



2-1-2010

Seven Pillars of a New Evidentiary Paradigm: The Food, Drug, and Cosmetic Act Enters the Genomic Era

Barbara J. Evans

Follow this and additional works at: <http://scholarship.law.nd.edu/ndlr>

Recommended Citation

Barbara J. Evans, *Seven Pillars of a New Evidentiary Paradigm: The Food, Drug, and Cosmetic Act Enters the Genomic Era*, 85 Notre Dame L. Rev. 419 (2010).

Available at: <http://scholarship.law.nd.edu/ndlr/vol85/iss2/1>

This Article is brought to you for free and open access by NDLScholarship. It has been accepted for inclusion in Notre Dame Law Review by an authorized administrator of NDLScholarship. For more information, please contact lawdr@nd.edu.

ARTICLES

SEVEN PILLARS OF A NEW EVIDENTIARY PARADIGM: THE FOOD, DRUG, AND COSMETIC ACT ENTERS THE GENOMIC ERA

*Barbara J. Evans**

To assess the impact of the March 2009 decision in Wyeth v. Levine, it is crucial to understand that the Supreme Court ruled on actions that the U.S. Food and Drug Administration (FDA) took under a statutory scheme that already had been amended by the time the case was decided. The Food and Drug Administration Amendments Act of 2007 (FDAAA) transformed drug regulation, adding significant new powers to develop evidence and make new types of decisions in the postmarket period. This article explores how the contours of drug regulation are likely to change after FDAAA, which is the most profound reworking of the U.S. drug regulatory framework in half a century.

FDAAA envisions heavy use, during the period after drugs are approved, of evidence from large observational studies that rely on interoperable health data networks. Understanding what was wrong with FDA's old evidentiary paradigm, which dates back to 1962, is essential to understanding its new one. Parts II and III of this article discuss the evidentiary limitations of premarket drug trials; important aspects of modern legal doctrine rest on misconceptions about their evidentiary power. Part IV then explores how scientific advances flowing from the Human Genome Project over the past decade further undermined FDA's old evidentiary paradigm. FDAAA was

© 2010 Barbara J. Evans. Individuals and nonprofit institutions may reproduce and distribute copies of this Article in any format, at or below cost, for educational purposes, so long as each copy identifies the author, provides a citation to the *Notre Dame Law Review*, and includes this provision and copyright notice.

* Associate Professor; Co-director, Health Law & Policy Institute; Director, Center on Biotechnology & Law, University of Houston Law Center, bjevans@central.uh.edu. J.D., Yale Law School; M.S., Ph.D. Stanford University; Post-doctoral Fellow, The University of Texas M.D. Anderson Cancer Center. This research has been supported by funding from the University of Houston Law Foundation.

Congress's response to these problems. Part V identifies seven pillars of the new evidentiary paradigm: seven novel propositions that reject foundational assumptions of twentieth-century drug regulation. Collapse of these assumptions sets off ripple effects in various doctrinal areas. Part VI provides two examples, with the aim of opening a scholarly debate about these and other impacts of FDA's new evidentiary paradigm.

INTRODUCTION	420
I. THE 1962 EVIDENTIARY PARADIGM FOR DRUG REGULATION ..	427
A. <i>The Gap Between Perception and Performance</i>	428
B. <i>The Role of Randomized, Controlled Trials</i>	431
C. <i>Evidentiary Paths Not Taken in 1962</i>	435
II. LIMITATIONS OF THE OLD EVIDENTIARY PARADIGM	439
A. <i>The Myth of the Premarket Safety Trial</i>	439
B. <i>Methodological Limitations of Premarket Safety Studies</i>	444
III. THE OLD PARADIGM: REPAIR OR REPLACE?	450
IV. THE CHALLENGES OF GENOMIC MEDICINE	458
A. <i>The Challenge of Predictive and Preventive Medical Technologies</i>	459
B. <i>The Challenge of Heterogeneous Drug Response</i>	468
V. SEVEN PILLARS OF THE NEW EVIDENTIARY PARADIGM	476
A. <i>Postmarket Evidence Development</i>	476
B. <i>The Flexible Best Evidence</i>	485
C. <i>Successive Improvement of the Risk-Benefit Ratio</i>	491
D. <i>Putting Evidence into the Hands of Innovators</i>	494
E. <i>Making Efficacy Regulation Effective</i>	497
F. <i>Piercing the Veil of Average Safety and Efficacy</i>	500
G. <i>Making Evidence Consequential</i>	503
VI. UNRESOLVED DOCTRINAL ISSUES	515
A. <i>Preemption of Suits Against Manufacturers After FDAAA and Wyeth v. Levine</i>	515
B. <i>Federal Intrusion on Medical Practice</i>	517
CONCLUSION	523

INTRODUCTION

The Food and Drug Administration Amendments Act of 2007¹ (FDAAA) transforms the evidentiary basis of medical product regulation by the U.S. Food and Drug Administration (FDA). FDAAA augments premarket clinical studies with new sources of evidence about the risks and benefits of drugs. FDAAA envisions heavy use, during

¹ Pub. L. No. 110-85, 121 Stat. 823 (codified as amended in scattered sections of 21 U.S.C.).

the postmarket period, of large observational studies that rely on interoperable health data networks. This shift will affect diverse areas of legal doctrine.

Product liability and medical malpractice are obvious candidates for impact. Questions about the preemptive effect of FDA's drug regulation loom large after FDAAA, notwithstanding the U.S. Supreme Court's March 2009 decision in *Wyeth v. Levine*.² "The major doctrinal question here is whether various forms of regulatory action by the FDA have the effect of preempting suits brought pursuant to state tort law."³ *Wyeth v. Levine* held that FDA's drug approval and labeling decisions did not preempt a failure-to-warn suit against the drug's manufacturer. That result seems fair, given that the evidence on which FDA has been basing these decisions is severely limited in ways this article explores.⁴ Congress enacted FDAAA to address these very limitations. FDAAA expands FDA's postmarket evidence development and empowers FDA to take new forms of regulatory action in response to the new flows of evidence. Some of these new regulatory actions may have the preemptive effect that the Court, in *Wyeth v. Levine*, found lacking in FDA's pre-FDAAA approval and labeling decisions. In assessing the impact of *Wyeth v. Levine*, it is crucial to understand that the Supreme Court was ruling on actions taken under a statutory scheme that had been substantially amended by the time the case was decided.

FDAAA directs FDA to address certain matters that Congress previously regarded as state medical practice issues. There may be constitutional questions as FDA flexes new powers to restrict clinical use of drugs and does so in ways that deny seriously ill patients access to FDA-approved pharmaceutical products—a problem that appears already to be occurring. FDA-imposed use restrictions raise the specter that physicians, as well as manufacturers, might argue preemption—or its close cousin, the regulatory compliance defense⁵—in drug-injury malpractice suits. In other areas, FDAAA breathes new life into classic infrastructure regulatory problems by requiring evidence that can only be generated with a massive, networked informational infrastructure that does not yet exist and will have to be financed,

2 129 S. Ct. 1187 (2009).

3 Richard A. Nagareda, *FDA Preemption: When Tort Law Meets the Administrative State*, 1 J. TORT L. art. 4, at 1 (2006), <http://www.bepress.com/jtl/voll/iss1/art4>.

4 See discussion *infra* Part II (discussing the limitations of the evidence on which FDA was relying prior to passage of FDAAA in 2007).

5 Robert L. Rabin, *Poking Holes in the Fabric of Tort: A Comment*, 56 DEPAUL L. REV. 293, 300 (2007).

built, and administered.⁶ The new forms of evidence, as they become available, may reshape cultural and commercial norms that have favored avid consumption of prescription drugs⁷ in recent decades, and they may alter popular expectations of medical privacy.⁸ The new research methodologies may require new structures for research oversight and new principles for the ethical conduct of biomedical research. This article touches on these broader impacts to open a scholarly dialogue about them. However, it focuses on how the broad, conceptual contours of drug regulation have changed and why. Understanding what was wrong with FDA's old evidentiary paradigm is key to appreciating the promise and the perils of its new one.

FDAAA is a profound change in law. The Federal Food, Drug, and Cosmetic Act⁹ ("FDCA" or "1938 Act") has been amended more than one hundred times¹⁰ since it became law in 1938. Most of these amendments were "tweaks" but, occasionally, Congress fundamentally altered the regulatory paradigm. FDAAA was a "sweeping overhaul of both the FDCA and the Public Health Service Act."¹¹ Once called

6 See generally Barbara J. Evans, *Congress' New Infrastructural Model of Medical Privacy*, 84 NOTRE DAME L. REV. 585 (2009) (discussing infrastructure regulatory aspects of FDAAA).

7 See, e.g., Press Release, U.S. Dep't of Health & Human Servs., *New Efforts to Help Improve Medical Products for Patient Safety and Quality of Medical Care* (May 22, 2008) [hereinafter HHS, *New Efforts*], available at <http://www.hhs.gov/news/press/2008pres/05/20080522a.html> (noting that Americans not on Medicare average thirteen prescriptions per year, while those on Medicare average about twenty-eight per year). Those Medicare beneficiaries who consider themselves in poor health receive about forty-five prescriptions per year. *Id.* (citing CTRS. FOR MEDICARE & MEDICAID SERVS., U.S. DEP'T OF HEALTH & HUMAN SERVS., *MEDICARE CURRENT BENEFICIARY SURVEY* (2004), available at <http://www.cms.hhs.gov/MCBS/Downloads/A04%20Ric%201.pdf>).

8 See Evans, *supra* note 6, at 610–22 (discussing privacy impacts of FDAAA).

9 Pub. L. No. 75-717, 52 Stat. 1040 (1938) (codified as amended at 21 U.S.C. §§ 301–399 (2006)).

10 See U.S.C.A. index (West Supp. 2009) (popular name table), Federal Food, Drug, and Cosmetic Act (FDCA) (Copeland Pure Food and Drugs Act) (Food, Drug, and Cosmetic Act) (FFDCA) (Humphrey-Durham Act); see also Peter Barton Hutt, *The State of Science at the Food and Drug Administration*, 60 ADMIN. L. REV. 431, 436, 467 tbl. 1 (2008) (counting statutes that directly affected FDA—including both amendments to the Food, Drug, and Cosmetic Act and other statutes that established duties to be performed by FDA—and finding ninety-plus such statutes enacted between 1938 and 1987 and one hundred statutes after 1988).

11 David A. Kessler & David C. Vladeck, *A Critical Examination of the FDA's Efforts to Preempt Failure-to-Warn Claims*, 96 GEO. L.J. 461, 467 (2008).

“the biggest reforms since at least 1997,”¹² FDAAA is in fact the most momentous shift in drug regulation in half a century. The 1962 Drug Amendments¹³ sparked the last big shift by introducing the modern concept of drug approval. Before 1962, new drugs were deemed approved for sale unless FDA, within a limited period of time, found grounds to keep them off the market.¹⁴ Since 1962, drugs await affirmative approval before they can be sold.¹⁵ It was after 1962 that “FDA became responsible for making benefit-risk decisions” about drugs.¹⁶ Congress directed FDA, for the first time, to require “substantial evidence”¹⁷ of efficacy as well as safety.¹⁸ FDA interpreted¹⁹ this language to require the familiar three-phase clinical trial process²⁰ through which drugs now pass before FDA approval.

