 4, 2004), <http://www.fda.gov/bbs/topics/NEWS/2004/NEW01016.html>.

drug but which would help drug sales indirectly by raising consumer awareness of the condition that the drug is intended to treat. These documents do not, however, mention FDA commitment to increased enforcement of FDA pharmaceutical advertising regulations.⁴⁰

One could argue that some lenience in enforcing existing advertising regulations might be permitted because the patient must work with the physician, a presumably knowledgeable intermediary, to obtain prescription drugs. Although this mechanism offers another level of protection, physicians frequently misperceive the relative benefits of medications and thus may provide inaccurate information to the patient.⁴¹ Regardless of how DTCA and its regulation evolve over time, the shared decision making encounter between patient and physician will remain a vital part of the pathway for actual use of the medication. Thus, improvements in the societal regulatory process must be made simultaneously as the individual patient-physician shared decision making process is enhanced.

IV. PREFERENCE-SENSITIVE CONDITIONS, DECISION AIDS, AND SHARED DECISION MAKING

Preference-sensitive conditions are ones in which the optimal treatment choice depends upon patients' preferences or values for benefits, harms, and uncertainties. I have previously described a shared decision making ideal that incorporates scientific evidence, clinical judgment, and patient preferences. In actual clinical practice, preference-sensitive conditions are often difficult, controversial, and time-consuming problems for physicians to address adequately with patients. For example, prostate specific antigen screening, discussion of surgical versus medical management of certain cancers, and aggressive treatment of chronic illnesses in frail older persons each share these complexities.⁴²

One way to explore patient preferences is through patient decision aids.⁴³ Compared to standard patient education materials, patient decision aids tend to be more interactive and go beyond merely imparting facts since the goal is to help the patient identify the best clinical decision for him or herself. While they have

40. See sources cited *supra* note 39.

41. Marshall H. Chin et al., *Differences in Generalist and Specialist Physicians' Knowledge and Use of Angiotensin-Converting Enzyme Inhibitors for Congestive Heart Failure*, 12 J. GEN. INTERNAL MED. 523 (1997).

42. See, e.g., Louise C. Walter & Kenneth E. Covinsky, *Cancer Screening in Elderly Patients: A Framework for Individualized Decision Making*, 285 JAMA 2750 (2001); U.S. PREVENTIVE SERVS. TASK FORCE, U.S. DEPT. HEALTH & HUM. SERVS., GUIDE TO CLINICAL PREVENTIVE SERVICES (2d ed. 1996).

43. See *supra* note 8.

been used in the research setting, diffusion of such aids has been limited in general clinical practice.⁴⁴ However, several factors make patient aids attractive as part of a program to help patients choose medications for preference-sensitive conditions and other situations where the appropriate decision is unclear. First, patient decision aids can impart information and explore relevant factors in much more detail than is possible in brief DTCA commercials. Second, given the complexity and time-consuming nature of understanding and discussing some medical decisions, aids can supplement the actual patient-physician dialogue about choice of medication. In addition, most physicians receive relatively little training on how best to conduct such discussions and may not be aware of the most recent scientific evidence. The patient decision aid can serve as a reference tool for both patient and physician and can provide patients with the information and communication skills necessary to engage in a meaningful conversation with their physicians.

One of the best studied clinical decision aids exploring choice of a drug was for postmenopausal hormone replacement prior to the publication of literature demonstrating negative cardiovascular effects for these medications.⁴⁵ Patient decision aid tools described benefits and risks of postmenopausal hormone replacement with information individualized for the specific patient's risk stratum. The aids also described the probabilities of disease with and without hormone replacement therapy, and helped clarify the patient's values regarding how they rated the relative benefits, risks, and uncertainties of the therapy.⁴⁶ Compared to patients who received an information pamphlet, patients with the decision aid had more realistic expectations of the benefits and risks, lower decisional conflict, and higher perceived acceptability of the intervention.

Patient decision aids may be used for a variety of clinical situations. For example, either Cox-2 inhibitors or less expensive non-steroidal anti-inflammatory agents can be used for arthritis. These two different drug classes

44. Patient decision aids might be used prior to a physician visit or after an initial visit. They could also be used at home or in the doctor's office.

45. Annette M. O'Connor et al., *A Decision Aid for Women Considering Hormone Therapy After Menopause: Decision Support Framework and Evaluation*, 33 PATIENT EDUC. COUNS. 267 (1998). Prior to the publication of literature demonstrating negative cardiovascular effects for these medications, the decision to prescribe postmenopausal hormone replacement therapy involved consideration of osteoporosis prevention, treatment of symptoms, and, at that time, possible cardiovascular protection versus risk of breast cancer and endometrial cancer. See Herbert B. Peterson et al., *Hormone Therapy: Making Decisions in the Face of Uncertainty*, 164 ARCHIVE INTERNAL MED. 2308 (2004).

46. Annette M. O'Connor et al., *Randomized Trial of a Portable, Self-Administered Decision Aid for Postmenopausal Women Considering Long-Term Preventive Hormone Therapy*, 18 MED. DECISION MAKING 295, 295-298 (1998).

have different side effect profiles. Factors to consider when choosing between these two drug classes are pain relief, gastrointestinal side effect profiles in high-risk patients, and, now, uncertainty over potential negative cardiovascular events in the wake of the removal of rofecoxib (Vioxx) and valdecoxib (Bextra) from the market.⁴⁷ Even within-class medication choices could be the subject of patient decision aids. For example, statins for hypercholesterolemia could be compared based upon relative efficacy, cost to the patient or society, and possibly length of time on the market, since there is likely more uncertainty regarding possible side effects of newer medications since they have been used in significantly fewer patients. A Cochrane review of patient decision aids found that patients who used decision aids had greater knowledge, more realistic expectations, and lower decisional conflict.⁴⁸ The study also found that patients using decision aids were more active in decision making, and demonstrated improved agreement between values and choices.⁴⁹ Cost-effectiveness of decision aids has not been studied in great detail, but United Kingdom trials of menorrhagia, menopause, and benign prostatic hyperplasia reported cost-neutral or cost-saving patient decision aid interventions.⁵⁰

O'Connor et al. describe a process for decision support for preference-sensitive conditions that involves brief counseling and referral to intensive decision support as needed.⁵¹ The goals of brief counseling are to clarify the decision by discussing benefits, harms, uncertainties, and costs; clarify values;

47. On May 21, 1999, the FDA granted approval to Merck to place rofecoxib (Vioxx) on the market. By 2001, concerns had been raised about possible cardiovascular side effects in medications of this drug class. A later trial demonstrated increased rates of myocardial infarction or stroke in patients taking rofecoxib. Concern has been raised whether Merck and the FDA should have taken earlier action to investigate possible adverse cardiovascular effects of rofecoxib or else withdraw the drug from the market. See Jennifer Couzi, *Withdrawal of Vioxx Casts a Shadow over Cox-2 Inhibitors*, 306 SCIENCE 384, 384-85 (2004); Debabrata Mukherjee et al., *Risk of Cardiovascular Events Associated with Selective Cox-2 Inhibitor*, 286 JAMA 954, at 954 (2001); Eric J. Topol, *Arthritis Medicines and Cardiovascular Events—"House of Coxibs,"* 293 JAMA 366, 366 (2005); Eric J Topol, *Failing the Public Health—Rofecoxib, Merck, and the FDA*, 351 NEW ENG. J. MED. 1707, 1707 (2004). The FDA also persuaded Pfizer to remove valdecoxib (Bextra) from the market because of possible cardiovascular side effects. See Gardiner Harris, *FDA Announces Strong Warnings for Painkillers: Pfizer Drug Withdrawn*, N.Y. TIMES, Apr. 8, 2005, at A1.

48. Annette O'Connor et al., *Decision Aids for People Facing Health Treatment or Screening Decisions*, 1 COCHRANE DATABASE SYS. REV. CD001431 (2003).

49. See MICHAEL F. DRUMMOND ET AL., *METHODS FOR THE ECONOMIC EVALUATION OF HEALTH CARE PROGRAMMES* (2d ed. 1997).

50. Annette M. O'Connor et al., *supra* note 43, at 67.

51. Annette M. O'Connor et al., *Risk Communication in Practice: The Contribution of Decision Aids*, 327 BRIT. MED. J. 736, 737-39 (2003).

and attempt to determine if benefits exceed harms; and screen for problems in the decision making process including decisional uncertainty, knowledge deficits, lack of clarity regarding values, and problems with support. The intensive phase of decision support involves assessing patient needs as well as their barriers to decision making. Then interventions are tailored to specific barriers such as insufficient information, difficulty clarifying what values are important to the patient, and lack of communication skills required for interacting most effectively with his or her physician.

While patient decision aids and decision support have promise, several significant challenges exist. Key problems include identifying the most useful situations to use aids and identifying patients who are comfortable judging the risks involved. For example, some patients prefer a paternalistic, authoritarian approach from their physicians, while others want to play an active role as full partners in their care who make the final decisions. Other concerns are whether fair, relevant, useful, timely, and comprehensible information can be ensured in decision aids. In particular, decision aids need to be understandable when describing probabilities and risk information for patients. Additional challenges include architectural design issues with multimedia decision aids and finding the best mix of text, graphics, and audio to communicate information most effectively.

Trying to include individual and societal costs into the decision making process for patients presents another challenge for decision aids. Societal economic costs are frequently difficult to incorporate into the individual decision making process if the patient has insurance that insulates him or herself from true costs. In contrast, out-of-pocket costs to the patient may be more feasible to include. For example, tiered pharmaceutical insurance plans offer a choice of medications that have differential cost implications for patients and these options can be explicitly presented to the patient. Logistically, however, it may be a challenge to provide the time and equipment necessary to offer decision aids. In addition, the patient decision cannot be viewed in isolation. Physicians need to be trained to have the communication skills needed to facilitate this aided patient decision making process.

V. RECOMMENDATIONS FOR PATIENT DECISION AIDS

Given these challenges in developing and implementing patient decision aids, I offer several practical recommendations to make aids as useful as possible:

- 1) Maintain the flexible conceptual model that incorporates patient preferences, clinical judgment, and scientific evidence into the shared decision making process and weighs each differently depending upon the individual patient and clinical situation.

2) Create educational materials matched to the literacy and numeracy levels of the users as well as their sociocultural context.

3) Allow the patient to navigate the tool so that he or she can acquire the information most important to him or herself. Patients have diverse needs and learning styles: Some patients might want comprehensive information such as every possible side effect of a medication, while others may prefer a simpler approach learning only about the most frequent or severe side effects.

4) Highlight individualized risk-stratified data to patients. Data derived from individuals who are similar to the patient are more applicable than general population data.⁵²

5) Incorporate both quantitative and qualitative data. Patients learn complementary information from numbers and stories. For example, quantitative population outcome data in conjunction with testimonials from patients who have made different medication choices provide a more complete picture. Some patients tend to think in a reductionist manner, disaggregating the individual components of a decision, while others take a more holistic approach trying to get an overall feel for the issue.

6) Make patients aware of both objective and subjective criteria. Some patients might choose the medication or treatment approach that minimizes the risk of mortality, while others will factor in the type of risk. For example, some patients fear cancer more than cardiovascular disease, and might choose a medication that has low cancer risk in exchange for a proportionately higher risk of cardiovascular mortality.

7) Update data regularly so that the aids remain current.

8) Design decision aids for older persons. Medication choice issues are particularly common in older persons, and thus architectural issues pertaining to limited vision, orthopedics (e.g., use of computer mouse), and cognitive status are vital.

9) Train providers in communication and behavior change, ethical issues of patient autonomy versus paternalism, and equity issues in resource allocation. The issues surrounding choice of medications are often complex. Therefore, patient decision aids should not be viewed within a vacuum. An informed, guided discussion between provider and patient is critical.

10) Make the cost ramifications explicit. Different perspectives are possible, including the patient's out-of-pocket costs, costs to the health system, and costs to society. The appropriate perspective to take depends upon what the policymaker's goal is, whether minimizing individual burden or maximizing societal cost-effectiveness. Even though both patients and physicians believe that

52. DAVID L. SACKETT ET AL., *CLINICAL EPIDEMIOLOGY: A BASIC SCIENCE FOR CLINICAL MEDICINE* (2d ed. 1991).

discussions about out-of-pocket costs are important, costs have frequently not been explicitly incorporated into the individual patient's decision-making process.⁵³

11) Encourage more thought and research on determining ideal outcomes for patient decision aids.

12) Fund development of decision aids from multiple interested parties including pharmaceutical companies, insurance companies, and the government.

The challenges raised in my recommendations are significant but solvable with resources and will.

CONCLUSION

We have entered an exciting new era in which patient empowerment and shared decision making are important components. We need to preserve free speech and the flow of information while, at the same time, protecting consumers. In addition, while patient empowerment sounds attractive as a general concept, we need to think creatively about how to facilitate a process in which patients become educated about their choices in a comprehensible way and make choices that reflect their true values and wishes. The challenge is creating incentives and regulations that will ensure a fair process to inform patients and facilitate shared decision making, leading to optimal, cost-effective outcomes.⁵⁴ Appropriately regulated and enforced direct-to-consumer advertising could lead to more informed, empowered patients. Decision aids can help patients define their own values and preferences, and engage in better discussions and make wiser decisions with their physicians.