12 Mark McClellan, *Drug Safety Reform at the FDA—Pendulum Swing or Systematic Improvement?*, 356 NEW ENG. J. MED. 1700, 1700 (2007) (describing FDAAA as “the biggest reforms since at least 1997”).

13 Drug Amendments of 1962 (Harris-Kefauver Act), Pub. L. No. 87-781, 76 Stat. 780 (codified as amended in scattered sections of 21 U.S.C.).

14 See *Wyeth v. Levine*, 129 S. Ct. 1187, 1195 (2009); Barry S. Roberts & David Z. Bodenheimer, *The Drug Amendments of 1962: The Anatomy of a Regulatory Failure*, 1982 ARIZ. ST. L.J. 581, 585; see also S. REP. NO. 87-1744, at 3–5 (1962), reprinted in 1962 U.S.C.C.A.N. 2884, 2890–91 (describing the earlier premarket notification process that the 1962 Drug Amendments replaced); Hutt, *supra* note 10, at 434 (noting that the 1962 Drug Amendments replaced premarket notification for drugs with premarket approval).

15 See Hutt, *supra* note 10, at 434–35.

16 INST. OF MED., UNDERSTANDING THE BENEFITS AND RISKS OF PHARMACEUTICALS 48 (Leslie Pray rapporteur, 2007) [hereinafter IOM, UNDERSTANDING BENEFITS] (reporting presentation of Peter Barton Hutt), available at http://www.nap.edu/openbook.php?record_id=11910.

17 See Drug Amendments of 1962, Pub. L. No. 87-781, § 102, 76 Stat. at 781 (defining “substantial evidence” as “evidence consisting of adequate and well-controlled 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”).

18 See *id.* § 102, 76 Stat. at 781–82 (amending 21 U.S.C. § 355(d) to allow drug applications to be denied for lack of substantial evidence of efficacy, and amending 21 U.S.C. § 355(e) to allow previously granted approvals to be withdrawn if new information reveals that a drug does not have its claimed effects).

19 See Hearing Regulations and Regulations Describing Scientific Content of Adequate and Well-Controlled Clinical Investigations, 35 Fed. Reg. 7250, 7250 (May 8, 1970) (amending 21 C.F.R. §§ 130.12(a)(5), 130.14 to clarify the terms “substantial evidence” and “adequate and well-controlled clinical investigation[s]”).

20 See Investigational New Drug Application, 21 C.F.R. § 312.21 (2009); see also Geoffrey M. Levitt et al., *Human Drug Regulation*, in 2 FUNDAMENTALS OF LAW AND REGULATION: AN IN-DEPTH LOOK AT THERAPEUTIC PRODUCTS 159, 165–66 (David G.

These requirements contributed to expansive growth of clinical trial activity late in the twentieth century. Between 1962 and 1980, the cost to develop a new drug rose from \$6.5 million to \$70 million,²¹ with much of the increase attributed to the cost of FDA's required clinical studies.²² In 2007, the Institute of Medicine reported various estimates ranging from \$500 million to \$2 billion as the cost of developing one new drug.²³ These data suggest a clear upward trend although historical and recent figures are not strictly comparable.²⁴ Robert Califf traces how FDA's clinical trial requirements led to a "complex maze of rules . . . and ultimately created an entirely new industry: the contract research organization (CRO)"²⁵ to help drug sponsors navigate FDA's premarket requirements. As of March 2008, the CRO industry had estimated annual revenues of \$17.8 billion, employed 100,000 people worldwide, and was experiencing annual revenue growth of 14–16%.²⁶

In their day, the 1962 amendments marked the most significant shift in regulatory philosophy not only since the 1938 Act, but since the 1906 Pure Food and Drug Act²⁷ that had preceded it.²⁸ Another important shift was the 1976 Medical Device Amendments²⁹ which authorized FDA to regulate medical devices, a category that later has

Adams et al. eds., 1999) [hereinafter FUNDAMENTALS OF LAW AND REGULATION] (describing the three phases of clinical investigations FDA typically requires).

21 Roberts & Bodenheimer, *supra* note 14, at 606.

22 See *id.*; see also HENRY GRABOWSKI, DRUG REGULATION AND INNOVATION 36 (1976) (finding a rough doubling of costs by the early 1970s); SAM PELTZMAN, REGULATION OF PHARMACEUTICAL INNOVATION 2 (1974) (same).

23 COMM. ON THE ASSESSMENT OF THE U.S. DRUG SAFETY SYS., INST. OF MED., THE FUTURE OF DRUG SAFETY 32 (Alina Baciu et al. eds., 2007) [hereinafter IOM, FUTURE OF DRUG SAFETY], available at http://books.nap.edu/openbook.php?record_id=11750.

24 Methodologies for estimating drug development costs are subject to ongoing controversy. See *id.* at 32. Recognizing that many drugs are abandoned for every one drug that does reach the market, modern estimates of drug development costs may include costs for drugs that were abandoned prior to FDA approval as part of the average cost of developing the one successful drug that does reach the market. It is unclear whether the figures reported for 1962 and 1980 reflect this same practice. See Roberts & Bodenheimer, *supra* note 14, at 606.

25 Robert M. Califf, *Clinical Trials Bureaucracy: Unintended Consequences of Well-Intentioned Policy*, 3 CLINICAL TRIALS 496, 497 (2006).

26 See Ass'n of Clinical Research Orgs., CRO Industry at a Glance, <http://www.acrohealth.org/industry-atagance.php> (last visited Oct. 16, 2009).

27 Pub. L. No. 59-384, 34 Stat. 768 (codified as amended in scattered sections of 21 U.S.C.).

28 See Alan H. Kaplan, *Fifty Years of Drug Amendments Revisited: In Easy-to-Swallow Capsule Form*, 50 FOOD & DRUG L.J. 179, 181 (1995).

29 Pub. L. No. 94-295, 90 Stat. 539 (codified as amended at 15 U.S.C. § 55 (2006) and scattered sections of 21 U.S.C.).

acquired significance in genomic medicine since it includes diagnostic and genetic tests. FDAAA belongs in this class of paradigm-shifting amendments.

If the 1962 amendments heralded the age of faith in clinical trials, then FDAAA tolls its twilight. That, in a nutshell, is the shift of regulatory philosophy now underway. It is important to qualify what that statement means: Declaring the twilight of premarket clinical trials is not the same thing as saying that the sun already has set, or will ever set, on them. Clinical trials will still be performed and, indeed, FDAAA expanded FDA's authority to require them in the postmarket as well as the premarket period.³⁰ Clinical trials will continue to be an important—often, the primary—source of data about medical products. Yet FDAAA accepts that clinical trials have intrinsic limitations and cannot, by themselves, ensure the safety and efficacy of FDA-approved products. FDAAA is a response to this reality. Premarket trials will continue as before, but there has been a pragmatic reassessment of their evidentiary value. One might refer to this as the nation's *Trialdämmerung*—its loss of faith in evidence from premarket drug trials. Derived from “clinical trial” and the German word, *Dämmerung* (meaning “dimming” or “twilight”³¹), the term deliberately alludes to *Götterdämmerung* with its sense of a “collapse (as of a . . . regime) marked by catastrophic violence and disorder; broadly DOWNFALL.”³²

The cataclysmic event, in this case, was a confluence of two trends. First, deadly drug safety problems³³ earlier this decade stirred doubts whether FDA's old evidentiary paradigm was capable of protecting the public's health.³⁴ Dating back to 1962, that paradigm relied heavily on risk-benefit data from premarket clinical trials.³⁵ The methodological limitations of clinical trials³⁶ were never concealed by FDA³⁷ and had long been known to biomedical statisti-

30 See discussion *infra* Part V.A.1.

31 Merriam Webster's Online Dictionary, *Götterdämmerung*, <http://merriam-webster.com/dictionary/gotterdammerung> (last visited Jan. 12, 2010).

32 *Id.*

33 See discussion *infra* Part I.

34 See Kessler & Vladeck, *supra* note 11, at 468 (noting that aspects of FDAAA suggest “that Congress did not share the FDA's view that it is capable of adequately safeguarding the public health on its own”).

35 See discussion *infra* Part I.

36 See discussion *infra* Part II.

37 See Kenneth L. Melmon, *Attitudinal Factors that Influence the Utilization of Modern Evaluative Methods*, in INST. OF MED., MODERN METHODS OF CLINICAL INVESTIGATION 135, 142 (Annetine C. Gelijns, ed., 1990) [hereinafter IOM, MODERN METHODS], available at http://www.nap.edu/openbook.php?record_id=1550 (pointing out that

cians,³⁸ but they were not widely appreciated by the public, academics, or the medical profession.³⁹ Problems with the 1962 drug approval framework already were apparent by the mid-1970s and a Joint Commission on Prescription Drug Use was formed in 1976 with Dr. Kenneth L. Melmon as its chair.⁴⁰ The Commission's voluminous 1980 report set out the limitations of clinical trial evidence⁴¹ and called for a "comprehensive system of post-marketing drug surveillance" remarkably similar to what Congress authorized twenty-seven years later in FDAAA.⁴² Drug safety scandals this decade displayed the problem in a way even non-statisticians could grasp: "safe and effective" drugs were killing people, lots of people. Widely held beliefs that FDA regulation was "comprehensive"⁴³ and involved "rigorous cost-benefit analysis along the dimensions of safety and efficacy"⁴⁴ seemed to support preemption of tort suits but presumed FDA was using evidence that *could* support rigorous analysis. This presumption was wrong.⁴⁵

The second trend was a series of scientific advances flowing from the Human Genome Project⁴⁶ which, as explored below,⁴⁷ exposed

FDA knows the limitations of premarket clinical trials and has made efforts to explain their practical limitations).

38 See IOM, UNDERSTANDING BENEFITS, *supra* note 16, at 29 (reporting presentation of Dr. Brian L. Strom).

39 See *id.*; Melmon, *supra* note 37, at 142.

40 See JOINT COMM'N ON PRESCRIPTION DRUG USE, FINAL REPORT (1980).

41 See Brian L. Strom et al., *Efficacy Research on Marketed Drugs: Issues of Validity*, in JOINT COMM'N ON PRESCRIPTION DRUG USE, *supra* note 40, at app. VII; see also William L. Finkle, *Sample Size Requirements for Detection of Drug-Induced Disease*, in JOINT COMM'N ON PRESCRIPTION DRUG USE, *supra* note 40, at app. V. (describing intrinsic methodological limitations of clinical trial evidence on which FDA bases its drug approval decisions).

42 JOINT COMM'N ON PRESCRIPTION DRUG USE, *supra* note 40, at 7–93 (describing the Joint Commission's proposed postmarket surveillance system); *id.* at 94–108 (foreseeing, in 1980, that privacy and data security would be major concerns with such a system).

43 *Wyeth v. Levine*, 129 S. Ct. 1187, 1219 (2009) (Alito J., dissenting); see also Richard A. Epstein, *Why the FDA Must Preempt Tort Litigation: A Critique of Chevron Deference and a Response to Richard Nagareda*, 1 J. TORT L. art. 5, at 3 (2006), <http://www.bepress.com/jtl/vol1/iss1/art5> ("I think the right rule in all cases should be that, in light of comprehensive FDA regulation of new and existing drugs, these actions should be preempted . . ."); Rabin, *supra* note 5, at 304 ("FDA regulation is comprehensive, at least as far as prescription drugs are concerned.").

44 Rabin, *supra* note 5, at 304.

45 See discussion *infra* Parts II–III.

46 For background on the Human Genome Project, see Nat'l Human Genome Research Inst., All About the Human Genome Project (HGP), <http://www.genome.gov/10001772> (last visited Oct. 20, 2009).

47 See discussion *infra* Part IV.

flaws in FDA's approach to effectiveness as well as safety. These two trends converged just as technological advances were starting to offer practical alternatives to FDA's old evidentiary approach.⁴⁸ This Article explores these trends and speculates what the new paradigm—which already has been authorized by Congress but which is still, very much, in the early stages of implementation—may look like. Part V outlines seven pillars of FDA's new evidentiary paradigm: seven novel propositions that reject foundational assumptions of the drug regulatory framework in place since 1962. Collapse of these assumptions sets off ripple effects in various doctrinal areas.

If fully implemented, FDAAA will change the way regulators manage—and the public thinks about—risks and benefits of the roughly \$230 billion⁴⁹ of prescription drugs Americans buy each year. It will promote more realistic expectations of what can (and cannot) be learned from the \$60–\$80 billion⁵⁰ Americans invest each year in drug-related research and it will shape policy on new types of research to receive greater emphasis in the future. However, that “if” is a big one: *if* fully implemented.⁵¹ In 2006, the Institute of Medicine remarked on FDA's record of poor follow-through on proposed initiatives in recent years.⁵² FDAAA sketches out a blueprint for a bold new regulatory framework, but implementation has only just begun.

I. THE 1962 EVIDENTIARY PARADIGM FOR DRUG REGULATION

This Part briefly surveys the track record of FDA's drug approval process after 1962 and describes its evidentiary requirements, to lay groundwork for exploring its limitations in Part II.

48 See discussion *infra* Parts I.C and V.A.2.

49 See Ctrs. for Medicare & Medicaid Servs., U.S. Dep't of Health & Human Servs., National Health Expenditures by Type of Service and Source of Funds (2007), *available at* <http://www.cms.hhs.gov/NationalHealthExpendData/downloads/nhe2007.zip>.

50 See Press Release, PhRMA, R&D Spending by U.S. Biopharmaceutical Companies Reaches Record Levels in 2008 Despite Economic Challenges (Mar. 10, 2009), *available at* http://www.phrma.org/news_room/press_releases/r&d_spending_by_u.s._biopharmaceutical_companies_reaches_record_levels_in_2008_despite_economic_chal (reporting private-sector investment in drug and vaccine development of \$65.2 billion in 2008); see also OFFICE OF COMM'NS & PUB. LIAISON, NAT'L. INSTS. OF HEALTH, PUB. NO. 08-05, NIH ALMANAC 2008–2009, at 401 (2008), *available at* http://www.nih.gov/about/almanac/Almanac_2008_2009.pdf (showing NIH appropriations of \$28.5 billion in 2006 and \$29 billion in 2007, figures which include various forms of biomedical research not all of which are related to pharmaceuticals).

51 See Kessler & Vladeck, *supra* note 11, at 469 (noting that FDAAA is “dauntingly complex” and expressing concern whether implementation is adequately resourced).

52 See IOM, FUTURE OF DRUG SAFETY, *supra* note 23, at 17–18.

A. *The Gap Between Perception and Performance*

“Over time, a large segment of the public has developed the belief that FDA-approved drugs carry no risk.”⁵³ Short of such extreme credulity, many people know FDA-approved drugs carry some risk but may overestimate the amount and quality of risk-benefit data available at the time drugs are approved. The clinical trials on which FDA relies “have inherent methodological limitations in the evaluation of adverse effects.”⁵⁴ The late Dr. Melmon decried the fact that “academics and the medical profession shirk responsibility for understanding limitations of the regulatory process.”⁵⁵ “Shirking” may be too strong a term. Limitations of the evidence FDA uses often are described in statistical terminology that can thwart even sincere attempts to grasp the impact.

Between 1969 and 2002—that is, well after the 1962 Drug Amendments were in place but before the 2004 scandal with rofecoxib (Vioxx)⁵⁶—more than seventy-five drugs and drug products were withdrawn from the market for safety reasons.⁵⁷ Black-box warnings had to be added to many other drugs.⁵⁸ Other countries experienced similar problems. Between 1963 and 2004, Canada had forty-one drug withdrawals that appear to have been safety-related.⁵⁹ Drug withdrawals are rare relative to the total number of drugs FDA approves. Roughly 11,000 FDA-approved drugs are available in the United States (including prescription and over-the-counter drugs).⁶⁰

53 IOM, UNDERSTANDING BENEFITS, *supra* note 16, at 7; *see also id.* at 1 (“[M]any individuals believe that drugs approved by the U.S. Food and Drug Administration (FDA) carry no risks.”).

54 LAWRENCE M. FRIEDMAN ET AL., FUNDAMENTALS OF CLINICAL TRIALS 182 (3d ed. 1998).

55 Melmon, *supra* note 37, at 142.

56 *See infra* notes 249–51 and accompanying text.

57 Diane K. Wysowski & Lynette Swartz, *Adverse Drug Event Surveillance and Drug Withdrawals in the United States, 1969–2002*, 165 ARCHIVE INTERNAL MED. 1363, 1363 (2005).

58 *See* Karen E. Lasser et al., *Timing of New Black Box Warnings and Withdrawals for Prescription Medications*, 287 JAMA 2215, 2216 (2002) (reporting that more than 10% of new chemical entities approved between 1975 and 1999 were given black-box warnings or withdrawn).

59 Joel Lexchin, *Drug Withdrawals from the Canadian Market for Safety Reasons, 1963–2004*, 172 CANADIAN MED. ASS'N J. 765, 765 (2005).

60 *See* Wyeth v. Levine, 129 S. Ct. 1187, 1202 (2009) (estimating the number of drugs at 11,000); BENCT D. FURBERG & CURT D. FURBERG, EVALUATING CLINICAL RESEARCH 115 (2d ed. 2007) (estimating the number of drugs at 10,000). *But see* Ctr. for Drug Evaluation & Research (CDER), Drugs@FDA: FDA Approved Drug Products, <http://www.accessdata.fda.gov/Scripts/cder/DrugsatFDA> (last visited Nov. 10, 2009) (listing approved products and reporting 5996 prescription and over-the-

FDA approved 8076 new drug applications from 1938–1962 and an additional 3638 from 1963–2006.⁶¹ Seventy-five safety-related withdrawals, out of 3638 approvals granted after the 1962 Drug Amendments, is roughly 2%. Pillans detected a recent increase in the rate of drug withdrawals and notes that there are competing explanations, such as laxer approval standards, higher overall levels of drug availability and consumption, or improved systems for detecting safety problems.⁶² A current estimate is that 20% of all drugs receive a black-box warning and about 4% are withdrawn for safety reasons.⁶³

Consistent with an upward trend, the last ten years saw a spate of incidents where serious risks came to light after drugs were approved.⁶⁴ Beyond the Vioxx problem, there was fulminant liver failure with troglitazone (Rezulin),⁶⁵ suicidal ideation in youngsters taking antidepressants (selective serotonin reuptake inhibitors or SSRIs),⁶⁶ rhabdomyolysis and kidney failure in people taking chole-

counter drugs and biologic therapies as of Nov. 10, 2009). The 5996 figure reportedly includes most prescription and over-the-counter drugs and therapeutic biologics approved for use in the United States since 1939 and excludes products that were not approved through new drug applications (NDA) or biologic license applications (BLA), drugs that failed to achieve FDA approval, drugs approved outside the United States but not in the United States, and dietary supplements. *See id.* (follow link for “FAQ” and scroll to questions three and four). The apparent discrepancy with higher estimates is unexplained but likely reflects drugs that were approved through alternative pathways, including pre-1938 drugs that were generally recognized as safe and roughly 4000 drugs approved 1938–1962 for safety alone, which subsequently went through the Drug Efficacy Study Implementation (DESI) process to confirm efficacy. *See* CTR. FOR DRUG EVALUATION & RESEARCH, U.S. FOOD & DRUG ADMIN., THE HISTORY OF DRUG REGULATION IN THE UNITED STATES 15 (2006), available at <http://www.fda.gov/downloads/AboutFDA/WhatWeDo/History/ProductRegulation/PromotingSafeandEffectiveDrugsfor100Years/UCM114469.pdf>.

61 U.S. Food & Drug Admin., Summary of NDA Approvals & Receipts, 1938 to the Present, <http://www.fda.gov/AboutFDA/WhatWeDo/History/ProductRegulation/SummaryofNDAApprovalsReceipts1938tothepresent/default.htm> (last visited Nov. 9, 2009).

62 Peter Ian Pillans, *Clinical Perspectives in Drug Safety and Adverse Drug Reactions*, 1 EXPERT REV. CLINICAL PHARMACOLOGY 695, 700 (2008).

63 Mei Sheng Duh et al., *The Role of Epidemiology in Drug Safety Litigations*, FOOD & DRUG L. INST. UPDATE, NOV./DEC. 2007, at 31, 31.

64 *See* Alastair J.J. Wood et al., *Making Medicines Safer—The Need for an Independent Drug Safety Board*, 339 NEW ENG. J. MED. 1851 (1998).

65 *See* Gerald A. Faich & Richard H. Moseley, *Troglitazone (Rezulin) and Hepatic Injury*, 10 PHARMACOEPIDEMIOLOGY & DRUG SAFETY 537 (2001); David Graham et al., *Incidence of Idiopathic Acute Liver Failure and Hospitalized Liver Injury in Patients Treated with Troglitazone*, 98 AM. J. GASTROENTEROLOGY 175 (2003).

66 *See* U.S. Food & Drug Admin., Labeling Change Request Letter for Antidepressant Medications, <http://www.fda.gov/Drugs/DrugSafety/InformationbyDrugClass/ucm096352.htm> (last visited Nov. 9, 2009).

terol-lowering cerivastatin (Baycol),⁶⁷ QT prolongation and ventricular arrhythmias with the gastroesophageal reflux drug cisapride (Propulsid),⁶⁸ cardiac problems with the antihistamines astemizole (Hismanal) and terfenadine (Seldane),⁶⁹ and a suite of health risks for women using hormone replacement therapy (HRT).⁷⁰ The Governmental Accountability Office (GAO) counted ten safety-related withdrawals between 2000 and March 2006.⁷¹ More recent problems have included cardiovascular risks with diabetes drugs;⁷² reports of esophageal cancer and necrotic jawbones in patients taking osteoporosis drugs;⁷³ elevated blood pressure among patients taking popular hyperactivity drugs;⁷⁴ this list is far from exhaustive. "It has been estimated that as many as half of all new drugs have at least one serious adverse effect that is unknown at the time of drug approval."⁷⁵

67 See FURBERG & FURBERG, *supra* note 60, at 118 (noting differences in rates of rhabdomyolysis between cerivastatin and other marketed statins); Judy Staffa et al., *Cerivastatin and Reports of Fatal Rhabdomyolysis*, 346 NEW ENG. J. MED. 539, 539–40 (2002); see also U.S. Food & Drug Admin., FDA Public Health Advisory For Crestor (Rosuvastatin) (Mar. 2, 2005), available at <http://www.fda.gov/Drugs/DrugSafety/PublicHealthAdvisories/ucm051756.htm> (reviewing potential safety issues with the cholesterol-lowering drug Crestor).