53. G. Caleb Alexander et al., *Patient-Physician Communication About Out-of-Pocket Costs*, 290 JAMA 953, 955-57 (2003).

54. For a more complete discussion of cost-effectiveness analysis, see MARTHE GOLD ET AL., *COST-EFFECTIVENESS IN HEALTH AND MEDICINE* (1996).

The Effects and Role of Direct-to-Physician Marketing in the Pharmaceutical Industry: An Integrative Review

Puneet Manchanda, M.Phil., Ph.D.* and Elisabeth Honka†

INTRODUCTION

The pharmaceutical industry plays a vital role in the world's economy, as well as in ensuring the welfare of its citizens. In the United States, this industry constitutes a large and important part of the economy. In 2002, health care expenditure in the United States reached \$1.6 trillion, accounting for fifteen percent of total GNP.¹ This percentage is also growing over time—it was seven percent in 1970.² An important component of the health care industry is the pharmaceutical industry—in 2002, its size was estimated at \$193 billion.³ While the pharmaceutical industry is driven by innovation, it spends more money on marketing than on research and development.⁴ For example, this industry spends more than any other U.S. industry on its sales force (\$7 billion annually) and on media advertising (\$2.8 billion annually).⁵

Pharmaceutical companies typically direct their marketing efforts toward physicians and, as of late, directly to patients (consumers). The marketing efforts directed at physicians comprise personal selling through sales representatives

* Associate Professor of Marketing, Graduate School of Business, University of Chicago. He would like to acknowledge research support from the Kilts Center for Marketing at the Graduate School of Business, University of Chicago.

† Doctoral Student, Graduate School of Business, University of Chicago. She would like to acknowledge financial support from the German Academic Exchange Council (DAAD).

1. *What's Driving Health Care Costs and the Uninsured: Hearing Before the Senate Comm. on Health, Educ., Labor & Pensions*, 108th Cong. 38 (2004) (statement of Douglas Holtz-Eakin, Director, Cong. Budget Office).

2. *Id.*

3. PHARM. RESEARCH & MFRS. OF AM. (PHRMA), PHARMACEUTICAL INDUSTRY PROFILE 2004 44 (2004), <http://www.phrma.org/publications/publications//2004-03-31.937.pdf>.

4. FAMILIES USA FOUND., OFF THE CHARTS: PAY, PROFITS AND SPENDING BY DRUG COMPANIES 3 (2001), <http://www.familiesusa.org/site/DocServer/offthecharts.pdf?docID=823>.

5. DICK R. WITTINK, ANALYSIS OF ROI FOR PHARMACEUTICAL PROMOTION (ARPP) (2002), http://www.rxpromoroi.org/arpp/media/arpp_handout_0927.pdf.

(detailing);⁶ sampling (provision of drugs at no cost); physician meetings and events; and advertisements in medical journals.⁷ Since 1997, a change in the legal environment that allowed direct-to-consumer advertising (DTCA) has resulted in a 350% increase in expenditures for such advertising between 1996 and 2001.⁸ However, the biggest chunk of marketing expenditure is directed toward detailing.⁹ Historically, detailing has been the pharmaceutical industry's primary promotional instrument.¹⁰ Our aim in this Article is to provide an integrative review of the academic research on the effect and role of detailing. We highlight the main findings that arise from the medical, legal, economics, and marketing literature. Finally, we propose an explanation of the pervasiveness of detailing over a drug's life. We conclude by proposing how an increase in the efficiency and effectiveness of this expenditure can benefit firms, physicians, and patients.

As noted above, we attempt to provide an integrative review of the literature on detailing. As a result, we need to provide organizational criteria in order to deal with the large number of studies on the subject. We use two such criteria to organize this review: the outcome variable and the nature of the data collected by the researcher. The outcome variable is the variable that is affected by detailing, which can range from "softer" variables, such as physician attitudes, to "harder" variables, such as drug sales. The nature of data collected can be survey data or actual behavioral (market) data. While we believe that these two criteria are important, we also describe the extant literature using all relevant criteria in the form of tables in the Appendix.¹¹ We first examine physician attitudes toward

6. For an excellent overview of the evolution of modern detailing in the United States, see Jeremy E. Greene, *Attention To 'Details': Etiquette and the Pharmaceutical Salesman in Postwar America*, 34 SOC. STUD. SCI. 271 (2004).

7. STEPHEN P. BRADLEY & JAMES WEBER, *THE PHARMACEUTICAL INDUSTRY: CHALLENGES IN THE NEW CENTURY* 7 (Harvard Bus. Sch., Working Paper No. 9-703-489, 2004).

8. *Id.*

9. WITTINK, *supra* note 5, at 6-7.

10. BRADLEY & WEBER, *supra* note 7, at 8-9.

11. There have been other such integrative articles. See, e.g., Dale B. Christensen & Patricia J. Bush, *Drug Prescribing: Patterns, Problems and Proposals*, 15a SOC. SCI. & MED. 343 (1981); Richard J. Plumridge, *A Review of Factors Influencing Drug Prescribing* (pt. 1), 13 AUSTL. J. HOSP. PHARMACY 16 (1983). But not all include detailing as an independent variable, see, e.g., Dennis W. Raisch, *A Model of Methods for Influencing Prescribing* (pts. 1 & 2), 24 DICP, ANNALS PHARMACOTHERAPY 417, 537 (1990), even the ones that do not differentiate between detailing as a general source of information, detailing's function in new product introductions, and its influence on physician prescribing, see, e.g., James R. Williams & Paul J. Hensel, *Changes in Physicians' Sources of Pharmaceutical Information: A Review and Analysis*, 11 J. HEALTH CARE MARKETING 46 (1991). Most other literature reviews cover a very broad set of variables that affect physician prescribing. See, e.g., T.S. Caudill & Nicole Lurie, *The Influence of Pharmaceutical Industry Advertising on Physician Prescribing*, 22 J. DRUG ISSUES 331 (1992); Elina Hemminki, *Review of*

detailing using studies from the medical literature. As the purported reason for the existence of detailing is that it provides information to physicians, we then examine whether the medical community indeed perceives it as such and if these perceptions have changed over time. We then look at whether detailing affects stated and actual prescription behavior. Finally, we examine the role of detailing over the life cycle of a drug with a special emphasis on its effects in the early, awareness-building stage. We conclude by integrating the main findings into a coherent explanation of the role of detailing.

Based on our analysis we draw the following major conclusions. First, it seems that physicians have negative (at one extreme) to neutral (at the other) attitudes toward pharmaceutical sales representatives. The variance in this attitude is explained by a variety of factors. Some of the important factors are the quality of informational and educational support provided via detailing, detailer style, and the physician's practicing environment. However, detailing exists and flourishes in spite of this attitude as it provides an inexpensive and convenient source of information. Interestingly, the importance of detailing as a source of information has declined over the past five decades, as it is no longer the most important source of information.

Second, not only is detailing an important source of information, it affects physician prescription behavior in a positive and significant manner. More important, this seems to occur over the length of the drug's life cycle. This is puzzling considering that over a drug's life cycle, most information about the drug is likely to be disseminated early on—a fact confirmed by physician surveys. Thus, detailing's effect should diminish over the life cycle of a drug. There is no obvious explanation for the fact that detailing has a positive and significant effect late in the drug life cycle. Based on our analysis and industry observations, our explanation is that in addition to providing a "reminder effect," constant interaction builds a stock of goodwill between a detailer (or the firm) and the physician, translating into positive physician prescription behavior. This goodwill is not based on purely objective and rational factors but on social and cultural norms. Its character changes from informative to more persuasive in the

Literature on the Factors Affecting Drug Prescribing, 9 SOC. SCI. & MED. 111 (1975); Russell R. Miller, *Prescribing Habits of Physicians: A Review of Studies on Prescribing of Drugs* (pts. 1-8), 7 DRUG INTELLIGENCE & CLINICAL PHARMACY 492, 557 (1973), 8 DRUG INTELLIGENCE & CLINICAL PHARMACY 81 (1974); J.P. Rovers, *The Doctor's, the Druggist's, and the Detail Rep's Dance: Who Leads, Who Follows*, 37 CAN. FAM. PHYSICIAN 100 (1991); Dennis B. Worthen, *Prescribing Influences: An Overview*, 7 BRIT. J. MED. EDUC. 109 (1973). In other words, reviews concentrating on detailing as a factor influencing physician attitudes and prescribing behavior are relatively rare. Also noteworthy is Joel Lexchin, *Doctors and Detailers: Therapeutic Education or Pharmaceutical Promotion?*, 19 INT'L J. HEALTH SERVS. 663 (1989), which critically discusses doctors, detailers, and their relationships.

later stages of the drug life cycle. The evolution of goodwill in this manner reflects the deepening relationship between the physician and the pharmaceutical sales representative.

Finally, detailing is clearly here to stay. Although physicians claim to tolerate it as a necessary evil, detailing evidently has an impact on prescription behavior via both a subjective and an objective path. From the industry perspective, pharmaceutical firms continue to invest heavily in this mode of promotion—they have more than doubled their 1997 sales force to about 90,000 in 2002.¹² Thus, one possible approach that could be beneficial to all concerned parties—patients, physicians, firms, and policy makers—would be to ensure that this large expenditure on detailing is carried out in the most efficient manner possible. We conclude the Article by providing suggestions on how this could be carried out.

I. REVIEW OF PAST STUDIES

A. Physician Attitudes Toward Detailing

In this Section, we focus our attention on physician attitudes as documented (mostly) in the medical literature. We focus on general attitudes toward detailing and detailers and attitudes toward gifts. We then look at studies that provide an explanation for the formation of these attitudes. (Tables 1a-1c provide a more detailed overview of the studies discussed.)

1. Physician Attitudes Toward Detailers

A series of studies document that physician attitudes toward detailing and pharmaceutical sales representatives are mostly negative. First, Poirier et al. surveyed physicians on their attitudes toward pharmaceutical marketing practices.¹³ They found that only 24% of the physicians were satisfied with detailing and 48% were dissatisfied.¹⁴ These skeptical attitudes were confirmed by the finding that only 20% of the physicians believed in the accuracy and objectivity of presented information, while 44% did not.¹⁵ Nevertheless, 56% admitted that representatives could influence formulary decisions if efficacy,

12. *Pushing Pills*, THE ECONOMIST, Feb. 15, 2003, at 61.

13. Therese I. Poirier et al., *Pharmacists' and Physicians' Attitudes Toward Pharmaceutical Marketing Practices*, 51 AM. J. HOSP. PHARMACY 378 (1994).

14. *Id.* at 379.

15. *Id.*

toxicity, and cost were the same, while 28% disagreed with this statement.¹⁶ Strang et al. surveyed Canadian general practitioners and specialists on their attitudes toward sales representatives.¹⁷ Ninety-two percent of the physicians thought that drug promotion was a major goal of sales representatives, while only 37% saw physician education as a major goal of sales efforts.¹⁸ Forty-seven percent of the physicians thought that sales representatives provide all information to describe a drug, while 80% thought that detailers overemphasized the effectiveness of a drug.¹⁹

In 1996 Caudill et al. surveyed physicians about their attitudes toward the educational value and behavioral influence of pharmaceutical sales representatives.²⁰ Physicians agreed that sales representatives provided useful and accurate information about newly and already established drugs, but only slightly agreed that they performed an important teaching function.²¹ Physicians strongly agreed that sales representatives should be banned from making presentations where the physicians practice.²² McKinney et al. examined physicians' attitudes toward detailing and its potential for ethical compromise.²³ They found that physicians had somewhat negative attitudes toward the educational and informational value of detailing activities, but also acknowledged sales representatives' support for conferences and speakers.²⁴

Hopper et al. collected information on the effects of an educational intervention aimed at training physicians in interactions with sales representatives.²⁵ They surveyed residents and faculty before and after the intervention. Before the intervention, physicians slightly agreed that contact with detailers was not beneficial, but strongly disagreed that it might influence their

16. *Id.*

17. David Strang et al., *National Survey on the Attitudes of Canadian Physicians Toward Drug-Detailing by Pharmaceutical Representatives*, 29 ANNALS ROYAL C. PHYSICIANS & SURGEONS CAN. 474 (1996).

18. *Id.* at 476.

19. *Id.*

20. T.S. Caudill et al., *Physicians, Pharmaceutical Sales Representatives, and the Cost of Prescribing*, 5 ARCHIVES FAM. MED. 201 (1996).

21. *Id.* at 204.

22. *Id.*

23. W. Paul McKinney et al., *Attitudes of Internal Medicine Faculty and Residents Toward Professional Interaction with Pharmaceutical Sales Representatives*, 264 JAMA 1693 (1990).

24. *Id.* at 1695.