68 See Diane Wysowski et al., *Postmarketing Reports of QT Prolongation and Ventricular Arrhythmia in Association with Cisapride and Food and Drug Administration Regulatory Actions*, 96 AM. J. GASTROENTEROLOGY 1698 (2001).

69 See Amalia M. Issa, *Clinical Applications of Pharmacogenomics to Adverse Drug Reactions*, 1 EXPERT REV. CLINICAL PHARMACOLOGY 251, 253 (2008).

70 See Press Release, U.S. Food & Drug Admin., FDA Updates Hormone Therapy Information for Post Menopausal Women (Feb. 10, 2004), available at <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2004.ucm108243.htm>.

71 U.S. GOV'T ACCOUNTABILITY OFFICE, DOC. NO. GAO-06-402, DRUG SAFETY: IMPROVEMENT NEEDED IN FDA'S POSTMARKET DECISION-MAKING AND OVERSIGHT PROCESSES 10 (2006) [hereinafter GAO, DRUG SAFETY], available at <http://www.gao.gov/new.items/d06402.pdf>.

72 CTR. FOR DRUG EVALUATION & RESEARCH, U.S. FOOD & DRUG ADMIN., GUIDANCE FOR INDUSTRY: DIABETES MELLITUS—EVALUATING CARDIOVASCULAR RISK IN NEW ANTIDIABETIC THERAPIES TO TREAT TYPE 2 DIABETES (2008) [hereinafter FDA, DIABETES MELLITUS], available at <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071627.pdf>.

73 Jill U. Adams, *Osteoporosis Drug Fosamax Linked to Serious Diseases*, L.A. TIMES, Jan. 12, 2009, <http://articles.latimes.com/2009/jan/12/health/he-closer12>.

74 See Meeting Transcript of Feb. 9, 2006, Drug Safety & Risk Mgmt. Advisory Comm., Ctr. for Drug Evaluation & Research, at 35–40, 220–26, available at <http://www.fda.gov/ohrms/dockets/ac/06/transcripts/2006-4202t1.pdf> (statements of Drs. Kate Gelperin & Steve Nissen).

75 FURBERG & FURBERG, *supra* note 60, at 8.

David Slavin has remarked that high public trust is typically associated with low perceived risk and vice versa.⁷⁶ Late-discovered risks and drug withdrawals both tend to be seen as regulatory failures, as evidence that regulators made mistakes.⁷⁷ Public confidence in FDA fell from 80% in the 1970s to 61% in 2000; 56% in 2004; and 36% in 2006.⁷⁸ Certainly, there have been instances where premarket studies (or early clinical experience in the postmarket period) gave early signals of a safety problem on which regulators failed to press. Such allegations were voiced in connection with Vioxx.⁷⁹ Obviously, if safety signals exist but elicit a flaccid regulatory response, then that is a problem that needs to be corrected. However, too much focus on regulator fault obscures a deeper problem: “In reality, almost all postmarketing safety issues involve rare adverse events that could not have been detected prior to marketing.”⁸⁰ The problem may not be failure of the regulator; it may be failure of the evidentiary paradigm.

B. *The Role of Randomized, Controlled Trials*

FDA’s post-1962 evidentiary paradigm concentrated research and regulatory effort in the premarket period, before products are approved.⁸¹ Congress gave FDA little power to require ongoing studies past the point of drug approval,⁸² and the agency was given few

76 IOM, UNDERSTANDING BENEFITS, *supra* note 16, at 29 (reporting presentation of Dr. David E. Slavin).

77 IOM, FUTURE OF DRUG SAFETY, *supra* note 23, at 2; *see also* IOM, UNDERSTANDING BENEFITS, *supra* note 16, at 29 (noting that the “public misunderstands drug safety, believing that postmarketing discovery of adverse drug reactions means ‘somebody messed up’”).

78 Hutt, *supra* note 10, at 443 (citing Harris Poll survey results reported by Bill Hubbard and Steven Grossman, April 11, 2007).

79 *See, e.g.*, Richard Horton, *Vioxx, the Implosion of Merck, and Aftershocks at the FDA*, 364 LANCET 1995 (2004); Pillans, *supra* note 62, at 700; Eric J. Topol, *Failing the Public Health—Rofecoxib, Merck, and the FDA*, 351 NEW ENG. J. MED. 1707 (2004).

80 IOM, UNDERSTANDING BENEFITS, *supra* note 16, at 29 (reporting presentation of Dr. Brian L. Strom); *see also* SALLY ROBINSON ET AL., INST. OF MED., EMERGING SAFETY SCIENCE: WORKSHOP SUMMARY 2 (2008) [hereinafter IOM, EMERGING SAFETY SCIENCE], available at http://books.nap.edu/openbook.php?record_id=11975 (“With current methods, it is unlikely that rare safety problems will be identified prior to approval.”).

81 Barbara J. Evans & David A. Flockhart, *The Unfinished Business of U.S. Drug Safety Regulation*, 61 FOOD & DRUG L.J. 45, 53–54 (2006); *see also* Kessler & Vladeck, *supra* note 11, at 485 (citing FDA’s relative staffing of pre- and postmarket functions to demonstrate the agency’s heavy historical emphasis on premarket regulation).

82 *See generally* GAO, DRUG SAFETY, *supra* note 71 (evaluating the weakness of FDA’s powers during the postmarket period and proposing reforms); *see also* discussion *infra* Part V.A.1 (describing FDA’s limited powers during the postmarket period prior to passage of FDAAA).

tools—other than the “blunt instrument”⁸³ of revoking a previously granted approval—with which to manage late-emerging risks. By 2005, FDA was employing 1000 people to review approval applications for a few dozen new drugs each year, but only had 100 professional employees engaged in postmarket monitoring of 3000 approved prescription drugs plus about 8000 over-the-counter medications.⁸⁴ This front-loaded paradigm rested on a conceit that premarket studies can identify a drug’s material risks and benefits before it goes on the market. This is now known not to be the case. “At the time of approval, the benefit-risk profile of a typical drug is not fully understood. It is only after approval and widespread use that the profile will become fully understood.”⁸⁵ If approval is a necessary step to discover a drug’s risks, then risks will not be known at the time of approval. That is the intrinsic flaw in the 1962 evidentiary paradigm.

Randomized, controlled clinical trials⁸⁶ (RCTs) occupy a central position in that paradigm.⁸⁷ When FDA was implementing the 1962 Drug Amendments, RCTs had only recently emerged as the preferred method for evaluating medical treatments. The first modern, multicenter randomized, controlled clinical trial was reported in 1948.⁸⁸

83 IOM, *FUTURE OF DRUG SAFETY*, *supra* note 23, at 153; *see also* discussion *infra* Part V.G (discussing FDA’s lack of authority, prior to FDAAA, to require labeling changes to manage new risks discovered during the postmarket period).

84 Kessler & Vladeck, *supra* note 11, at 485.

85 IOM, *UNDERSTANDING BENEFITS*, *supra* note 16, at 4.

86 The term “clinical trial” is variously defined but the common elements of most definitions are as follows: The study is prospective, following study subjects forward in time from a defined baseline point. FRIEDMAN ET AL., *supra* note 54, at 2; FURBERG & FURBERG, *supra* note 60, at 11. The baseline point need not correspond to a particular calendar date but might be defined as a stage of illness that all subjects exhibit on whichever dates they enter the study. FRIEDMAN ET AL., *supra* note 54, at 2. Concurrent groups of study subjects receive either an intervention (one or more treatments that are under study) or a control (either a placebo or an alternative treatment with which the intervention is being compared). If a clinical trial is randomized, subjects are assigned randomly to receive either the intervention or the control. *Id.* at 2; FURBERG & FURBERG, *supra* note 60, at 11. There is a predefined primary question to be answered and, at most, a limited number of predefined secondary questions that are related to the primary question. FRIEDMAN ET AL., *supra* note 54, at 16–17.

87 *See* 21 C.F.R. § 314.50(d)(5)(ii) (2009) (calling for data from controlled clinical studies in new drug applications).

88 FRIEDMAN ET AL., *supra* note 54, at 1 (citing Streptomycin in Tuberculosis Trials Comm., Med. Research Council, *Streptomycin Treatment of Pulmonary Tuberculosis*, 2 BRIT. MED. J. 768, 769–78 (1948)); Califf, *supra* note 25, at 496 (same). *But see* FRIEDMAN ET AL., *supra* note 54, at 1 (noting the use of randomization via coin toss in a 1931 clinical trial, J.B. Amberson et al., *A Clinical Trial of Sanocrysin in Pulmonary Tuberculosis*, 25 AM. REV. TUBERCULOSIS 401 (1931), which also was the first reported study to use blinding).

Until FDA began requiring RCTs, it was not uncommon to assess efficacy by comparing people treated with a drug to historical controls—that is, to people treated in earlier years with the then-standard care.⁸⁹ Safety data were not well standardized and sometimes consisted only of laboratory and animal studies.⁹⁰ Efficacy studies can be biased when they rely on historical controls. For example, patients on a newer drug may get well because of unrelated changes in patient care, or their improvement may reflect unexplained trends in the prevalence and severity of disease.⁹¹ RCTs, which include contemporaneous control subjects, avoid these historical biases.

Randomization⁹² eliminates additional types of bias, such as allocation bias that might arise if investigators picked people to receive the intervention instead of the control and subliminally assigned healthier-looking subjects to receive the intervention.⁹³ Double-blinding is an additional enhancement that eliminates ascertainment bias (the tendency to perceive improvement in patients known to be receiving the intervention) by keeping the investigators and subjects unaware of their treatment assignments.⁹⁴ A trial can be an RCT without being blinded. For example, blinding may be impossible in a trial where patients are randomly assigned to receive psychoanalysis versus a drug to treat their depression; people will know which one they received.⁹⁵ Blinding is feasible in most clinical drug trials, which typically compare one pill to another. Thus, in FDA drug studies, “RCT” generally means a double-blinded RCT.

89 See generally Henry Sacks et al., *Randomized Versus Historical Controls for Clinical Trials*, 72 AM. J. MED. 233 (1982) (providing a meta-analysis of results from RCTs and trials that had used historical controls in contexts where both had been performed).

90 Lynne Kessler Lechter, *Regulatory Overkill and the AIDS Patient*, 1 ALB. L.J. SCI. & TECH. 131, 145–46 (1991) (discussing deficiencies of premarket safety data in the period after the 1938 Act but before the 1962 amendments).

91 See FRIEDMAN ET AL., *supra* note 54, at 48–49 (discussing an unexplained decline in coronary heart disease in the general population over the past twenty years, which could create a false appearance that new treatments are effective if they were compared to historical controls).

92 For a definition of randomization, see *supra* note 86.

93 FURBERG & FURBERG, *supra* note 60, at 14–15.

94 *Id.* at 15.

95 Cf. Harlan Weisman et al., *Broader Post-Marketing Surveillance for Insights on Risk and Effectiveness*, in ROUNDTABLE ON EVIDENCE-BASED MED., INST. OF MED., THE LEARNING HEALTHCARE SYSTEM: WORKSHOP SUMMARY 128, 130 (LeighAnne Olsen et al. eds., 2007) [hereinafter IOM, LEARNING HEALTHCARE], available at http://books.nap.edu/openbook.php?record_id=11903 (noting the inherent difficulty in blinding certain types of clinical trials).

FDA's premarket evidentiary requirements for drugs include preclinical investigations (laboratory and animal studies)⁹⁶ followed by three phases of clinical studies in humans. The required phase I and phase II studies,⁹⁷ though often referred to as "trials," are not actually RCTs in a strict sense.⁹⁸ Phase III drug trials truly are RCTs.⁹⁹ FDA also may require RCTs for a subset of medical devices which, due to their novelty and/or significant risk, fall under FDA's premarket approval (PMA) requirements.¹⁰⁰ Devices that are less novel and risky pass through a less-rigorous clearance process¹⁰¹ that generally does not require full-fledged clinical trials, but does require premarket research to support the risk classification and to validate any analytical or clinical claims their sponsors plan to make about them.

Califf notes that "it was only in the 1970s that the clinical trial became a widely-accepted tool for the biomedical enterprise."¹⁰² The fact that FDA interpreted the 1962 amendments as requiring their widespread use may have contributed to this growing acceptance in the 1970s. Today, RCTs are regarded by many as the "gold standard"¹⁰³ for evaluating new drugs. However, this statement is subject

96 21 C.F.R. § 312.23(a)(8) (2009); see also Levitt et al., *supra* note 20, at 160.

97 See COMM. ON IDENTIFYING & PREVENTING MEDICATION ERRORS, INST. OF MED., PREVENTING MEDICATION ERRORS 55-56 (Philip Aspden et al. eds., 2007) [hereinafter IOM, PREVENTING MEDICATION ERRORS], available at http://books.nap.edu/openbook.php?record_id=11623 (describing phase I and phase II studies); FRIEDMAN ET AL., *supra* note 54, at 3-6 (describing and distinguishing the various phases of clinical investigations).

98 FRIEDMAN ET AL., *supra* note 54, at 3-5; see also *supra* note 86 (defining RCT).

99 FRIEDMAN ET AL., *supra* note 54, at 5; IOM, PREVENTING MEDICATION ERRORS, *supra* note 97, at 56.

100 See 21 U.S.C. § 360e (2006); 21 C.F.R. §§ 814, 860.7 (2009); see also Howard M. Holstein & Edward C. Wilson, *Developments in Medical Device Regulation*, in 2 FUNDAMENTALS OF LAW AND REGULATION, *supra* note 20, at 257, 282 (noting that evidentiary standards for devices subject to premarket approval are potentially less rigorous than for drugs, but that FDA often requires RCTs).

101 See FDCA §510(k), 21 U.S.C. § 360(k); 21 C.F.R. §§ 807.81-807.100; see also Nagareda, *supra* note 3, at 8-12 (comparing and contrasting the evidentiary requirements for 510(k) and PMA devices).

102 Califf, *supra* note 25, at 496.

103 See, e.g., Weisman et al., *supra* note 95, at 130 ("Randomized clinical trials with blinding are the gold standard for drug evaluations of safety and efficacy but may not be possible in device studies."); Califf, *supra* note 25, at 496 ("As randomized controlled trials have become the 'gold standard' for medical research, a complex regulatory structure for the conduct of clinical trials has emerged."); Jonathan B. Perlin & Joel Kupersmith, *Information Technology and the Inferential Gap*, 26 HEALTH AFF. 192, 192 (2007) (noting that "[r]andomized controlled trials are the current gold standard of evidence development" but discussing limits to their generalizability to real-world patient care); T. Pincus, *Limitations of Randomized Controlled Clinical Trials to*

to certain qualifications. Because RCTs eliminate biases to the internal validity of study results (such as the allocation, ascertainment, and historical biases mentioned earlier) they are correctly regarded as the highest-quality evidence of drug *efficacy*.¹⁰⁴ However, the same may not be true with respect to drug safety,¹⁰⁵ and it is increasingly clear that RCTs may not provide a complete picture of a drug's effectiveness.¹⁰⁶ Efficacy (how well a drug works for carefully chosen trial subjects in the ideal setting of a clinical trial) and effectiveness (how well it works in real patients in the actual health-care setting) can be two different things.¹⁰⁷

C. Evidentiary Paths Not Taken in 1962

The perceived strength of medical evidence varies depending on the study design.¹⁰⁸ Efforts to define a hierarchy of evidence generally accord the highest ranking to RCTs and properly performed meta-analyses of multiple RCTs.¹⁰⁹ Just below this is evidence from well-

Depict Accurately Long-Term Outcomes in Rheumatoid Arthritis, 57 ZEITSCHRIFT FÜR RHEUMATOLOGIE 46, 46 (1998) (noting that RCTs are regarded as the "gold standard" but criticizing their limitations); David L. Sackett et al., *Evidence Based Medicine: What It Is and What It Isn't*, 312 BRIT. MED. J. 71, 72 (1996) (noting that while RCTs are the "gold standard," some questions about therapy can be answered by other methods or may be too time-sensitive to await trial results (internal quotation marks omitted)).

104 IOM, PREVENTING MEDICATION ERRORS, *supra* note 97, at 57 ("While randomized controlled trials are considered the gold standard for assessing efficacy, they rarely provide all the information needed in clinical practice."); John Concato et al., *Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs*, 342 NEW ENG. J. MED. 1887, 1887 (2000) ("Randomized, controlled trials . . . have become the gold standard for assessing the effectiveness of therapeutic agents."); FURBERG & FURBERG, *supra* note 60, at 17 ("[R]andomized controlled trials are the gold standard for evaluating the efficacy of medical interventions").

105 See *infra* Part II.

106 See *infra* Part IV.

107 Leslie L. Roos et al., *Strengths and Weaknesses of Health Insurance Data Systems for Assessing Outcomes*, in IOM, MODERN METHODS, *supra* note 37, at 47, 50–51; see also Pillans, *supra* note 62, at 700 (noting that real-world practice is "far removed from the premarketing clinical trial conditions").

108 See U.S. PREVENTIVE SERVS. TASK FORCE, GUIDE TO CLINICAL PREVENTIVE SERVICES (2d ed. 1996) (comparing the quality of evidence produced by various types of study design); Evidence-Based Med. Working Group, *Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine*, 268 JAMA 2420, 2420–25 (1992) (same); Sylvan B. Green & David P. Byar, *Using Observational Data from Registries to Compare Treatments: The Fallacy of Omnimetrics*, 3 STAT. MED. 361, 361 (1984) (same).

109 FRIEDMAN ET AL., *supra* note 54, at 42; Concato et al., *supra* note 104, at 1888. But see Duh et al., *supra* note 63, at 32–33 (discussing potential flaws if meta-analyses are not carefully conducted).

designed controlled trials that lack randomization.¹¹⁰ Farther down in the rankings are various study designs¹¹¹ that, collectively, are called “observational studies.”¹¹² These include individual case reports (lowest in the hierarchy of evidence) describing, for example, what happened to one patient who took a particular drug; case series that compile multiple such reports; and uncontrolled experiments that produce results too dramatic to be the product of chance (for example, early uses of penicillin to treat bacterial disease). They also include cross-sectional studies, case-control studies, cohort and registry studies,¹¹³ which enjoy intermediate positions in the hierarchy of evidence. The term “outcomes research” also is used to refer to observational studies that follow groups of patients over time to observe their health outcomes or analyze (prospectively or retrospectively) factors that may contribute to those outcomes.¹¹⁴ Evidence from large, well-designed cohort or case-control studies enjoys a fairly high ranking, just below controlled trials that lack randomization.¹¹⁵

Concato et al. credit a landmark 1982 study¹¹⁶ with bolstering the perceived superiority of RCTs.¹¹⁷ That study identified a group of observational studies that had overstated the effectiveness of drugs relative to results from RCTs. However, those observational studies had

110 See Concato et al., *supra* note 104, at 1888; *supra* note 108.

111 Concato et al., *supra* note 104, at 1888.

112 FURBERG & FURBERG, *supra* note 60, at 29.

113 For definitions of these terms, see Concato et al., *supra* note 104, at 1888; FURBERG & FURBERG, *supra* note 60, at 29–33. Cross-sectional studies examine a group of patients at a single point in time without follow-up (for example, examining long-term users of a particular drug and well-matched nonusers to measure differences in their current health condition). Retrospective case-control studies compare “cases” (people who now have a medical condition) and “controls” (people without the condition), to explore whether differences in their past drug exposures or behaviors may help explain their current health differences. Prospective cohort studies track groups of people—sometimes very large groups—forward through time, recording data about their characteristics, behaviors, exposures, and eventual health outcomes. FURBERG & FURBERG, *supra* note 60, at 29–33. Registry studies are similar to cohort studies in that they are prospective studies, often using large electronic databases, of patients diagnosed with a particular disease or who received a particular treatment, to observe their eventual outcomes over the course of time. *Id.* at 33. Qualitative methods, such as surveys of patients’ perception of their well-being after treatment, also are used in observational research. *Id.*

114 Agency for Healthcare Research & Quality, Outcomes Research Fact Sheet (2000) [hereinafter AHRQ Fact Sheet], available at <http://www.ahrq.gov/clinic/outfact.htm>.

115 Concato et al., *supra* note 104, at 1888 (citing U.S. PREVENTIVE SERVS. TASK FORCE, *supra* note 108).

116 Sacks et al., *supra* note 89, at 233–34.

117 Concato et al., *supra* note 104, at 1887.

relied on historical controls.¹¹⁸ Well-designed observational studies using contemporaneous control subjects can produce average results that are “remarkably similar” to RCTs.¹¹⁹ Still, observational studies can have problems, even when historical controls are not an issue.¹²⁰

For example, retrospective studies that inquire into past events or health conditions are subject to errors of recall, though this is not a problem in prospective observational studies.¹²¹ There can be allocation bias in observational studies that follow outcomes in patients who were treated with a particular drug, since the choice of therapy was not randomized.¹²² Patients and their doctors decide which drug to take or whether to take a drug at all, and patients with particular prognostic factors may disproportionately select that particular drug.¹²³ A related issue is indication bias, in which doctors vary the choice of drug depending on the severity of a patient’s illness.¹²⁴ This can cause the drugs used in more severe cases to appear less effective. Alternatively, there can be a “healthy patient” bias—a tendency of insured, better-educated, and healthier people to take particular drugs. The appearance, based on observational studies, that HRT reduces heart-attack risk turned out simply to reflect the fact that healthier, wealthier women (who have a lower risk of heart disease than other women) were the ones who had access to doctors and HRT.¹²⁵ Careful study design (for example, making sure observa-

118 Sacks et al., *supra* note 89; Concato et al., *supra* note 104, at 1887.

119 Concato et al., *supra* note 104, at 1887, 1890, 1892 (basing conclusion on ninety-nine studies in five treatment areas).