25. John A. Hopper et al., *Effects of an Educational Intervention on Residents' Knowledge and Attitudes Toward Interactions with Pharmaceutical Representatives*, 12 J. GEN. INTERNAL MED. 639 (1997).

prescribing in negative ways.²⁶ However, physicians were rather neutral about whether interactions were likely to influence the prescribing behavior of other physicians in negative ways.²⁷ Residents believed significantly more than faculty that sales representatives sometimes use unethical marketing practices and that the residents have too much contact with the detailers.²⁸ Two items of the post-intervention survey were found to have statistically significant differences between the intervention and nonintervention resident groups: Participating residents more strongly believed than nonintervention residents that sales representatives may use unethical marketing practices and that interaction with detailers is likely to influence the prescribing of other physicians in negative ways.²⁹

Other studies have documented more neutral physician attitudes to detailing and pharmaceutical sales representatives. Andaleeb and Tallman's examination of physicians' relationships with sales representatives showed that although physicians viewed sales representatives as an important source of information, they thought they could also get the needed information from another source.³⁰ The study found that physicians had friendly relationships with sales representatives and did not distrust them, but did not consider them a vital part of their practice. Selling methods were not viewed as manipulative, nor were sales representatives perceived negatively.³¹ The median overall attitude toward sales representatives was also reported as neutral in a study by Thomson et al. based on a survey of general practitioners in New Zealand.³² One specific attribute of this study was that only 77% of the physicians reported having access to colleagues.³³ Physicians also tended to see more sales representatives if colleagues' advice was less readily available. Eighty-seven percent of the respondents reported having seen detailers; one physician would have liked to see sales representatives, but was never visited because of the isolated location of his practice.³⁴ The reasons given most often for seeing sales representatives were

26. *Id.* at 640.

27. *Id.*

28. *Id.*

29. *Id.* at 641.

30. Syed S. Andaleeb & Robert F. Tallman, *Relationships of Physicians with Pharmaceutical Sales Representatives and Pharmaceutical Companies: An Exploratory Study*, 13 HEALTH MARKETING Q. 79, 84-85 (1996).

31. *Id.*

32. A.N. Thomson et al., *Attitudes of General Practitioners in New Zealand to Pharmaceutical Representatives*, 44 BRIT. J. GEN. PRAC. 220 (1994).

33. *Id.* at 221.

34. *Id.* at 221.

practical prescribing information, samples, a feeling of politeness, or pressure.³⁵ Relative to all respondents, practitioners favorably disposed to detailers saw more sales representatives. Also relative to physicians in smaller practices, physicians in larger practices saw fewer detailers.

2. Physician Attitudes Toward Gifts

Another dimension on which physicians have very strong attitudes is the practice of gift-giving from pharmaceutical sales representatives to physicians. As part of the detailing process, sales representatives often not only give samples, but also give trinkets, books, or meals. Sixty-seven percent of the faculty and 77% of the residents in the McKinney et al. study indicated that they believed that physicians could be compromised by accepting gifts from sales representatives.³⁶ Specifically, the authors found that 50% of the faculty and 42% of the residents perceived gifts of \$100 or more to be likely to compromise a physician's judgment.³⁷ Keim et al. questioned residents and directors in emergency medicine about their interactions with the biomedical industry³⁸ and found that 74% of the residents who responded to the survey believed that representatives "sometimes cross ethical boundaries by giving gifts to physicians."³⁹ While 75% of the program directors believed that marketing techniques of sales representatives affected residents' prescribing, only 49% of the residents believed the same to be true.⁴⁰ However, in a 1997 study Madhavan et al. found that doctors slightly agreed that pharmaceutical companies gave gifts to physicians to influence their prescribing, but disagreed that, in general, gift-giving influenced most physicians' prescribing behavior.⁴¹ The physicians surveyed strongly disagreed that they themselves could be influenced in their prescribing behavior by the gifts they receive.⁴² Aldir et al. also reported that physicians disagreed that their prescribing was influenced by gifts such as lunches or dinners, but the physicians surveyed admitted that their prescribing

35. *Id.*

36. McKinney et al., *supra* note 23.

37. *Id.* at 1695.

38. Samuel M. Keim et al., *Beliefs and Practices of Emergency Medicine Faculty and Residents Regarding Professional Interactions with the Biomedical Industry*, 22 ANNALS EMERGENCY MED. 1576 (1993).

39. *Id.* at 1578.

40. *Id.*

41. S. Madhavan et al., *The Gift Relationship Between Pharmaceutical Companies and Physicians: An Exploratory Survey of Physicians*, 22 J. CLINICAL PHARMACY & THERAPEUTICS 207, 212 (1997).

42. *Id.*

might be affected by sample giving.⁴³ Reeder et al. surveyed chief residents in emergency medicine programs about their attitudes surrounding their “gift relationship” with pharmaceutical companies.⁴⁴ One-fifth of the chief residents believed that accepting gifts could affect their own prescription habits.⁴⁵

While the studies above suggest that gifts are not generally acceptable, the ones that asked about the value of the gift found that gifts below a certain threshold—typically \$100—are acceptable.⁴⁶ Aldir et al. also found that the majority of physicians agreed that gifts above \$100 were inappropriate, but found no relationship between physicians’ values regarding gifts and their attitudes regarding scientific information provided by the pharmaceutical industry.⁴⁷

3. Antecedents of Physician Attitudes

While the studies described above have expressed attitudes, there is relatively little research on the antecedents (or causes) of this attitude formation. A 1991 study by Lagace et al. showed that the salesperson’s ethical behavior and expertise positively affected physician attitudes (especially trust and satisfaction).⁴⁸ It also found that the frequency of visits did not significantly affect satisfaction.⁴⁹ Brotzman and Mark provided an alternative set of antecedents;⁵⁰ they argued that regulatory policies affect physicians’ attitudes toward sales representatives.⁵¹ By comparing residents from free and restricted programs,⁵² Brotzman and Mark found those from free programs to be twice as likely to view overall interactions, educational information, and extracurricular

43. Rodolfo E. Aldir et al., *Practicing and Resident Physicians’ Views on Pharmaceutical Companies*, 16 J. CONTINUING EDUC. HEALTH PROFS. 25, 31 (1996).

44. Mike Reeder et al., *Pharmaceutical Representatives and Emergency Medicine Residents: A National Survey*, 22 ANNALS EMERGENCY MED. 1593 (1993).

45. *Id.* at 1595.

46. See Aldir et al., *supra* note 43; McKinney et al., *supra* note 23; Reeder et al., *supra* note 44.

47. Aldir et al., *supra* note 43, at 29.

48. Rosemary Lagace et al., *The Relevance of Ethical Salesperson Behavior on Relationship Quality: The Pharmaceutical Industry*, 11 J. PERS. SELLING & SALES MGMT. 39, 44 (1991).

49. *Id.*

50. Gregory L. Brotzman & David H. Mark, *The Effect on Resident Attitudes of Regulatory Policies Regarding Pharmaceutical Representative Activities*, 8 J. GEN. INTERNAL MED. 130 (1993).

51. *Id.* at 132.

52. In a free program, residents’ access to sales representatives is not overseen by the facility. However, in a restricted program, the quality and quantity of contact between residents and sales representatives is determined by the policies of the facility. This restriction usually results in much less access relative to that in a free program.

activities as beneficial, and four times more likely to view detailing as helpful.⁵³ Physicians from free programs had more contacts with sales representatives and, as measured by eight categories, they were more likely to feel that gift acceptance was appropriate.⁵⁴ However, in contrast, Ferguson et al. found no differences in the likelihood of meeting with sales representatives or accepting samples between internists from hospitals with and without regulatory policies.⁵⁵ Andaleeb and Tallman also identified factors that influenced physicians' attitudes toward sales representatives.⁵⁶ They found that physicians' attitudes were influenced by the information and educational support they receive, selling techniques, and their volume of patients.⁵⁷ The more informational and educational support from sales representatives and the higher the number of patients, the more favorable were physicians' attitudes toward sales representatives.⁵⁸ In contrast, a manipulative and aggressive selling style was associated with an unfavorable attitude.⁵⁹

B. Detailing as a Source of Information

The classic role of detailing is to provide (medical) information to a physician. This information ranges from awareness-building to detailed technical information. The importance of detailing as one of physicians' sources of information about drugs has often been investigated, as is outlined in Table 2. These studies were perceptual by nature and asked physicians how much importance they attributed to either detailing in general or its certain aspects.

In general, physicians perceive detailers to be useful sources of information. Fassold and Gowdey surveyed Canadian physicians, about one-half general practitioners and one-half specialists, on their reactions to drug promotions.⁶⁰ Forty-six percent of the respondents considered detailing the most informative and/or acceptable form of drug promotion.⁶¹ Among the general practitioners,

53. Brotzman & Mark, *supra* note 50, at 132.

54. *Id.* at 132.

55. Robert P. Ferguson et al., *Encounters with Pharmaceutical Sales Representatives Among Practicing Internists*, 107 AM. J. MED. 149 (1999).

56. Syed S. Andaleeb & Robert F. Tallman, *Physician Attitudes Toward Pharmaceutical Sales Representatives*, 20 HEALTH CARE MGMT. REV. 68 (1995).

57. *Id.* at 73.

58. *Id.*

59. *Id.*

60. R.W. Fassold & C.W. Gowdey, *A Survey of Physicians' Reactions to Drug Promotion*, 98 CAN. MED. ASS'N J. 701 (1968).

61. *Id.* at 702.

56% ranked it first while only 37% of the specialists did so.⁶² Only 13% considered detailing as the least informative and/or acceptable form of drug promotion.⁶³ Twenty-four percent of the physicians (18% specialists, 31% general practitioners) stated that detailing and other spoken forms of manufacturers' advertisements were their preferred choice of information on new drugs.⁶⁴ Another study by Henley et al. surveyed Iowa physicians on the frequency with which they use certain sources of drug information.⁶⁵ Pharmaceutical textbooks were ranked first, followed by drug salesmen.⁶⁶ Fifty-five percent of the physicians indicated that they relied on pharmaceutical representatives very often or often.⁶⁷ Twenty-seven percent indicated occasional use of this information source, and 17% seldom or never rely on detailers.⁶⁸ A 1976 study by Eaton and Parish surveyed general practitioners in Great Britain concerning how they gathered information and what sources they found useful.⁶⁹ Ninety-three percent of the respondents indicated seeing sales representatives at least once a week, and 67% thought they would lose an important source of information if they did not see any detailers.⁷⁰ While 90% of the physicians indicated that sales representatives were a helpful source to find out about the existence of a drug, only 51% said they were a helpful source in finding out about the usefulness of a drug.⁷¹ Reeder et al. found that 80% of the respondents thought their residency program benefited from interaction with pharmaceutical representatives, usually through the presentation of new clinical data.⁷² Finally, Connelly et al. studied knowledge resources of family physicians and found that they regarded detailers to provide information that was less extensive and credible than secondary (e.g., *Physicians' Desk Reference*, medical texts, *Index Medicus*) and primary sources (colleagues).⁷³ In terms of information availability, searchability, understandability, and applicability, information from detailers was regarded as higher than information from secondary sources such as

62. *Id.*

63. *Id.*

64. *Id.* at 703.

65. Scott Henley et al., *Dissemination of Drug Information*, 42 HOSPITALS 99 (1968).

66. *Id.* at 100.

67. *Id.*

68. *Id.*

69. Gail Eaton & Peter Parish, *Sources of Drug Information Used by General Practitioners: Prescribing in General Practice*, 26 J. ROYAL C. GEN. PRAC. 58 (Supp. 1976).

70. *Id.* at 61.

71. *Id.* at 62-63.

72. Reeder et al., *supra* note 44, at 1595.

73. Donald P. Connelly et al., *Knowledge Resource Preferences of Family Physicians*, 30 J. FAM. PRAC. 353 (1990).

research articles, *Index Medicus*, and a computerized bibliography.⁷⁴

The underlying assumption in the above studies is that physicians are good at extracting relevant information from detailers. However, as this is usually not part of medical school training, Shaughnessy et al. investigated whether physicians would benefit from such training.⁷⁵ They developed a curriculum to teach hospital faculty and residents to evaluate information provided by pharmaceutical representatives.⁷⁶ After receiving this training, physicians had generally positive attitudes toward the detailers' services and did not feel overly influenced by them relative to pre-training. This effect, while statistically significant, was small in magnitude.⁷⁷ Samourai and Avorn summarize a series of studies that also show that education of physicians about detailing leads to more accurate and cost-effective prescription outcomes.⁷⁸

In contrast, some studies have found detailers lacking in this regard. Williams et al. found that a minority (19%) of Canadian physicians viewed detailers to be an important source of information (though a quarter of high prescribing physicians found them to be an important source).⁷⁹ Caudill et al. also asked physicians to rate sales representatives as a source of information on the three dimensions of credibility, availability, and applicability.⁸⁰ The mean responses were all nearly neutral, and there was a significant positive correlation between the three measures.⁸¹ Fassold and Gowdey's 1968 study asked physicians to grade sales representatives on several characteristics.⁸² While detailers were rated good or excellent with respect to personality, reliability, and honesty by 86%, 65%, and 69% of the physicians respectively, sales representatives' general knowledge, knowledge of drugs, and usefulness was rated fair or poor by 67%, 63%, and 59% of the practitioners, respectively.⁸³

A more interesting question is the importance of detailing as an information source relative to other information sources. A study by Kalb tried to assess the

74. *Id.* at 356 fig. 1.

75. Allen F. Shaughnessy et al., *Teaching Information Mastery: Evaluating Information Provided by Pharmaceutical Representatives*, 27 FAM. MED. 581 (1995).

76. *Id.*

77. *Id.* at 584.

78. Stephen B. Soumerai & Jerry Avorn, *Principles of Educational Outreach ('Academic Detailing') To Improve Clinical Decision Making*, 263 JAMA 549 (1990).

79. A. Paul Williams et al., *The Physician as Prescriber: Relations Between Knowledge About Prescription Drugs, Encounters with Patients and the Pharmaceutical Industry, and Prescription Volume*, 3 HEALTH & CAN. SOC'Y 135, 164 (1995).

80. Caudill et al., *supra* note 20, at 203.

81. *Id.*

82. Fassold & Gowdey, *supra* note 60.

83. *Id.* at 704.

relative importance of six information sources for physician prescribing.⁸⁴ When directly asked whether sales representatives were the primary motivation in their prescribing habits, only 13% of the physicians felt this way.⁸⁵ When asked to rank the six information sources they relied on for making prescribing decisions, physicians rated sales representatives as fourth on average, whereby the score was not significantly different from the third source, company reputation.⁸⁶ Gambrill and Bridges-Webb surveyed general practitioners on their most recent, regular, and most useful sources of information about therapeutics and prescribing.⁸⁷ Journals were ranked first on all three criteria, followed by sales representatives.⁸⁸ Strickland-Hodge and Jeqson surveyed general practitioners in Great Britain about their usage of information sources.⁸⁹ The sales representative was ranked seventh on a general evaluation as a source of information, but fourth on its general usefulness among twenty sources.⁹⁰ Hatton et al. studied physicians' sources of information about teratogenic effects of drugs (drug use during pregnancy).⁹¹ They asked physicians to indicate their general drug information sources and sources used for specific information about potential teratogenicity of drugs. In both cases, sales representatives were ranked fifth, but the mean use rate was only about one-half in the second case.⁹² Bower and Burkett conducted a survey in 1987 to learn about factors influencing prescribing of generic drugs.⁹³ Thirty-two percent of the physicians indicated that they rely a great deal on sales representatives as a source of information and 61% of the physicians reported relying to some extent.⁹⁴ In Eaton and Parish's study, physicians ranked articles and partners ahead of detailing.⁹⁵

84. Clifford C. Kalb, *Psychological Motivations in Physician Prescribing Habits*, 13 MED. MARKETING & MEDIA 43 (1978).

85. *Id.* at 49.

86. *Id.* at 52.

87. J. Gambrill & C. Bridges-Webb, *Use of Sources of Therapeutic and Prescribing Information by General Practitioners*, 9 AUSTL. FAM. PHYSICIAN 482 (1980).

88. *Id.* at 483.

89. B. Strickland-Hodge & M.H. Jeqson, *Usage of Information Sources by General Practitioners*, 73 J. ROYAL SOC'Y MED. 857 (1980).

90. *Id.* at 859.

91. Randy C. Hatton et al., *Physicians' Sources of Information About Teratogenic Effects of Drugs*, 16 DRUG INFO. J. 148 (1982).

92. *Id.* at 150.

93. Anthony D. Bower & Gary L. Burkett, *Family Physicians and Generic Drugs: A Study of Recognition, Information Sources, Prescribing Attitudes, and Practices*, 24 J. FAM. PRAC. 612 (1987).

94. *Id.* at 613.

95. Eaton & Parish, *supra* note 69, at 63.

Given the rich availability of information sources to physicians over the last two or three decades, it is possible that detailing, while important (as the studies above have documented), may be losing out to other sources over time. In 1991 Williams and Hensel reviewed twenty empirical analyses between 1952 and 1986 and conducted a meta-analysis of these studies about drug information sources, their importance, or use by physicians.⁹⁶ They classified all possible sources of information into four categories. These categories were commercial sources (direct mail, journal advertising, and detailing), noncommercial sources (journal articles, meetings, conventions, pharmacists, and colleagues), personal sources which require a face-to-face contact (detailing, colleagues, pharmacists, and conventions/meetings/conferences), and nonpersonal sources (journal articles, journal advertising, and direct mail). They found that commercial sources declined in importance over time and personal sources gained in importance, while the difference for nonpersonal sources was insignificant.⁹⁷ The importance of detailing specifically has declined over time. While it was mostly ranked first in studies in the 1950s, results from the 1970s or later (there were no studies between 1959-1970) ranked it the fourth to seventh most important source of information.⁹⁸ The new most important sources were colleagues and journal articles; pharmacists and other sources also gained more weight.⁹⁹ The observed declining ranking of detailing is congruent with lower reported means of detailing in studies where physicians had to rate the importance of sales representatives on a scale.¹⁰⁰

C. Physicians' Responsiveness Toward Detailing

Building on the previous discussion, the important question for physicians, pharmaceutical firms, and policymakers is whether detailing indeed influences prescription behavior (or sales). We begin by focusing on physicians' perceptions about this question (which we describe in greater detail in Table 3). We then look at studies that have examined this issue using behavioral (market) data.

1. Studies Using Perceptual Data

In one of the earliest studies of physicians' responsiveness, Caplow and Raymond found that detailing was a minimal factor in motivating physicians to

96. Williams & Hensel, *supra* note 11.

97. *Id.* at 55.

98. *Id.* at 54-55.

99. *Id.* at 54.

100. *Id.* at 55.

prescribe a drug.¹⁰¹ This is consistent with a 2000 study by Abratt and Lanteigne.¹⁰²

However, this message was somewhat less clear in other studies. For example, Pitt and Nel found physicians perceived sales calls as the third most dominant factor after personal experience with the product and recommendations from colleagues.¹⁰³ This information implied that physicians regarded detailing as more influential than seminars, conferences, ads in journals, samples, or direct mail. Lurie et al. surveyed internal medicine faculty and housestaff at teaching hospitals about the nature, frequency, and effects of their contacts with sales representatives.¹⁰⁴ Both faculty and housestaff averaged 1.5 brief conversations per month with sales representatives.¹⁰⁵ Twenty-five percent of faculty and 32% of residents reported having changed their practices at least once in the preceding year based on contact with a detailer.¹⁰⁶ But detailing activity also potentially influences prescribing through another channel: hospital formularies. Based on the suggestion of a sales representative, 20% of faculty and 4% of residents had recommended an addition to the formularies at least once during the past year.¹⁰⁷ Using stepwise logistic regression, Lurie et al. found that brief conversations, extended conversations, and free meals predicted a change in faculty prescribing practice.¹⁰⁸ Taylor and Bond studied the association between new prescriptions and factors of influence.¹⁰⁹ They collected prescription behavior of 189 British practitioners and asked them to indicate up to two influences. Pharmaceutical representatives were listed as the second most important source (20% of total number of times mentioned) and mostly influenced the prescription of anti-infective preparations and non-steroidal anti-inflammatory agents.¹¹⁰ Swanson et al. found that twenty-seven out of thirty-one family physicians felt that detailers

101. See Theodore Caplow & John J. Raymond, *Factors Influencing the Selection of Pharmaceutical Product*, 19 J. MARKETING 18, 20 (1954).

102. Russell Abratt & Julie Lanteigne, *Factors Influencing General Practitioners in the Prescription of Homeopathic Medicines*, 31 AFR. J. BUS. MGMT. 91, 94 (2000).

103. Leyland Pitt & Deon Nel, *Pharmaceutical Promotion Tools—Their Relative Importance*, 22 EUR. J. MARKETING 7, 10 (1988).

104. N. Lurie et al., *Pharmaceutical Representatives in Academic Medical Centers: Interaction with Faculty and Housestaff*, 5 J. GEN. INTERNAL MED. 240 (1990).

105. *Id.* at 241.

106. *Id.* at 242.

107. *Id.*

108. *Id.*

109. Ross J. Taylor & Christine M. Bond, *Change in the Established Prescribing Habits of General Practitioners: An Analysis of Initial Prescriptions in General Practice*, 41 BRIT. J. GEN. PRAC. 244 (1991).

110. *Id.* at 246.

affected their prescription behavior.¹¹¹ However, the physicians felt that this influence was small.¹¹² Strang et al. surveyed 262 practitioners, of whom 70% agreed that detailing affected their prescribing habits.¹¹³ Williams et al. also found a strong positive association between the number of visits by detailers and the number of prescriptions per week.¹¹⁴

Bower and Burkett found that family physicians who relied least on sales representatives were most likely to prescribe generic drugs (33%), while only 12% of those who said they relied "a great deal" on detailers prescribed generic drugs.¹¹⁵ Physicians who relied "some or not at all" on sales representatives as a source of information also recognized more generic and trade name drugs.¹¹⁶ Chren and Landefeld used survey data to test three hypotheses: whether physicians who interacted with drug companies were no more likely than other physicians to (1) make formulary requests; (2) request drugs manufactured by those companies; and (3) request drugs manufactured by those companies than drugs manufactured by other companies.¹¹⁷ They measured interaction with pharmaceutical companies in the following four forms: traditional detailing, acceptance of money to support attendance at educational symposia, acceptance of money to speak at educational symposia, and acceptance of money for research. The results demonstrate a strong, consistent, and specific association between physicians' behavior and many types of interactions with pharmaceutical companies, including detailing.¹¹⁸

From the discussion above, it seems that physicians are beginning to acknowledge that detailing has an impact on physician prescription behavior. However, the general perception that detailing has no effect on prescription behavior still persists. This perception may exist because physicians are unwilling to admit their reliance on detailing or their lack of awareness of such influence.¹¹⁹ Finally, Roughead et al. provided some insights into how and why

111. Rick W. Swanson et al., *Pharmaceutical Representatives—Educators or Product Marketers?*, 69 ACAD. MED. 128, 128 (1994).

112. *Id.*

113. Strang et al., *supra* note 17, at 476.

114. Williams et al., *supra* note 79, at 165.

115. Bower & Burkett, *supra* note 93, at 614.

116. *Id.* at 615.

117. Mary-Margaret Chren & C. Seth Landefeld, *Physicians' Behavior and Their Interactions with Drug Companies: A Controlled Study of Physicians Who Requested Additions to a Hospital Drug Formulary*, 271 JAMA 684 (1994).

118. *Id.* at 687.

119. See Jerry Avorn et al., *Scientific Versus Commercial Sources of Influence on the Prescribing Behavior of Physicians*, 73 AM. J. MED. 4 (1982). Not surprisingly, other studies have also documented contradictory statements made by physicians. For example, Ferguson et al. found

physicians were affected by detailing.¹²⁰ They used sixteen taped visits where sixty-four medicines were detailed. They found that the most common method, which was seen in all sixteen visits, was reciprocation where detailers gave gifts such as samples and printed material to physicians.¹²¹ Such gift-giving made the physicians feel bound to make a repayment and encouraged an automatic response. Social validation claims were used in 41% of the cases.¹²² The peer groups to whose established practices sales representatives referred when using social validation were mostly vaguely defined as "other doctors." Commitment acts appealed to the need and desire to be consistent in order to influence physicians' behavior. These acts were applied in 39% of details either in the form of a direct request to prescribe the product or in a series of questions or statements that gradually moved to agreement to prescribe the drug.¹²³ And last, detailers appealed to authority in the form of experts in 14% of the interactions.¹²⁴

2. Studies Using Market Data

Most of the studies about physicians' responsiveness to detailing have concentrated on either estimating sales response models to detailing (and other advertising tools) or estimating sales response models to the total marketing mix.

a. Detailing Response Models

We first focus on models that focus exclusively on modeling the impact of detailing on demand (dollar sales, market share, or number of prescriptions). Parsons and Vanden Abeele carried out one of the first studies estimating sales response to detailing.¹²⁵ They observed an established drug in the growth phase of a product class with ten products, none of which was dominant. Using time-varying coefficients, they estimated a multiplicative model with pooled data and

that physicians describing themselves as busy practitioners were significantly less likely to abstain from meeting sales representatives and that physicians with frequent contacts were virtually all busy practitioners, even though presumably busier physicians should have less time to meet detailers. See Ferguson et al., *supra* note 55.

120. E.E. Roughead et al., *Commercial Detailing Techniques Used by Pharmaceutical Representatives To Influence Prescribing*, 28 AUSTL. & N.Z. J. MED. 306 (1998).

121. *Id.* at 308.