120 Allocation bias is an issue since patients in observational studies have not been randomly assigned to receive the treatment under study; patients and their doctors made the choice of which drug to take or whether to take a drug at all, and patients with particular prognostic factors may disproportionately have chosen a particular drug. FURBERG & FURBERG, *supra* note 60, at 15, 35. A related issue is indication bias, in which, among patients with a given disease, different drugs are prescribed depending on the severity of illness. This can cause the drugs used in more severe cases to appear less effective. *Id.* at 35. However, there also can be a countervailing “healthy patient” bias—a tendency of insured, healthier people to take particular drugs. Thus, the appearance that hormone replacement therapy (HRT) reduced heart-attack risk turned out simply to reflect the fact that healthier, wealthier women (who have a lower risk of heart disease than other women) were the ones who had access to doctors and HRT. *Id.* at 35.

121 FURBERG & FURBERG, *supra* note 60, at 35.

122 *Id.* at 15, 35.

123 *Id.*

124 *Id.* at 35.

125 *Id.* A similar bias created a false appearance that statins reduce the rate of Alzheimer’s disease; both phenomena (the use of statins and resistance to dementia) are tied to higher education and wealth. *Id.*

tional groups are well matched in terms of severity of disease, income, and other prognostic factors) can help correct these problems. However, observational studies, even when well designed, can spot associations between events and outcomes but are not a good tool for proving causation.¹²⁶ Thus, they are useful for generating hypotheses about what may be causing what,¹²⁷ but confirming these hypotheses typically requires an RCT.

FDA's 1962 evidentiary paradigm, with its heavy reliance on premarket RCTs, was influenced by these concerns, but it also can be viewed as a response to the primitive state of 1960s information technology.¹²⁸ Premarket RCTs may indeed have been the best way to assess the risk-benefit characteristics of drugs under the technological constraints of that day. "However, recent advances in information technology have . . . fuel[ed] increased interest in the question of whether alternative analytical methods might offer sufficient validity to elevate observational analysis in the hierarchy of medical knowledge."¹²⁹

Observational research flowered after 1980 in response to several stimuli. One was concern about regional disparities in healthcare practices and outcomes in the United States.¹³⁰ Another was adoption of outcome-oriented accreditation standards by the Joint Commission on Accreditation of Health Care Organizations.¹³¹ The most significant factor was improved information technology and "an era of large volumes of data on platforms conducive to analyses."¹³² Modern cohort and registry studies often rely on large administrative databases, such as claims databases maintained by health insurers, and large clinical databases.¹³³ Expanded availability and improved qual-

126 *Id.* at 36; see also Robert M. Califf, *Evolving Methods: Alternatives to Large Randomized Control Trials*, in IOM, *LEARNING HEALTHCARE*, *supra* note 95, at 84, 84–85 (discussing observational analysis of prospective clinical databases but noting, "no amount of statistical analysis can substitute for randomization . . . when comparing alternative approaches to diagnosis or treatment"); Duh et al., *supra* note 63, at 34 (discussing difficulties establishing causation in the association between HRT and breast cancer).

127 Duh et al., *supra* note 63, at 34.

128 Evans & Flockhart, *supra* note 81, at 47.

129 Califf, *supra* note 126, at 84–85.

130 AHRQ, Fact sheet, *supra* note 114.

131 Fred D. Brenneman et al., *Outcomes Research in Surgery*, 23 *WORLD J. SURGERY* 1220 (1999).

132 Califf, *supra* note 126, at 95.

133 Brenneman et al., *supra* note 131, at 1220; see also PANEL ON PERFORMANCE MEASURES & DATA FOR PUB. HEALTH PERFORMANCE PARTNERSHIP GRANTS, NAT'L RESEARCH COUNCIL, *HEALTH PERFORMANCE MEASUREMENT IN THE PUBLIC SECTOR*

ity of these databases¹³⁴ fostered new approaches that can help reduce the problems seen in historical-control studies.¹³⁵ FDAAA calls for observational studies to play an expanded role in FDA's regulatory processes.¹³⁶ This shift requires advanced information technology¹³⁷ that simply was not available when FDA was developing its 1962 evidentiary paradigm.

II. LIMITATIONS OF THE OLD EVIDENTIARY PARADIGM

If RCTs are viewed as the "gold standard,"¹³⁸ then greater use of observational studies seemingly would weaken the evidence on which FDA relies. This appearance is false. The clinical trial evidence on which FDA has been basing its decisions is far weaker than many of us understand. Much of what the public thought FDA was doing for the past several decades, as it turns out, simply could not be done. FDA's regulation of drugs was comprehensive only in the public's hopes and imaginations. Expanding FDA's reliance on other forms of data will strengthen, rather than weaken, the evidentiary basis for FDA's decisions.

A. *The Myth of the Premarket Safety Trial*

One of the most common misconceptions about U.S. drug regulation is the belief that FDA requires RCTs to confirm the safety and efficacy of new drugs. This is almost never the case. Before FDA approves a drug, the agency requires rigorous, hypothesis-driven RCTs of *efficacy*, but not of safety. "[T]rials are generally not designed for the purpose of assessing adverse effects. The scientific standards that are used in evaluating an intervention for efficacy are rarely employed when evaluating possible adverse effects."¹³⁹

83–94 (Edward B. Perrin et al. eds., 1999), available at <http://www.nap.edu/openbook.php?isbn=0309064368> (discussing administrative databases and their use in observational research); Jed Weissberg, *Use of Large System Databases*, in IOM, *LEARNING HEALTHCARE*, *supra* note 95, at 46 (describing observational research with a large HMO clinical database).

134 Pillans, *supra* note 62, at 697.

135 FRIEDMAN ET AL., *supra* note 54, at 50.

136 See discussion *infra* Part V.A.

137 U.S. FOOD & DRUG ADMIN., U.S. DEP'T OF HEALTH & HUMAN SERVS., *THE SENTINEL INITIATIVE 6–7* (2008) [hereinafter FDA, SENTINEL INITIATIVE], available at <http://www.fda.gov/downloads/Safety/FDAsSentinelInitiative/UCM124701.pdf> (noting that information technology is key to modernizing FDA's drug safety framework).

138 See *supra* notes 103–04.

139 FRIEDMAN ET AL., *supra* note 54, at 170–71.

In 1983, FDA assessed the premarket studies of various drugs that had accounted for 30% of FDA approvals between 1975 and 1981.¹⁴⁰ The studies had taken an average of 8.2 years. Phase I and II studies consumed roughly half that time and provided essentially all the information that would be discovered about adverse effects before drug approval. No additional adverse reactions were detected in phase III studies, which were aimed at assessing efficacy.¹⁴¹ In subsequent years, FDA took various steps to improve reporting of adverse events during phase III trials.¹⁴² Thus it has become common, in recent years, to see attrition of promising new drug candidates after unexpected safety problems come to light during phase III trials.¹⁴³ Phase III trials now play an important role in safety assessment even though their primary focus is, and was intended to be, efficacy. Perhaps for this reason, many people believe phase III trials are RCTs of safety as well as efficacy.

A better description of a phase III drug trial is that it is an RCT of drug efficacy, accompanied by a small observational study of adverse events within the clinical trial population. Essential to the definition of an RCT is that there is a predefined primary question to be answered.¹⁴⁴ The question often is framed as a testable hypothesis (for example, that people who receive the interventional drug will

140 Melmon, *supra* note 37, at 143 (discussing FDA's 1983 survey).

141 *Id.*

142 See 21 C.F.R. § 314.50(d)(5)(vi) (2009) (describing safety information to be included in a new drug application); see also CTR. FOR DRUG EVALUATION & RESEARCH, U.S. FOOD & DRUG ADMIN., REVIEWER GUIDANCE: CONDUCTING A CLINICAL SAFETY REVIEW OF A NEW PRODUCT APPLICATION AND PREPARING A REPORT ON THE REVIEW (2005) [hereinafter FDA, CONDUCTING A CLINICAL SAFETY REVIEW], available at <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm072974.pdf> (detailing how safety data and adverse event reports from clinical trials are used and reported during FDA's review of a new drug application).

143 See U.S. FOOD & DRUG ADMIN., U.S. DEP'T OF HEALTH & HUMAN SERVS., INNOVATION OR STAGNATION: CHALLENGE AND OPPORTUNITY ON THE CRITICAL PATH TO NEW MEDICAL PRODUCTS, at ii (2004) [hereinafter FDA, INNOVATION OR STAGNATION], available at <http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/ucm113411.pdf>; see also Robert M. Califf, *A Broad Perspective*, in ROUNDTABLE ON TRANSLATING GENOMIC-BASED RESEARCH FOR HEALTH, INST. OF MED., DIFFUSION AND USE OF GENOMIC INNOVATIONS IN HEALTH AND MEDICINE: WORKSHOP SUMMARY 3, 9 (Lyla M. Hernandez rapporteur, 2008) [hereinafter IOM, DIFFUSION & GENOMIC INNOVATIONS], available at http://books.nap.edu/openbook.php?record_id=12148 (discussing abandonment of torcetrapib during phase III trials).

144 FRIEDMAN ET AL., *supra* note 54, at 16; see also FURBERG & FURBERG, *supra* note 60, at 108 (defining "statistical power" as the ability to detect prespecified intervention effects).

have a lower heart-attack rate than those receiving the control drug). A predefined question is essential if a trial is to produce statistically significant¹⁴⁵ results. The perils of post hoc analysis of trial data were nicely demonstrated by investigators who went back and analyzed various subgroups of patients within an old clinical trial data set. They waggishly announced their “discovery” that patients born under the sun signs Libra and Gemini had a higher risk of death when taking the interventional drug, whereas other sun signs got a distinct benefit from the drug.¹⁴⁶ When a trial is not designed to study a question, apparent answers may be spurious.

In addition to its primary question, an RCT also may address a limited number of predefined secondary questions (for example, whether the intervention reduces strokes as well as heart attacks, or whether it reduces heart attacks not only in all subjects tested but also among the female subjects).¹⁴⁷ The primary question is almost always a question about drug efficacy.¹⁴⁸ The same is generally true of the secondary questions in phase III drug trials.¹⁴⁹ In many cases, it would be impossible to define specific, testable hypotheses about drug safety in advance, since a given drug may produce many different types of adverse events, many of which are unforeseeable.¹⁵⁰ Evaluating safety risks “requires knowledge, or at least tentative ideas, about what effects might occur.”¹⁵¹ Without testable safety hypotheses, there can be no RCT of safety.

Dexfenfluramine, part of the “fen-phen” combination used in treating obesity, produced reports of heart-valve abnormalities within a year of its FDA approval and eventually was withdrawn¹⁵² after millions of patients had taken it.¹⁵³ Premarket trials had not caught this problem. With no expectation of heart-valve impacts, the trials had not included echocardiography that might have detected them.¹⁵⁴

145 If results are not statistically significant, it means that the observed difference between people in the intervention and control groups could have arisen by chance. FURBERG & FURBERG, *supra* note 60, at 107.

146 *Id.* at 112.

147 FRIEDMAN ET AL., *supra* note 54, at 17.

148 FURBERG & FURBERG, *supra* note 60, at 70 (“Clinical trials are not often conducted with the primary purpose of determining treatment safety.”).

149 FRIEDMAN ET AL., *supra* note 54, at 172 (noting that adverse events are seen as secondary or even tertiary response variables).

150 *Id.* at 18, 171.

151 *Id.* at 132.