122. *Id.*

123. *Id.*

124. *Id.*

125. Leonard J. Parsons & Piet Vanden Abeele, *Analysis of Sales Call Effectiveness*, 18 J. MARKETING RES. 107 (1981).

found sales call elasticity to be negative if no samples or handouts were additionally given out.¹²⁶

However, this study seems to be the only one that has not found a strong positive effect of detailing on sales. Cleary studied the impact of detailing on physician antibiotic prescribing at a university hospital.¹²⁷ He evaluated the effectiveness of sales representatives on the average number of new prescriptions, the average number of grams prescribed, and their dollar value.¹²⁸ He found a significant correlation between detailing and the number of new prescriptions, but not with the number of grams or dollar value.¹²⁹ He concluded that the latter two variables were less reliable measures of the impact of detailing. Leeflang et al. proposed a method to measure complex time lag structures and to select the most appropriate model.¹³⁰ They applied their procedure to sales representatives' activities in the pharmaceutical industry and found positive effects on sales.¹³¹ Rizzo also found that detailing stock positively affected sales, while current detailing was insignificant.¹³² Conducting a subgroup analysis for on-patent drugs only, the same pattern was confirmed.¹³³ Wosinska examined the effects of DTCA on the demand for drugs.¹³⁴ She found that detailing had a significant positive brand switching effect, even stronger than the one from DTCA.¹³⁵

Using a hierarchical model, Manchanda and Chintagunta studied physicians' response to detailing at the individual level.¹³⁶ They modeled the number of prescriptions as a function of detailing frequency and quality measured by the

126. *Id.* at 111.

127. John D. Cleary, *Impact of Pharmaceutical Sales Representatives on Physician Antibiotic Prescribing*, 8 J. PHARMACY TECH. 27 (1992).

128. *Id.* at 28.

129. *Id.*

130. Peter S.H. Leeflang et al., *Identification and Estimation of Complex Multivariate Lag Structures: A Nesting Approach*, 24 APPLIED ECON. 273, 281 (1992) (recommending the use of a geometric (multiplicative) lag).

131. *Id.*

132. John A. Rizzo, *Advertising and Competition in the Ethical Pharmaceutical Industry: The Case of Antihypertensive Drugs*, 42 J.L. & ECON. 89, 108 tbl. 3 (1999).

133. *Id.* at 110.

134. MARTA WOSINSKA, JUST WHAT THE PATIENT ORDERED? DIRECT-TO-CONSUMER ADVERTISING AND THE DEMAND FOR PHARMACEUTICAL PRODUCTS (Harvard Bus. Sch., Marketing Research Paper No. 02-04, 2002), <http://ssrn.com/abstract=347005>.

135. *Id.* at 18.

136. Puneet Manchanda & Pradeep K. Chintagunta, *Responsiveness of Physician Prescription Behavior to Salesforce Efforts: An Individual Level Analysis*, 15 MARKETING LETTERS 129 (2004).

number of provided samples.¹³⁷ Their results showed that both measures of detailing and their interaction effect positively affected the number of prescriptions.¹³⁸ They also investigated sales force effectiveness assuming partial knowledge of the response parameters.¹³⁹ Though most physicians responded positively to sales calls, they found that physicians were not detailed optimally. High-volume physicians were detailed to a greater extent than low-volume physicians without regard to their responsiveness to detailing.¹⁴⁰ Iizuka and Jin estimated the effects of DTCA in the prescription drug market.¹⁴¹ While they found that DTCA increases the number of visits to physicians' offices and had a market-expanding effect for a whole class of drugs, they found no significant effect of DTCA on physicians' choice of a specific brand.¹⁴² In contrast, detailing positively influenced doctors' brand choice.¹⁴³ Using a large-scale dataset, Mizik and Jacobson tried to pinpoint the effects of detailing and sampling as precisely as possible. They estimated fixed-effects distributed lag regression models for three different drugs and found that detailing, lagged up to the previous six months, was statistically significant.¹⁴⁴ In other words, past detailing affects current prescription behavior.

Most studies find a positive significant effect of detailing.¹⁴⁵ This effect is robust to differences in variable operationalization, model specification, data series, and estimation method. Table 4 shows that the effect of detailing is positive and significant across a wide variety of models and datasets.

b. Marketing Mix Models

We now focus on marketing mix models. Marketing mix models differ from the models described above as they include the effects of other marketing variables along with detailing in order to provide a more complete picture of sales and prescription behavior. Another advantage of these models is that they can pin down the effects of various instruments simultaneously.

137. *Id.* at 136.

138. *Id.* at 138-39.

139. Puneet Manchanda et al., *Response Modeling with Non-Random Marketing Mix Variables*, 41 J. MARKETING RES. 467 (2004).

140. *Id.* at 474.

141. Toshiaki Iizuka & Ginger Z. Jin, *The Effects of Direct-to-Consumer Advertising in the Prescription Drug Markets* (2003) (unpublished manuscript, on file with authors).

142. *Id.* at 11, 21.

143. *Id.* at 21.

144. Natalie Mizik & Robert Jacobson, *Are Physicians "Easy Marks"? Quantifying the Effects of Detailing and Sampling on New Prescriptions*, 50 MGMT. SCI. 1704, 1734 (2004).

145. Parsons & Vanden Abeele, *supra* note 125, is the one exception.

Berndt et al. investigated the effects of detailing, journal ads, DTCA, and pricing in an industry as well as market-share model.¹⁴⁶ For both models, they found detailing to have the largest positive significant effects among the marketing activities.¹⁴⁷ Gonul et al. measured the impact of price, detailing squared,¹⁴⁸ samples, and several interaction effects with physicians' characteristics on doctors' choice of drugs.¹⁴⁹ They found that detailing increased the prescription probability of a drug, while detailing squared decreased it.¹⁵⁰ The interaction effects between detailing and Medicare price were significant and negative, while detailing's effect with HMO insurance was insignificant.¹⁵¹ Wittink measured the effects of several promotional instruments on return on investment (ROI).¹⁵² He examined how ROI differed according to brand size and launch date and also provided detailed analyses for specific therapeutic categories.¹⁵³ He found that the average revenue impact estimates of detailing remained constant around one dollar for small brands; increased from \$1.20 if the brand was launched before 1994 to \$2.10 if the brand was launched between 1998 and 2000 for medium-sized brands; and from \$3.10 if the brand was launched before 1994 to \$11.60 if the brand was launched between 1998 and 2000 for large brands.¹⁵⁴ Based on these findings, he concluded that the most promising return target for additional resources was detailing for large brands launched after 1997.¹⁵⁵

In a 2004 study, Narayanan et al. examined the effects of detailing, DTCA, other marketing efforts such as meetings and events, price and their interactions with sales, and ROI.¹⁵⁶ They estimated both category sales and sales share models and found that detailing did not affect category sales, but did affect the market share.¹⁵⁷ They found long-term effects of detailing on revenues and

146. Ernst R. Berndt et al., *Information, Marketing, and Pricing in the U.S. Antiulcer Drug Market*, 85 AM. ECON. REV. 100 (1995).

147. *Id.* at 103-04.

148. Detailing squared represents the product of detailing with itself. The role of this term is to capture non-linear (diminishing) returns to detailing.

149. Fusun F. Gonul et al., *Promotion of Prescription Drugs and Its Impact on Physicians' Choice Behavior*, 65 J. MARKETING 79 (2001).

150. *Id.* at 86-87.

151. *Id.* at 87.

152. WITTINK, *supra* note 5.

153. *Id.* at 13-19.

154. *Id.* at 19.

155. *Id.* at 28.

156. Sridhar Narayanan et al., *Return on Investment Implications for Pharmaceutical Promotional Expenditures: The Role of Marketing Mix Interactions*, 68 J. MARKETING 90 (2004).

157. *Id.* at 97, 98.

significant interaction effects between marketing variables in the market share model.¹⁵⁸ Iizuka et al. found an insignificant interaction effect between detailing and DTCA advertising.¹⁵⁹

In general, these models all find that detailing has a positive and significant effect on sales, even after controlling for other marketing mix instruments. Most studies also find that the effect of detailing is largest relative to other marketing instruments. However, the results pertaining to detailing interactions (the joint effect of detailing and another marketing instrument) are not clear. Table 5 provides a detailed overview of these studies.

D. The Role of Detailing over the Product's Life Cycle

The discussion up to this point has shown evidence that while physicians are somewhat negatively predisposed toward detailers and detailing, they do perceive them as a source of information. There is also evidence that detailing has a positive and significant effect on prescription behavior for both physicians' perceptions and market data. An interesting question that arises particularly in pharmaceutical markets is whether the effect of detailing varies over a product's life cycle. When a new drug is launched, not much is known about its efficacy in practice, which may make detailing more effective. Academic researchers have suggested this explanation. For example, Miller notes that detailing is likely to play a large role in the early and awareness-building phase of a new product's life.¹⁶⁰ Consistent with our approach, we first look at studies that examine physician perceptions about the role of detailing over the drug's life cycle and then at behavioral studies.

1. Studies Using Perceptual Data

Most studies in this area have found that detailing plays an important role in how physicians obtain information about newly launched products (see Table 6 for details). McCue et al. surveyed internists, surgeons, and general practitioners to find out their opinions about the accuracy, accessibility, and frequency of use of ten information sources for new drugs.¹⁶¹ While only about 36% of the physicians considered information from sales representatives to be accurate, 72% regarded it as accessible and 45% reported its frequent use.¹⁶² McCue et al. also

158. *Id.* at 99, 100.

159. Iizuka & Jin, *supra* note 141, at 23.

160. Miller, *supra* note 11, at 493.

161. Jack D. McCue et al., *Physicians' Opinions of the Accuracy, Accessibility, and Frequency of Use of Ten Sources of New Drug Information*, 79 S. MED. J. 441 (1986).

162. *Id.* at 442.

found that family practitioners and physicians with more than fifteen years in practice used sales representatives significantly more as a source of information than did internists, surgeons, or less-experienced physicians.¹⁶³ Stross examined the dissemination of information about the management of chronic airway obstruction in small community hospitals.¹⁶⁴ He surveyed internists and family physicians on information sources that were critical to changing their behavior. While sales representatives appeared irrelevant to the diagnosis of the illness, they were important in influencing decisions to use new drugs.¹⁶⁵ Differentiating between early and late adopters, 80% of the former cited sales representatives as their major source of information, while only 15% of the latter did so.¹⁶⁶ Stross explained the great role played by sales representatives in his study by the fact that there were no formal education programs on chronic airway obstruction in these hospitals.¹⁶⁷

Peay and Peay studied the adoption process of a specific new drug, temazepam.¹⁶⁸ Among those physicians who were familiar with this drug (71%), 40% reported to have first heard from detailers about the drug.¹⁶⁹ Thirty-seven percent of the doctors received additional information from detailers after first hearing about the drug and before prescribing it.¹⁷⁰ More than 42% of the physicians identified the detailers as the most influential information source in their first decision to prescribe temazepam.¹⁷¹ Sixty-one percent of the doctors familiar with temazepam reported contact with the detailers regarding the drug.¹⁷² They concluded that contact with detailers was the most consistent predictor of choice and quantity of prescriptions of temazepam.¹⁷³ In a follow-up study, Peay and Peay confirmed their finding for medium-risk drugs but found that among specialists who evaluated relatively high-risk drugs, the importance of detailers was ranked twelfth among fifteen potential sources.¹⁷⁴ Manning and Denson

163. *Id.*

164. Jeoffrey K. Stross, *Information Sources and Clinical Decisions*, 2 J. GEN. INTERNAL MED. 155 (1987).

165. *Id.* at 157.

166. *Id.* at 158.

167. *Id.*

168. Marilyn Y. Peay & Edmund R. Peay, *The Role of Commercial Sources in the Adoption of a New Drug*, 26 SOC. SCI. & MED. 1183 (1988).

169. *Id.* at 1185.

170. *Id.*

171. *Id.*

172. *Id.*

173. *Id.*

174. Marilyn Y. Peay & Edmund R. Peay, *Patterns of Preference for Information Sources in the Adoption of New Drugs by Specialists*, 31 SOC. SCI. & MED. 467, 470 (1990).

surveyed Californian general internists about how they learned about a specific new drug, cimetidine.¹⁷⁵ Fifty-six percent of these physicians named more than one information source.¹⁷⁶ Detailing was ranked sixth among seventeen sources from which practitioners first gained knowledge of the drug and learned about the principles of using it.¹⁷⁷ As a means to update information about cimetidine, detailing was ranked seventh.¹⁷⁸ Colleagues were ranked third on all three criteria.¹⁷⁹

Differentiating between the awareness and evaluation stage of a new drug, physicians ranked sales representatives first on the former and sixth on the latter among twelve sources in Strickland-Hodge and Jeqson's study.¹⁸⁰ Single-practice doctors cited detailers significantly more often for drug evaluation than did joint-practice doctors.¹⁸¹ The authors also found that "industrial information . . . was cited significantly more often by older, single-practice doctors who had a first degree only, did none of their own dispensing, and who did not specialize."¹⁸²

While most physicians note that detailing plays an important role in their understanding and adoption of new products, at least one study finds mixed results. Christensen and Wertheimer studied sources of information and influence on new drug prescribing by surveying pediatric and adult medicine practitioners working in a health maintenance organization.¹⁸³ When asked how they learned about the existence of two specific new drugs, detailing played only a minor role for one of the drugs, while it was most often identified as the first source of information for the second drug.¹⁸⁴ The authors provided three explanations for this result: differences in preferred information sources among physician specialties, differences in promotional practices for the two drugs, and "attributes or activities of the detailers involved."¹⁸⁵ For both new drugs, detailing was unimportant when the physicians were asked about the most important information source influencing their decision to prescribe a drug for the first

175. Phil R. Manning & Teri A. Denson, *How Internists Learned About Cimetidine*, 92 ANNALS INTERNAL MED. 690 (1980).

176. *Id.* at 690.

177. *Id.* at 691.

178. *Id.*

179. *Id.*

180. Strickland-Hodge & Jeqson, *supra* note 89, at 860.

181. *Id.* at 861.

182. *Id.* at 862.

183. Dale B. Christensen & Albert I. Wertheimer, *Sources of Information and Influence on New Drug Prescribing Among Physicians in an HMO*, 13A SOC. SCI. & MED. 313 (1979).

184. *Id.* at 316.

185. *Id.*

time.¹⁸⁶ However, this organization's policy allowed only for minimal contact with detailers. The presence of this policy may explain why detailing was ranked last among eleven as the most frequently used source of information concerning drug therapy.¹⁸⁷

2. Studies Using Market Data

In contrast to the studies above, market data-based studies examine the relationship between the sales performance of a new drug and detailing post-launch. Lilien et al. developed a repeat-purchase diffusion model to forecast and control the rate of sales for a new product using Bayesian estimation.¹⁸⁸ They noted two phenomena: Early prescribing doctors prescribed more, and the effectiveness of detailing decayed over time. Both phenomena were linked to decreasing returns to detailing spending over time.¹⁸⁹ Assuming similar market characteristics for all drugs, they found positive effects of detailing on sales.¹⁹⁰ Berndt et al. studied a diffusion process with consumption externalities.¹⁹¹ They estimated the effects of advertising on market share and simulated it until the market reached its equilibrium shares. They found a significant positive effect of detailing as well as detailing elasticities of about one.¹⁹² Manchanda et al. found that detailing had a significant and positive effect on the decision to adopt a drug even after controlling for the adoption behavior of "near" physicians.¹⁹³

Azoulay investigated "how different sources of information influence the diffusion of pharmaceutical innovations."¹⁹⁴ He found a significant positive effect of detailing on market share.¹⁹⁵ He also found support for the hypothesis that marketing plays an important informative role in increasing demand, but a

186. *Id.* at 317.

187. *Id.* at 315.

188. Gary L. Lilien et al., *Bayesian Estimation and Control of Detailing Effort in a Repeat Purchase Diffusion Environment*, 27 MGMT. SCI. 493 (1981).

189. *Id.* at 495.

190. *Id.* at 502.

191. Ernst R. Berndt et al., *Consumption Externalities and Diffusion in Pharmaceutical Markets: Antiulcer Drugs*, 51 J. INDUS. ECON. 243 (2003).

192. *Id.* at 262.

193. PUNEET MANCHANDA ET AL., *THE ROLE OF TARGETED COMMUNICATION AND CONTAGION IN PRODUCT ADOPTION* (Rutgers Bus. Sch. Marketing Dep't, Working Paper No. RBS-MKT-2004-02, 2004).

194. Pierre Azoulay, *Do Pharmaceutical Sales Respond to Scientific Evidence?*, 11 J. ECON. & MGMT. STRATEGY 551, 551 (2002).

195. *Id.* at 574.

relatively minor persuasive role.¹⁹⁶ Narayanan et al., who investigated the role of detailing over a product's life cycle, confirmed some of these results in their own study.¹⁹⁷ They hypothesized that early in the product's life cycle, detailing would play largely an informative role (i.e., it would reduce uncertainty about a product's efficacy) while later, detailing would play a more persuasive role.¹⁹⁸ They found this situation to be true using data on three new drugs in the antihistamine category.¹⁹⁹ Specifically, they found that the effect of detailing was larger on sales in the early stages when there was both an informative (indirect) and persuasive (direct) effect, as opposed to later stages, when there was only a persuasive effect.²⁰⁰ This result was also found in a subsequent study that examined the effects of detailing in the erectile dysfunction category using individual physician data.²⁰¹ Note that in both the perceptual and the market data-based studies, very little effort has been focused on understanding the exact information transfer during detailing over the life cycle. This area remains open for research.

II. DISCUSSION

At this point, it is worthwhile to try to summarize the main message from these studies. Note that given our broad span of studies and disciplines, it is hard to provide objective (or quantitative) findings. Thus, the following represents our subjective interpretation, based on all the studies discussed up to now, of the role and effects of detailing.

We first began by examining physician attitudes toward detailing and detailers. Broadly speaking, it seems that physicians have negative (at one extreme) to neutral attitudes (at the other) toward pharmaceutical sales representatives. The variance in attitude is explained by a variety of factors. First, the more informational and educational support provided by the representative and the higher the number of patients, the more favorable a physician's attitude toward sales representatives. Second, detailer style and detail content also affect attitude. For example, a manipulative and aggressive selling style is associated with an unfavorable attitude. The overemphasis of drug promotion versus

196. *Id.* at 583.

197. Sridhar Narayanan & Puneet Manchanda, *Temporal Differences in the Role of Marketing Communication in New Product Categories*, 42 J. MARKETING RES. (forthcoming 2005).

198. *Id.* (manuscript at 15).

199. *Id.* (manuscript at 14).

200. *Id.*

201. Sridhar Narayanan, Puneet Manchanda, & Pradeep K. Chintagunta, *Heterogeneous Learning and the Targeting of Marketing Communication for New Products* (Nov. 2004) (unpublished manuscript, on file with authors).

information delivery also tends to engender negative attitudes. Finally, it also seems that the physician's environment helps determine her attitude toward detailers. For example, physicians who have relatively little access to colleagues seem to have a less negative attitude toward detailers. Also, physicians in practices that restrict access to detailers tend to be more negative in their attitudes toward detailing and detailers. Attitudes toward gifting are mostly negative, though several studies note that gifts below a certain threshold are acceptable. A more disturbing finding is that these gifts induce reciprocal feelings among physicians.

Given this somewhat negative picture of the relationship between physicians and detailers, the question is why the practice of detailing persists. The answer seems to lie in the fact that detailing and interaction with detailers acts as an inexpensive and convenient source of information. Studies that have explicitly investigated this question seem to suggest that detailers (and detailing) do provide pertinent information. While physicians are aware of the potential conflicts of interest, they still find this information to be of some value. Two other interesting themes also emerge. First, relative to other sources of information, it is clear that detailing is not the most important source. The most important source of information seems to be either medical journals or other colleagues.²⁰² Second, to the extent that our studies are representative of each decade, the relative importance of detailing as a source of information has declined over the past five decades. More recent studies have found that it occupies a rank between four and seven in contrast to one or two.

However, from the patient, physician, firm, and policymaker's point of view, it is important to establish that detailing does have a significant effect on physician prescription behavior. Interestingly enough, many studies that have asked physicians this question find that physicians believe that it is likely that prescription behavior can be influenced by detailing. This opinion is supported by virtually all the studies that have investigated the effect of detailing (either in isolation or with other marketing instruments) using behavioral data either at the market or the individual physician level. While there seems to be little consensus about the size of the effect, it is clear that the effect is positive and significant in a statistical sense.

This result is somewhat puzzling, especially considering that over a drug's life cycle, most information about the drug is likely to be disseminated early on.²⁰³ This observation implies that if indeed the role of detailing is to provide information, its effect should die out soon after launch. However, we do not see

202. Given that these studies are all based on survey data, it should be noted that this reply represents the "correct" professional response.

203. See discussion *infra* Section I.D.

this result in the studies cited above. We carry this notion further and investigate the role of detailing for new products. As physicians typically need more information about new products, it is clear that detailing should play a larger role at the beginning of a drug's life cycle. The survey studies that have investigated this question seem to confirm that detailing does play an important role, especially in the early, awareness-building, phase of a new product's launch. Presumably, this effect should diminish as a drug enters the maturity phase of its life cycle.

Most of the perceptual studies confirm the importance of detailing in the early stages of the life cycle. These studies also confirm the diminishing role of detailing over the product's life cycle. In other words, these studies find that detailing has a positive, but decreasing, effect over the whole life cycle of a drug. While this finding helps us in confirming our hypothesis, we still need to explain the existence of a positive detailing effect in the late stages of the life cycle. Our explanation is that, in addition to providing a "reminder effect," the constant interaction builds a stock of goodwill between a detailer (or the firm) and the physician. This goodwill is not based on purely objective and rational factors but on social and cultural ones. Its character changes from informative to more persuasive in the later stages of the drug's life cycle. The evolution of goodwill in this manner reflects the deepening of the relationship between the physician and the pharmaceutical sales representative. Reports on the industry focus on using detailing to build lasting relationships with physicians, providing some support for our explanation.²⁰⁴

In conclusion, detailing is clearly here to stay. While physicians claim to tolerate it as a necessary evil, it evidently has an impact on prescription behavior via both a subjective and an objective path. They are therefore heavily invested in this mode of promotion. Thus, one possible approach that could be beneficial to all parties concerned—patients, physicians, firms, and policymakers—would be to ensure that this large expenditure on detailing is carried out in the most efficient manner possible. The application of economics and management science principles to the high-quality marketing data now available shows considerable potential for "optimizing" detailing expenditure. By "optimal," we mean that firms detail to the point where the marginal benefit is equal to marginal cost.

204. *Pushing Pills*, *supra* note 12; Martin E. Elling et al., *Making More of Pharma's Sales Force*, MCKINSEY Q., 2002 Issue 3, at 86. Note that our explanation of goodwill accumulation is based on three arguments. First, this goodwill accumulation represents the residual effect of detailing *after* the informational effects have died out. Thus, these effects do not have anything to do with objective information transfer. Second, this industry is based on building lasting relationships between physicians and manufacturers. Finally, we are unable to offer an alternative explanation that is consistent with the results.

From the physician's perspective, this means that detailing should be carried out at a level that provides physicians with the amount of information (and samples) that enables them to maximize the welfare of their patients. To this end, it may be useful to provide physicians training on how to use their relationship with detailers in the most effective manner possible. Similarly, firms could also investigate other, complementary, mechanisms that could improve the efficiency and effectiveness of their detailing practices. Thus, initiatives such as e-detailing are worth investigating. The benefit of more efficient use of detailing expenditure for consumers is somewhat indirect, as it arises when firms divert the savings to developing newer products. Finally, policymakers could suggest training and educational standards for detailers such that detailers act more as collaborative problem-solvers rather than as sales professionals.

CONCLUSION

This paper attempts to synthesize research on the role and effect of detailing in the pharmaceutical industry. Our sweep is broad in the sense that we have looked at papers across various disciplines spanning five decades of research. In terms of what this research has documented, it is clear that there is a two-sided relationship between physicians and detailers. There is also strong evidence that detailing affects physician (prescription) behavior in a positive and significant manner. While this relationship is tolerated by physicians and promoted aggressively by detailers, it is clear that it will continue in the foreseeable future. Based on our reading of the research, we propose a relatively simple explanation of why this relationship exists and matters in terms of prescription outcomes. The objective part of the relationship consists of awareness-building and information transfer and is prevalent in the early part of a drug's life cycle. The subjective part pertains to building social and personal relationships between physicians and detailers. It is therefore important that physicians, firms, and policymakers recognize this reality and take appropriate steps so as to make this relationship as efficient and effective as possible.