152 FURBERG & FURBERG, *supra* note 60, at 19–20.

153 Pillans, *supra* note 62, at 697 (reporting that seven million patients were exposed).

154 *Id.*

Even if FDA required RCTs of safety, it is unlikely the revealing hypothesis would have been put forward in this instance. Even when adverse effects are foreseen, the fact that one drug can have many potential risks means that premarket RCTs of safety would need to test a whole array of hypotheses (for example, hypotheses related to kidney function, liver function, heart-valve status, impact on eyesight, blood-clot risk, etc.).¹⁵⁵ RCTs that attempt to test multiple hypotheses may fail to yield statistically significant results unless they involve very large numbers of research subjects.¹⁵⁶ This is why RCTs can address only a limited number of primary and secondary questions.

Another problem concerns the ethics of safety-related RCTs. If the suspicion of a particular drug-related risk is strong enough to generate a testable hypothesis, an RCT may well be unethical.¹⁵⁷ When an important safety problem is suspected, such a trial would involve deliberately exposing test subjects to possible harm. Its ethical acceptability would depend on the expected balance of benefits and harms.¹⁵⁸ In a recent guidance document,¹⁵⁹ FDA called for phase II and III studies of type-2 diabetes drugs to examine specific safety-related variables bearing on cardiovascular risk.¹⁶⁰ The guidance recommends lengthening these trials from the typical three to six months to a minimum of two years to improve detection of such risks.¹⁶¹ It also calls for including subjects, such as elderly patients and patients with kidney impairment, who are at higher risk of experiencing cardiovascular events.¹⁶² “Until now, manufacturers had only to show that their drugs reduced blood sugar levels.”¹⁶³ That is, they only tested efficacy hypotheses.

155 FRIEDMAN ET AL., *supra* note 54, at 18.

156 *See id.* at 21, 123–25, 171, 307–08, 318; FURBERG & FURBERG, *supra* note 60, at 110.

157 FRIEDMAN ET AL., *supra* note 54, at 18; FURBERG & FURBERG, *supra* note 60, at 21.

158 *See* FRIEDMAN ET AL., *supra* note 54, at 18 (commenting that “rigorous, convincing demonstration of serious toxicity is usually not achieved because it is generally thought unethical to continue a study to the point at which a drug has been conclusively shown to be more harmful than beneficial”); *see also* 45 C.F.R. § 46.111 (2009) and related FDA regulations at 21 C.F.R. § 56.111 (2009) (requiring minimal risks to subjects and a reasonable relation of risks to anticipated benefits of the research).

159 FDA, DIABETES MELLITUS, *supra* note 72.

160 *Id.* at 3.

161 *Id.* at 4.

162 *Id.* at 3.

163 Susan J. Landers, *FDA Wants Closer Scrutiny of New Diabetes Drugs*, AM. MED. NEWS, Jan. 19, 2009, at 17, available at <http://www.ama-assn.org/amednews/2009/01/12/hlsa0112.htm>.

Ethics of this new guidance are delicate. In general, “any participant for whom the intervention is known to be harmful should not be admitted to the trial.”¹⁶⁴ In this instance, it may be justifiable to include high-risk patients and lengthen their exposure to experimental therapies, since the suspicion of cardiovascular risk is based on experience with the broader class of diabetes drugs¹⁶⁵ rather than being a particularized suspicion about the test drugs themselves. Moreover, the at-risk patients in the trial have a potential for direct, personal benefit as new therapies become available, since their disease is chronic in nature and will require ongoing treatment after the trial concludes.¹⁶⁶

As a general matter, FDA’s premarket studies do not include rigorous hypothesis-testing RCTs of drug safety. Instead, FDA requires what amount to small, premarket observational studies of adverse events among the people who participate in trials. FDA requires reporting of adverse events seen in the trial subjects.¹⁶⁷ All serious events, like deaths or hospitalizations, must be reported whether or not they are thought to be drug-related; the goal is to spot unknown or unexpected associations as well as those that were anticipated.¹⁶⁸ This inclusiveness makes sense, given the unpredictable nature of adverse events, but it generates a large flow of data, making it hard to spot associations that really may be drug-related.¹⁶⁹ Confirming causation would require an RCT. The trial population is too small to provide even high-quality observational evidence of safety.¹⁷⁰

The best available estimate of drug-related risks, at the time a drug is approved, may be the difference in reported adverse events between the control group and the intervention group in the phase III trial.¹⁷¹ Such estimates generally lack statistical significance and are subject to various confounding factors. For example, in one beta-blocker trial, 66% of the group receiving *the placebo* reported shortness of breath, even though only 6% of those people had any prior history

164 FRIEDMAN ET AL., *supra* note 54, at 35.

165 See Steven E. Nissen & Kathy Wolski, *Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death from Cardiovascular Causes*, 356 NEW ENG. J. MED. 2457, 2458 (2007) (discussing cardiovascular risks of the diabetes drug Avandia).

166 See FDA, DIABETES MELLITUS, *supra* note 72, at 2.

167 FRIEDMAN ET AL., *supra* note 54, at 180.

168 FURBERG & FURBERG, *supra* note 60, at 69.

169 *Id.* at 69–70.

170 See discussion *infra* Part II.B.1.

171 FURBERG & FURBERG, *supra* note 60, at 70.

of this problem.¹⁷² Such results can cast doubt on the reliability of events reported by the people who actually were taking the drug. Asking trial subjects questions about adverse events may bias their perceptions of whether events occurred.¹⁷³ Adverse event reports tend to be unsystematic, without consistent classifications and often lacking data on the severity (as opposed to the frequency) of events.¹⁷⁴ *Intrastudy* inconsistencies are not uncommon.¹⁷⁵ When publishing trial results, there is a tendency to downplay or even to fail to report adverse event rates.¹⁷⁶ The fact that safety findings tend to be of low evidentiary quality may contribute to this underreporting. Data on why subjects terminated their participation in a trial also are poorly reported.¹⁷⁷ Thus, it may be unknown whether people who quit a study did so because of side effects that did not rise to the level of a reportable adverse event.

B. *Methodological Limitations of Premarket Safety Studies*

Key methodological limitations of premarket safety studies include inadequate size, inadequate duration, and problems with the generalizability of results.¹⁷⁸ FDA has not concealed these limitations¹⁷⁹ but they remain poorly understood by the public, physicians, and academics.

1. Trial Size

Premarket drug trials are simply too small to detect rare adverse events, yet even rare risks can generate large numbers of casualties once a drug is marketed to millions of people.¹⁸⁰ The sample size for a clinical trial—how many subjects need to be included in the study to

172 FRIEDMAN ET AL., *supra* note 54, at 171 (citing Robert P. Byington, Beta-Blocker Heart Attack Trial Research Group, *Beta-Blocker Heart Attack Trial: Design, Methods, and Baseline Results*, 5 CONTROLLED CLINICAL TRIALS 382 (1984)).

173 *Id.* at 172–76 (discussing difficulties in ascertaining the rate of adverse events in a trial).

174 NAT'L CANCER POLICY FORUM, INST. OF MED., DEVELOPING BIOMARKER-BASED TOOLS FOR CANCER SCREENING, DIAGNOSIS, AND TREATMENT: WORKSHOP SUMMARY 54 (Margie Patlak & Sharyl Nass rapporteurs, 2006) [hereinafter IOM, BIOMARKER-BASED TOOLS], available at http://books.nap.edu/openbook.php?record_id=11768 (reporting presentation of Dr. Naomi Aronson).

175 FRIEDMAN ET AL., *supra* note 54, at 172.

176 *Id.* at 171; see also FURBERG & FURBERG, *supra* note 60, at 70 (discussing studies that show underreporting of adverse events in publications resulting from trials).

177 FURBERG & FURBERG, *supra* note 60, at 70.

178 See FRIEDMAN ET AL., *supra* note 54, at 182; Pillans, *supra* note 62, at 700.

179 Melmon, *supra* note 37, at 142.

180 IOM, FUTURE OF DRUG SAFETY, *supra* note 23, at 37–38.

produce statistically significant results—is based on the primary question¹⁸¹ which, in a phase III drug trial, is a question about drug efficacy. It typically requires a larger sample size to answer questions about safety than it takes to answer questions about a drug's efficacy. Thus premarket trials, which are designed to provide statistically significant results about efficacy, may fail to shed much light on safety.

The required sample size for a clinical trial depends on several factors. One of the most important factors is the underlying event rate.¹⁸² Suppose the primary question is whether the interventional drug reduces the rate of heart attacks. The event rate is how often people in the study would be expected to have a heart attack under ordinary circumstances (that is, in the absence of taking the interventional drug). It requires a larger sample size to study a rare event than a common one, since the question can only be studied as the event actually occurs.¹⁸³ The sample size also depends on how dramatically the interventional drug is expected to change the underlying event rate.¹⁸⁴ Fewer test subjects are needed if the interventional drug is expected to cut heart-attack rates in half, than if it is expected to cut heart-attack rates by 5%. This is simple common sense: it is harder—it requires more subjects—to be sure of a minor change than a dramatic one.

Sample size is extremely sensitive to these two variables: the underlying event rate and the expected change caused by taking the interventional drug. Here is the problem with safety evaluations: the underlying event rate for the primary efficacy question in a trial (for example, the heart-attack rate) may be quite different from the rate of possible drug-related adverse events (for example, drug-related kidney failure). Study subjects likely would be chosen to be in the trial because of their high risk of heart disease and might, for example, have an expected heart-attack rate of 20% (1-in-5) over the duration of the trial. The potential rate of drug-related kidney failure might be much lower, for example 1-in-10,000 or even 1-in-100,000. One or two thousand subjects may well be enough to give statistically significant answers to the efficacy question. If an adverse event has a 1-in-10,000 incidence rate, around 30,000 interventional subjects would be needed to have a 95% probability of spotting even one case of it;

181 FRIEDMAN ET AL., *supra* note 54, at 16; *see also id.* at 174–75 (“[I]nvestigators calculate sample sizes on the basis of primary response variables, not as a result of estimates of adverse effect frequency.”).

182 *Id.* at 103 fig. 7-2.

183 FURBERG & FURBERG, *supra* note 60, at 108.

184 *Id.*

65,000 would be needed to spot three cases.¹⁸⁵ Including control subjects, a safety trial would need to be much larger than the usual phase III trial, which typically includes 600–3000 subjects.¹⁸⁶ To detect the 1-in-20,000 chance of serious liver toxicity in bromfenac, an NSAID that stayed on the market for less than a year in 1997–1998, an estimated 60,000 subjects would need to have been exposed to the drug in trials.¹⁸⁷ By one estimate, three million trial subjects would have been required to detect the risk of aplastic anemia with the antibacterial drug chloramphenicol.¹⁸⁸ “Even if assessment of adverse effects is a major objective of the study, a trial’s size will not generally be increased in an effort to improve the likelihood of reliably detecting such effects.”¹⁸⁹

2. Trial Duration

Clinical trials are poor instruments for detecting long-term risks.¹⁹⁰ Phase III trials typically last one to four years and may include 1000 to 10,000 patients of whom only a few hundred patients typically receive the new drug for more than three to six months.¹⁹¹ “[O]nly the most profound and overt risks and side effects that occur immediately after taking a drug can be detected” (and then only if the side effect occurs fairly frequently, such as once in one hundred times the drug is taken).¹⁹² FDA increasingly is required to assess products intended for long-term use as America’s disease burden shifts from acute, infectious diseases toward chronic health conditions. Tens of millions of Americans regularly consume drugs for chronic conditions like depression, arthritis, diabetes, high cholesterol, and osteoporosis.¹⁹³ Yet very few clinical trials last more than four or five

185 Melmon, *supra* note 37, at 142; accord FURBERG & FURBERG, *supra* note 60, at 17.

186 IOM, FUTURE OF DRUG SAFETY, *supra* note 23, at 36.

187 *Id.* at 38.