APPENDIX

TABLE 1A: PHYSICIAN ATTITUDES TOWARD DETAILERS

| Paper | Type of Design | Type of Analysis | Other Variables Measured? | Hospital or Practice? | Physician Specialty | Number of Physicians | Results | Scale Used | Further Results/Comments |
|--------------------------|----------------|------------------|---------------------------|--|---|-----------------------------|---|------------------------------|---|
| Pointer et al. (1994) | Cross-section | Descriptive | No | - | - | 26 | 24% (48%) (dissatisfied with detailing; 20% (44%) believe (not) in accuracy & objectivity of information | - | 56% (28%) say representatives could (not) influence formulary decisions if efficacy, toxicity, & cost are the same |
| Strang et al. (1996) | Cross-section | Descriptive | No | - | General practitioners, specialists | 262 | 92% (37%) consider drugs promotion (education) major goal of detailing; 47% (80%) think detailers provide all information (over-emphasize effectiveness) | - | - |
| Caudill et al. (1996) | Cross-section | Descriptive | No | 46% solo practice, 42.6% group practice, 12.3% academic/hospital | Family medicine, general internal medicine | 446 | 3.59 (3.51) useful & accurate information about newly introduced (already established) drugs | 1-5 | 3.05 important teaching function; 4.21 banned from presentation where they practice |
| McKinney et al. (1990) | Cross-section | Descriptive | No | Hospital | - | 277 faculty & 240 residents | Important teaching function: residents 3.6, faculty 3.9; provide useful & accurate information about newly introduced (already established) drugs: residents 3.2, faculty 3.2 (residents 3.1; faculty 3.3); no impact on prescribing behavior: residents 2.5, faculty 2.3 | Yes/no questions & 1-5 scale | Same degree of contact whether gifts distributed or not: residents 3.2, faculty 2.0; acceptance of promotional items with no impact on prescribing behavior: residents 1.7, faculty 1.6 |
| Hopper et al. (1997) | Cross-section | Descriptive | No | Residents, faculty | Primary care internal medicine, internal medicine, pediatrics | 31 residents & 18 faculty | Contact not beneficial: 2.31; influence my (others') prescribing in negative ways: 4.27 (2.95) | 1-5 | Comparison of pre- & post-intervention: stronger belief in unethical practices & that one's prescribing is influenced in neg. ways |
| Andaleeb, Tallman (1996) | Cross-section | Descriptive | No | Hospital | 35 osteopathic doctors, 58 medical doctors, 2 doctors of pediatric medicine | 95 | Important source of information: 4.00; Could get information from another source: 4.35; friendly relationship: 4.55; trust: 4.00; representatives are manipulative: 3.50 | 1-6 | Total of 17 questions concerning attitude |
| Thomson et al. (1994) | Cross-section | Descriptive | No | Practice | General practitioners | 67 | Overall attitude toward detailers: 3 (median) | 1-5 | Provide a list of characteristics of detailers practitioners regard as pos./neg. |

TABLE 1B: PHYSICIAN ATTITUDES TOWARD GIFTS

| Paper | Type of Dataset | Type of Analysis | Other Variables Measured? | Hospital or Practice? | Physician Specialty | Number of Physicians | Results | Scale for Calculations | Further Results/Comments |
|----------------------|-----------------|----------------------------------|---------------------------|--|--|--|---|------------------------|---|
| Keim et al. (1993) | Cross-section | Descriptive | No | Hospital | Residents & directors in emergency medicine | 1,385 residents & 80 directors in emergency medicine | 75% believe accepting gifts is potentially ethically compromising | - | 75% directors believe detailing can affect residents' prescribing; only 49% of residents believe this |
| Madhavan (1997) | Cross-section | Principal component, correlation | No | 37.5% solo practice, 24.6% group practice, 16.7% hospital, 21.2% teaching hospital | Family medicine, internal medicine, obstetrics/gynecology, pediatrics, surgery, others | 283 | Gifts to influence prescribing: 3.17; physicians are (1 am) influenced in prescribing by gifts: 2.29 (1.22) | 0-6 | - |
| Aldir et al. (1996) | Cross-section | Descriptive | No | Residents, practice | Internal medicine, family medicine, OB/GYN | 521 | Influenced by lunches (dinners): 2.3 (2.4) practitioners/2.7 (2.8) residents | 10-point rating scale | 4.6% practitioners (6.0% residents) believe gifts may influence prescribing (yes/no question) |
| Reeder et al. (1993) | Cross-section | Descriptive | No | Hospital | Chief residents | 87 | 20% accepting gifts can influence own prescribing | - | - |

TABLE 1C: ANTECEDENTS OF PHYSICIAN ATTITUDES

| Paper | Type of Dataset | Type of Analysis | Other Variables Measured? | Which Other Variables Measured? | Hospital or Practice? | Physician Specialty | Number of Physicians | Results | Scale for Calculations |
|--------------------------|-----------------|-----------------------------|---------------------------|---|-----------------------|---|----------------------|--|------------------------|
| Lagace et al. (1991) | Cross-section | Regression | No | - | - | - | 90 | Trust (satisfaction) on ethical behavior: .209 (.212); expertise: .554 (.450); no. of meetings: 270 (.098); relationship duration: .064 (-.031); years in practice: .128 (.093) | - |
| Brotzman, Mark (1993) | Cross-section | Descriptive | Yes | <i>Physicians' Desk Reference</i> , AMA drug evaluations, medical letter, journal ads | Residents | Family medicine residents | 265 | Residents from free programs twice as likely to view overall interactions, educational information, & extracurricular activities as beneficial, four times more likely to view detailing as helpful, more likely to view gift acceptance as appropriate. | 1-3 |
| Ferguson et al. (1999) | Cross-section | Descriptive | No | - | Hospital | Internists | 346 | 85% (80%) physicians in restricted (free) programs had seen detailers; 71% (71%) physicians in restricted (free) programs accepted samples | - |
| Andaleeb, Tallman (1995) | Cross-section | Factor analysis, regression | No | - | Hospital | 35 osteopathic doctors, 58 medical doctors, 2 doctors of pediatric medicine | 95 | Regression: support: .384; style: -.190; peers: .128; own practice: .012; length of practice: .004; volume of patients: .005 | 1-6 |

TABLE 2. SOURCE OF GENERAL INFORMATION

| Paper | Type of Dataset | Type of Analysis | Other Variables Measured? | Which Other Variables Measured? | Hospital or Practice? | Physician Specialty | Number of Physicians | Results | Scale for Calculations | Ranking Among All Considered Variables | Comments/Further Results |
|---------------------------|-----------------|------------------|---------------------------|--|--------------------------------|--|----------------------|---|------------------------|---|---|
| Fassold, Cowdey (1968) | Cross-section | Descriptive | Yes | Direct mail, meetings, journals | - | General practitioners, specialists | 531 | 46% (13%) physicians considered detailing as the most (least) informative &/or acceptable form of drug promotion; 56% (37%) general practitioners (specialists) considered detailing as the most informative &/or acceptable form of drug promotion | - | - | Manufacturers spoken advertising as preferred source of information for new drugs for 24% physicians (31% general practitioners, 18% specialists); assessments of detailers on 6 attributes |
| Henley et al. (1968) | Cross-section | Descriptive | Yes | Pharmaceutical & medical textbooks, pharmaceutical & medical periodicals, pharmaceutical literature, other physicians, pharmacists, others | - | 47.5% general practitioners, 52.5% diverse specialists | 300 | 55.1% use detailers very often or often, 26.8% occasionally, 17.4% seldom or never | - | 2/9 | - |
| Eaton, Parish (1976) | Cross-section | Descriptive | Yes | Industry/professional sources, DHSS | - | General practitioners | 453 | 89.6% (50.8%) use detailers to find out about existence (usefulness) of a drug | - | 1/18 for drug existence, 8/18 for drug usefulness | 93% see detailers at least once a week, 67% felt would lose important source of information if not seeing detailers |
| Reeder et al. (1993) | Cross-section | Descriptive | No | - | Hospital | Chief residents | 87 | 80% consider interaction with detailers because of presentation of new clinical data | - | - | - |
| Connelly et al. (1990) | Cross-section | Regression | Yes | Numerous variables | 7 in hospital, 113 in practice | Family physicians | 126 | Detailers rated low on extensiveness, credibility, clinical availability, searchability, understandability, & clinical applicability | 5-point scale | - | - |
| Shaughnessy et al. (1995) | Cross-section | Descriptive | No | - | Residents | - | 12 | No impact on prescribing behavior: 3.1 (pre-test), 3.3 (post-test); gifts without influence: 1.8 (pre-test), 2.3 (post-test) | 1-5 | - | Total of 10 questions concerning attitude (pre- & post-test) |

TABLE 2 CONTINUED: SOURCE OF GENERAL INFORMATION

| Paper | Type of Dataset | Type of Analysis | Other Variables Measured? | Which Other Variables Measured? | Hospital or Practice? | Physician Specialty | Number of Physicians | Results | Scale for Credibility Judgments | Ranking Among All Considered Variables | Comments/Further Results |
|---------------------------------|-----------------|------------------|---------------------------|--|--|--|----------------------|---|---------------------------------|---|--|
| Williams et al (1995) | Cross-section | Descriptive | Yes | Meetings & conventions, articles, CME, pharmacists & pharmacologists, seminars | - | - | 852 | 18.7% (25.1% all (heavy prescribing) physicians view detailing as an important source of information | 1-5 | 6/6 | Pos. correlation (.37) between prescription volume & detailing importance |
| Caudill et al (1996) | Cross-section | Descriptive | No | - | 46% solo practice, 42.6% group practice, 12.3% academic/hospital | Family medicine, general practitioners, internal medicine | 446 | 2.76 on credibility; 3.14 on availability; 3.17 on applicability of information provided by detailers | 1-5 | - | Pos. correlation between the three measures |
| Kalib (1978) | Cross-section | Descriptive | Yes | Journal articles, colleagues, company reputation, journal advertising, direct mail | Practice | 65% general practitioners, 35% diverse specialists | 204 | Asked separately 13% responded detailers as primary source of information | - | 4/6 | 57% view most important function of detailing as providing information about new drugs & dosages |
| Gambrell, Bridges-Webb (1980) | Cross-section | Descriptive | Yes | Journals, consultants & colleagues, drug company literature, clinical meetings | - | General practitioners | 104 | Most recent source: 22%, regular sources: 56%, most useful source: 17% | - | Most recent source: 2/5, regular source: 2/5, most useful source: 2/5 | - |
| Strickland-Hodge, Jenson (1980) | Cross-section | Descriptive | Yes | Several variables (categorized as industry/professional, active/passive) | - | General practitioners | 252 | - | - | General evaluation: 7/20, general usefulness: 4/20 | - |
| Haron et al (1982) | Cross-section | Descriptive | Yes | Various variables | 84.3% private, 9% academic, 6.7% other | 39.2% OB/GYN, 29.5% family practice, 17.5% pediatrics, 13.8% other | 166 | General information: 7.3%, specific information: 3.5% | - | General information: 5/10, specific information: 5/10 | - |
| Bower, Burkett (1987) | Cross-section | Descriptive | Yes | Articles, advertising, colleagues | - | Family physicians | 317 | 31.9% rely a great deal, 60.5% to some extent, & 7.6% not at all on detailers as source of information on new drugs | - | - | - |

TABLE 3: PERCEPTUAL RESPONSE TO DETAILING

| Paper | Type of Data/Analysis | Type of Analysis | Other Marketing Variables Measured? | Which Other Marketing Variables Measured? | Detailing Operation-ization | Dep. Variable Operation-ization | Number of Physicians Practiced? | Specialty? | Results | Scale for Calculations | Ranking Among All Considered Variables | Further Results/Comments |
|------------------------|-----------------------|----------------------------------|-------------------------------------|--|-----------------------------|----------------------------------|---------------------------------|-----------------------|---|------------------------|---|--|
| Caplow, Raymond (1954) | Cross-section | Descriptive | Yes | Direct mail, journal ads & articles, conferences, colleagues | Physicians' opinion | Physicians' opinion | 182 | Practice | Drugs adopted usually using more than one information source; relative importance of sources in inducing initial & continued drug usage different | - | 1st mentioned: 31.1%; 2nd: 14.5%; 3rd: 11.9%; confidence in dealers as motivating drug prescription factor for .08% | - |
| Abatt, Lantagne (2000) | Cross-section | Descriptive, rankings | Yes | Numerous variables | Physicians' opinion | Physicians' opinion | 278 | Practice | General practitioners | 0-4 | Marketing factors: 5/8; professional & marketing factors: 11/13 | Homeopathic medicine |
| Pitt, Nel (1988) | Cross-section | Averages | Yes | Colleagues, conferences, ads, samples, direct mail | Physicians' opinion | Physicians' opinion | 210 | General practitioners | 1.28 (mean) | - | 3/7 | No significant differences between sex & length of time in practice |
| Lurie et al. (1990) | Cross-section | Descriptive, logistic regression | No | Self-reported type & freq. of contacts with dealers | Physicians' opinion | Self-reported change in behavior | 240 faculty, 131 residents | Internal medicine | Among faculty change in practice on brief conversations .016, extended conversations .036, & free meals .089; among faculty request for formulary addition on brief conversations .014, honorarium 178, & research support 1.02; among residents change in practice on brief conversations .049 | - | - | - |
| Taylor, Bond (1991) | Cross-section | Descriptive | Yes | Several professional information sources, mailings/ads | Physicians' opinion | Physicians' prescriptions | 189 | General practitioners | Major influence on change in prescribing habit for 20% | - | Professional & commercial factors: 2/12 | Drugs most influenced by detailing: anti-infective preps. & non-steroidal anti-inflammatory agents |
| Swanson et al. (1994) | Cross-section | Descriptive | No | - | Physicians' opinions | Physicians' opinions | 31 | Family physicians | 87% feel detailers influence their prescription behavior; mean influence 3.50 | 10-point scale | - | 100% of detailers feel they influence behavior of physicians; mean influence 6.01 |

TABLE 3 CONTINUED: PERCEPTUAL RESPONSE TO DETAILING

| Paper | Type of Dataset | Type of Analysis | Other Marketing Variables Measured? | Which Other Marketing Variables Measured? | Detailing Operationalization | Dep. Variable Operationalization | Number of Physicians | Hospital or Practice? | Specialty? | Results | Scale for Calculations | Ranking Among All Considered Variables | Further Results/Comments |
|-------------------------|-----------------|---------------------------|-------------------------------------|--|------------------------------|--|----------------------|-----------------------|-------------------------------------|---|------------------------|--|---|
| Strang et al. (1996) | Cross-section | Descriptive | No | - | Given by survey | Proportion of physicians agreeing with a statement | 262 | - | General practitioners, specialists | Detailing affects prescribing: 2.32; 70% (strongly) agreed with this statement | 1-5 | - | Various questions concerning attitude toward drug-detailing process |
| Williams et al. (1995) | Cross-section | Descriptive | Yes | Meetings & conventions, articles, CME, pharmacists & pharmacologists, seminars | No. of visits | No. of prescriptions written | 852 | - | - | Correlation: .37 | - | - | - |
| Bower, Burkett (1987) | Cross-section | Descriptive | Yes | Articles, advertising, colleagues | Physicians' opinion | Physicians' | 317 | - | Family physicians | Heavily (some or not at all) relying on detailers doctors prescribe mostly generic drugs: 11.8% (33.2%); heavily (some or not at all) relying on detailers doctors have confidence in generic drugs: 56.9% (64.3%) | - | - | Generic drug prescription |
| Chen, Landefeld (1994) | Cross-section | Nested case-control study | No | - | Physicians' opinion | Request that a drug is added to hospital formulary | 105 | Hospital | - | No difference in freq. of general interaction, but case physicians shared more expensive meals; significantly more likely to request formulary addition after meeting with specific detailer whether accepted money to attend symposia, speak at symposia, or perform research or not | - | - | 81% met with detailers |
| Avorn et al. (1982) | Cross-section | Descriptive | Yes | Drug ads, patients' preference, scient. papers, colleagues, own training | Physicians' opinion | Physicians' | 85 | Practice | Internal medicine, general medicine | Detailers perceived as minimal/moderate/very important source of influence by 34%/26%/20% | - | - | Measuring archival/perceptual bias |
| Roughhead et al. (1998) | Cross-section | Descriptive | No | - | Visit's tape record | - | 115 | - | General practitioners | Reciprocal 100% of visits, social validation 41% of visits, commitment/consistency 39% of visits, authority 14% of visits | - | - | - |

TABLE 4: DETAILING RESPONSE MODELS

| Paper | Type of Dataset | Level | Whose Heterogeneity | Other Marketing Variables Measured? | Which Other Marketing Variables Measured? | Detailing Operationalization | Dep. Variable Operationalization | Data Point Freq. | Model | Interaction Effects Estimated? | Specification | Estimation Method | Detailing Coefficient | Sig. Effect at $p = ?$ |
|--------------------------------------|-----------------|------------|---------------------|-------------------------------------|--|---|---|------------------|-------------------------------------|--------------------------------|---|---------------------------------------|---|------------------------|
| Parsons et al. (1981) | Panel | Aggregate | None | Yes | Sampling, direct mail, journal ads | No. of product calls reported by salespeople | Wholesale unit sales | Month | Sales | Yes | Double-log | OLS | -.148 | .001 |
| Cleary (1992) | Panel | Aggregate | - | Yes | Journal articles & ads | No. of new prescriptions/grams prescribed | - | Month | - | No | Descriptive | - | - | - |
| Leeftang et al. (1992) | Time series | Aggregate | None | Yes | Expenditure on magazine advertising, direct mail | Estimated expenditure based on time spent with practitioner | Market share | Month | Market share (geometric lag model) | No | Multinomial logit | - | .081 | .01 |
| Rizzo (1999) | Panel | Individual | Drugs' | Yes | Price | Expenditure | Annual dollar sales (nominal)/price of an initial dose of the drug (nominal) | Year | Category model | Yes | Log-linear (alternative model specifications) | OLS | .160 & .170 | Insig. |
| Rizzo (1999) | Panel | Individual | Drugs' | Yes | Price | Cumulative expenditure | Annual dollar sales (nominal)/price of an initial dose of the drug (nominal) | Year | Category model | Yes | Log-linear (alternative model specifications) | OLS | Between .260 & .280 | .01 |
| Wosnska (2002) | Panel | Individual | Drugs' | Yes | DTCA | Expenditure | No. of new prescriptions | Month | Market share | Yes | Mixed logit | - | .185 & .152 | .01 |
| Marchanda, Chintagunta (2004) | Panel | Individual | Physicians' | Yes | Samples | No. of visits | Average no. of prescriptions | Quarter | No. of Prescriptions | Yes | Poisson | Bayesian estimation | .880 | - |
| Marchanda, Rossi, Chintagunta (2004) | Panel | Individual | Physicians' | No | - | No. of visits | No. of prescriptions | Month | Negative binomial | No | - | Bayesian estimation | .190 in the short-run | - |
| Marchanda, Rossi, Chintagunta (2004) | Panel | Individual | Physicians' | No | - | No. of visits | No. of prescriptions | Month | Negative binomial | No | - | Bayesian estimation | -6.100 in the long-run | - |
| Iturza, Jin (2003) | Panel | Individual | Drugs' | Yes | DTCA, journal, samples | Estimated expenditure based on time spent with practitioner | Market share | Month | Market share | Yes | Multinomial logit (log specification) | - | -.300 | .01 |
| Iturza, Jin (2003) | Panel | Individual | Drugs' | Yes | DTCA, journal, samples | Estimated expenditure based on time spent with practitioner | Cumulative detailing expenditures estimated based on time spent with practitioner | Market share | Month | - | Yes | Multinomial logit (log specification) | - | - |
| Mizlik, Jacobson (2003) | Panel | Individual | Physicians' | Yes | Samples | No. of visits | No. of new prescriptions | Month | Fixed effects distributed lag model | No | IV | OLS | Last 4-6 months sig. cumulative coefficient .129, .151 & .165 | .001 |

TABLE 5: MARKETING MIX MODELS

| Paper | Type of Dataset | Whose Heterogeneity? | Other Marketing Variables Measured? | Which Other Marketing Variables Measured? | Detailing Operationalization | Dep. Variable Operationalization | Data Point Freq. | Model | Interaction effects estimated? | Specification | Estimation | Detailing Coefficient | Sig. Effect at p=? |
|-------------------------|-----------------|----------------------|-------------------------------------|---|---|--|--|-------|--------------------------------|--------------------------------|------------|--|--------------------|
| Berndt et al. (1995) | Panel | Aggregate | None | Yes | DTCA, journal advertising, price | Relative cumulative detailing incumbent minutes to (allowing for spillovers) | No. of patient days of therapy | Month | Categorical model | Double-log (IV) | NL-2SLS | .553 | .001 |
| Berndt et al. (1995) | Panel | Aggregate | None | Yes | DTCA, journal advertising, price | Relative cumulative detailing incumbent minutes to (allowing for spillovers) | No. of patient days of therapy | Month | Market share | Double-log (IV) | NL-2SLS | .649 | .001 |
| Gonul et al. (2001) | Panel | Aggregate | - | Yes | Samples, price | Discounted cumulative time | No. of prescriptions | - | Market share | Multinomial logit (linearized) | ML | .1085 | .001 |
| Witink (2002) | Panel | Aggregate | - | Yes | DTCA, journal advertising, meeting & events | Detailing expenditure | Estimated increase in revenue for \$1 increase in independent variable | Month | Linear Regression | - | OLS | ~.1 (small brands), 1.2 (medium brands), 3.1 (large brands), all increasing with more recent launch date | - |
| Narayanan et al. (2004) | Panel | Individual | Brands | Yes | DTCA, price/expenditure | Detailing | No. of prescriptions | Month | Categorical sales | Linear | IV | -.0093 | Insig. |
| Narayanan et al. (2004) | Panel | Individual | Brands | Yes | DTCA, price/expenditure | Detailing | No. of prescriptions | Month | Market share | Mixed logit | IV | .2853 | .05 |

TABLE 6: PERCEPTUAL RESPONSE OVER PRODUCT LIFE CYCLE

| Page | Type of Data | Type of Analysis | Other Variables Measured? | Which Other Variables Measured? | Hospital or Practice? | Physician Specialty | Number of Physicians | Results | Ranking Among All Considered Variables | Further Results/Comments |
|----------------------------------|---------------|----------------------------|---------------------------|---|-----------------------|---|----------------------|--|---|--|
| McCue et al. (1986) | Cross-section | Descriptive | Yes | Numerous variables | - | Internists, surgeons, & general practitioners | 119 | Accurate: 36%, accessible: 71.8%; accurate & accessible: 27.7%; frequently used: 45.5% Detailers as most valuable information source on new treatment: 3%-35%; detailing as major information source for 80% (15% of early (late) adopters) | Accurate: 10/10; accessible: 5/10; accurate & accessible: 9/10; frequently used: 4/10 | - |
| Stross (1987) | Cross-section | Descriptive | Yes | Formal/informal CME, journals, others | Hospital | Internists, family physicians | 85 | 40% (68.2% of those aware of temazepam cited detailers (commercial sources) as source of first news; 37.1% (87.1%) received additional information from detailers (commercial sources); 42.6% (59.3% of those prescribing viewed detailers (commercial sources) as most influential in first decision to prescribe | - | Detailers without influence on diagnostic tests |
| Pay, Pay (1988) | Cross-section | Descriptive, multi-variate | Yes | Various professional & commercial factors | Practice | General practitioners, specialists | 124 | Target drug adoption: detailing source of first/most useful/most important information source for 5.0%/34.3%/2.3% | - | 71% aware of temazepam, 47.6% had prescribed it; doctors with contact to drug more likely to rate it as moderate advance, more likely to prescribe it/prescribe it earlier, & routinely in preference of alternatives |
| Pay, Pay (1990) | Cross-section | Descriptive | Yes | Various professional & commercial factors | Hospital & community | Specialists | 156 | Mean usefulness rating: 12/15 | Mean usefulness rating: 12/15 | Usual adoption procedure: detailing source of first/most useful/most important information source for 21.9%/14.7%/5.8% |
| Manning, Denson (1980) | Cross-section | Descriptive | Yes | Numerous variables | - | General internists | 449 | First knowledge: 4%; principles of drug usage: 7%; information update: 5% | First knowledge: 6/17; principles of drug usage: 6/17; information update: 7/17 | Medical journals & CME as primary indicated sources of learning about cimetidine |
| Strickland-Hodge, Jorgson (1980) | Cross-section | Descriptive | Yes | Numerous variables | - | General practitioners | 252 | Drug A: first/second/third source of information: 4/6/2; Drug B: first/second/third source of information: 1/1/4; most important source in decision to prescribe Drug A/B 11/5 | Awareness: 1/12; evaluation: 6/12 | Single-practice doctors cite detailers more often than joint-practice doctors; industrial information cited more often by older, non-specialized, single-practice doctors with a first degree only, who did none of their dispensing |
| Christensen, Wertheimer (1979) | Cross-section | Descriptive | Yes | Numerous variables | HMO | Pediatrics, adult medicine practitioners | 29 | - | Most frequently used general information source: 11/11 | - |

TABLE 7: MARKET BEHAVIORAL RESPONSE OVER PRODUCT LIFE CYCLE

| Paper | Type of Dataset | Level | Whose Heterogeneity | Other Marketing Variables Measured? | Which Other Marketing Variables Measured? | Detailing Operationalization | Dep. Variable Operationalization | Data Point | Model | Specification | Estimation | Detailing Coefficient | Sig. Effect at |
|--|-----------------|------------|---------------------|-------------------------------------|---|------------------------------|----------------------------------|------------|---------------------------------|--|---------------------|-----------------------|----------------|
| | | | | | Word-of-mouth among physicians, competitive detailing (assumed equal) | - | No. of prescribing doctors | Quarter | Repeat-purchase diffusion model | 2-stage estimation | Bayesian estimation | - | - |
| Lilien et al. (1981) | Panel | Aggregate | None | Yes | | | | | | Linear & semi-log (saturation level) | | | - |
| Berndt et al. (2000) | Panel | Aggregate | None | Yes | Price | Cumulative detailing minutes | Real price per day of therapy | Month | Category sales | | GMM | - | .001 |
| Berndt et al. (2000) | Panel | Aggregate | None | Yes | Price | Cumulative detailing minutes | Market share | Month | Market share | Multinomial logit (equilibrium shares), IV | SUR, 3SLS, GMM | Elasticities ~1.00 | .001 |
| Manchanda, Xie, Youn (2004) | Panel | Individual | Physicians' | Yes | Samples, contagion | No. of calls | Adoption | Month | Choice | Multinomial logit | Bayesian estimation | Positive | .01 |
| Azoulay (2002) | Panel | Individual | Drugs' | Yes | Price, journal advertising | Cumulative detailing minutes | Total sales | Month | Market share | Multinomial logit | OLS, 2SLS, GMM | Between .582 & 1.081 | .001 |
| Narayanan, Manchanda, Chintagunta (2004) | Panel | Aggregate | Physicians' | Yes | Samples, price, DTCA, expenditure on meetings & events | Detailing expenditure | No. of new prescriptions | Month | Market share | Multinomial logit | Bayesian estimation | .0116 | .01 |
| Narayanan, Manchanda (2004) | Panel | Individual | Physicians' | Yes | Patient requests | No. of calls | Brand choice | Month | Choice | Multinomial logit | Bayesian estimation | Positive | .01 |