188 Melmon, *supra* note 37, at 142; see also The Merck Manuals Online Medical Library, Chloramphenicol, <http://www.merck.com/mmpe/sec14/ch170/ch170d.html> (last visited Oct. 16, 2009) (characterizing the risk of irreversible idiosyncratic aplastic anemia as less than 1:25,000 and noting that due to these effects, chloramphenicol is no longer the drug of choice except in serious infections caused by a few multidrug-resistant pathogens that still respond to this drug).

189 FRIEDMAN ET AL., *supra* note 54, at 175; FURBERG & FURBERG, *supra* note 60, at 18.

190 FURBERG & FURBERG, *supra* note 60, at 19.

191 FRIEDMAN ET AL., *supra* note 54, at 181; IOM, PREVENTING MEDICATION ERRORS, *supra* note 97, at 56.

192 IOM, PREVENTING MEDICATION ERRORS, *supra* note 97, at 56.

193 FURBERG & FURBERG, *supra* note 60, at 19; Duh et al., *supra* note 63, at 31.

years.¹⁹⁴ A short-duration trial obviously cannot assess risks and benefits, thirty or forty years hence, of giving cholesterol-lowering statins to eight-year-olds, as recommended in a recent set of clinical guidelines.¹⁹⁵ Assessing safety via RCT would require a multi-decade trial.

Americans have never been patient about long regulatory delays in getting new medical products approved. The average time from drug creation to FDA approval was two-and-a-half years before the 1962 amendments.¹⁹⁶ By 1980, the lag had increased to seven to thirteen years¹⁹⁷ and a series of governmental studies attributed the delay largely to the 1962 amendments.¹⁹⁸ In 1984, Congress passed the Hatch-Waxman Amendments¹⁹⁹ to mitigate the impact these long delays were having on drug patent life. Between 1987 and 1992, FDA implemented several policies aimed at reducing delays that were allegedly hindering access to promising new drugs for AIDS and cancer.²⁰⁰ Recent years have seen patients pressing federal courts for relief when FDA's clinical trial process delays access to desired, unapproved therapies.²⁰¹ Clinical trials are too short to answer questions about long-term risks, but they already seem too long to patients desperate for new therapeutic options.²⁰²

194 FURBERG & FURBERG, *supra* note 60, at 19.

195 Alan Zarembo, *Use of Statins in Children Debated*, L.A. TIMES, July 9, 2008, at A13.

196 Lechter, *supra* note 90, at 157–58.

197 Roberts & Bodenheimer, *supra* note 14, at 586 (citing H. SUBCOMM. ON SCI., RESEARCH & TECH. OF THE COMM. ON SCI. & TECH., 96TH CONG., REPORT ON THE FOOD AND DRUG ADMINISTRATION PROCESS FOR APPROVING NEW DRUGS 3 (Comm. Print 1980)).

198 See Lechter, *supra* note 90, at 158; Roberts & Bodenheimer, *supra* note 14, at 589–96.

199 Drug Price Competition and Patent Term Restoration (Hatch-Waxman) Act, Pub. L. No. 98-417, 98 Stat. 1585 (1984).

200 21 C.F.R. § 312.34 (2009) (addressing the treatment IND mechanism); New Drug, Antibiotic, and Biological Drug Product Regulations; Accelerated Approval, 57 Fed. Reg. 58,942 (Dec. 11, 1992); Expanded Availability of Investigational New Drugs Through a Parallel Track Mechanism for People With AIDS and Other HIV-Related Disease, 57 Fed. Reg. 13,250 (Apr. 15, 1992).

201 See, e.g., *Abigail Alliance for Better Access to Developmental Drugs v. von Eschenbach*, 495 F.3d 695 (D.C. Cir. 2007); *Gunvalson v. PTC Therapeutics, Inc.*, No. 08-3559, 2008 WL 4003377 (D.N.J. Aug. 21, 2008).

202 See generally IOM, *FUTURE OF DRUG SAFETY*, *supra* note 23, at 122 (discussing the view that patients, particularly those with fatal diseases, should not be hindered in accessing drugs they wish to try).

3. Generalizability

Clinical trial results often are not generalizable to the larger population that will consume a drug once it is approved.²⁰³ Real patients may be younger, older, healthier, sicker, or more or less at risk than the carefully chosen trial subjects on whom new drugs are tested. If a treatment depends on the skill with which it is administered (for example, an implantable medical device where outcomes vary depending on the surgeon's skill) or on follow-up care (such as close monitoring to detect adverse events before they inflict serious injury), clinical trials may give unduly rosy assessments of its performance in routine clinical settings.

David Eddy points out that all study designs, including RCTs, have biases; the biases are simply different for different study designs.²⁰⁴ RCTs reduce biases to the *internal* validity of a study—biases that might make the interventional drug look better than it actually is in the setting of the clinical trial.²⁰⁵ Internal biases include the allocation and ascertainment biases discussed earlier, which randomization and blinding help eliminate.²⁰⁶ In contrast, generalizability goes to the *external* validity of study results—whether conditions in the trial are biased relative to the external clinical setting where patients, medical techniques, and other factors may be different.²⁰⁷ Certain observational study designs, such as large cohort or case-control studies that sample data from a large number of patients in various healthcare settings, are superior to RCTs in avoiding biases to external validity.²⁰⁸ Large prospective registries that track people who take a drug once it is approved are more representative of the real population than limited-eligibility RCTs.²⁰⁹ The notion that RCTs have fewer biases than observational studies is inaccurate; they reduce *internal* biases.

Clinical trials admit subjects who are, on average, less likely to suffer adverse events than other people.²¹⁰ High-risk patients are deliberately excluded from trials²¹¹ both for commercial reasons (to

203 FRIEDMAN ET AL., *supra* note 54, at 37.

204 David M. Eddy, *Should We Change the Rules for Evaluating Medical Technologies?*, in IOM, MODERN METHODS, *supra* note 37, at 117, 124–26.

205 *Id.* at 124.

206 See discussion *supra* Part I.

207 Eddy, *supra* note 204, at 124.

208 *Id.* at 125.

209 FRIEDMAN ET AL., *supra* note 54, at 50.

210 *Id.* at 37; see also *id.* at 181 (“People most likely to develop adverse effects are generally excluded from clinical trials.”).

211 *Id.*

make the interventional drug look good) and for ethical reasons (to minimize risks to research subjects). A Finnish study looked at 400 patients hospitalized for gastric ulcer and asked which of the patients would have met eligibility criteria to participate in a typical prospective drug trial.²¹² Seventy-one per cent of the highest-risk patients (defined as those who, over a five to seven year period, eventually died or experienced serious complications from their ulcers) would have been excluded from the typical clinical drug trial.²¹³ FDA's recent guidance calling for higher-risk subjects to be included in diabetes drug trials²¹⁴ is an exception to the usual rule, which is that drugs are approved based on small observational safety studies (clinical trial adverse event statistics) in a carefully selected population of *low-risk patients*.

A related point is that FDA conceives drug safety²¹⁵ as a benefit-risk ratio: a drug is "safe" if its benefits compare favorably to its risks.²¹⁶ If the trial subjects are more likely than other patients to experience *benefits* from taking the drug, then its benefit-risk ratio will be overstated and so will be its safety.²¹⁷ Suppose a drug is expected to reduce rate of heart attacks by 20%. The trial eligibility criteria

212 See I. Kääriäinen et al., *What Fraction of Hospital Ulcer Patients Is Eligible for Prospective Drug Trials?*, 186 SCANDINAVIAN J. GASTROENTEROLOGY 73 (1991).

213 *Id.* at 75.

214 FDA, DIABETES MELLITUS, *supra* note 72; see discussion *supra* Part II.A.

215 See FDCA § 505(d), 21 U.S.C. § 355(d) (2006) (requiring a drug to be "safe for use under the conditions prescribed, recommended, or suggested in the proposed labeling thereof").

216 D. Throckmorton, Acting Deputy Dir., Ctr. for Drug Evaluation & Research, Presentation, *Efficacy Biomarkers: Efficacy/Risk Assessment*, at slide 4 (Oct. 6, 2005) (on file with author); see also CTR. FOR DRUG EVALUATION & RESEARCH, FOOD & DRUG ADMIN., GUIDANCE FOR INDUSTRY: E2E PHARMACOVIGILANCE PLANNING 2 (2005), available at <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073107.pdf> ("The decision to approve a drug is based on its having a satisfactory balance of benefits and risks within the conditions specified in the product labeling."); CTR. FOR DRUG EVALUATION & RESEARCH, U.S. FOOD & DRUG ADMIN., MANUAL OF POLICIES AND PROCEDURES 6010.3, at § 1 (2004), [hereinafter CDER, MAPP 6010.3] available at <http://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/StaffPoliciesandProcedures/ucm080121.pdf>; INT'L CONFERENCE ON HARMONISATION OF TECHNICAL REQUIREMENTS FOR REGISTRATION OF PHARM. FOR HUMAN USE, DOC. NO. ICH E3, GUIDELINE FOR INDUSTRY: STRUCTURE AND CONTENT OF CLINICAL STUDY REPORTS 35 (1996), available at <http://www.emea.europa.eu/pdfs/human/ich/013795en.pdf> (stating that "the relationship of risks and benefits should be briefly summarized and discussed" when making conclusions about a drug's efficacy and safety).

217 See David M. Kent & Rodney A. Hayward, *Limitations of Applying Summary Results of Clinical Trials to Individual Patients: The Need for Risk Stratification*, 298 JAMA 1209, 1209 (2007) ("There is growing awareness that the results of randomized clinical tri-

select subjects who are at high risk of heart attack (for example, a 40% annual risk of suffering a heart attack). For these people, the drug offers an eight-percentage-point reduction (20% x 40%) in heart-attack risk. This eight-point reduction is a large benefit and compares favorably to the drug's small risk of adverse events, so the drug is deemed safe. It may not be safe for patients at lower risk of having a heart attack. A patient with a 5% heart-attack risk can only expect to get a one-point reduction in that risk (20% x 5%). This one-point benefit may not compare favorably to the drug's risks. Drug safety depends not only on a patient's level of risk for adverse effects, but on the patient's level of risk for the condition the drug is intended to treat.²¹⁸ Trial subjects tend to differ from other patients in both these risk factors.

The generalizability of trial results can be improved by making certain changes to the trial design.²¹⁹ However, when the purpose of a trial is to generate data to support an application for FDA approval, there are limits to how much generalizability can be attained. Exclusion criteria for a phase III drug trial serve various objectives of which generalizability is only one. Other objectives include complying with FDA's ethical requirement that risks to research subjects be minimized,²²⁰ and selecting subjects in whom efficacy can be tested clearly and expeditiously. As a result, phase III drug trials are susceptible to problems with generalizability.

III. THE OLD PARADIGM: REPAIR OR REPLACE?

Even today, faith in the old paradigm runs deep. There is a tendency to assume that more evidence development means more *premarket* evidence development—as if that were the only methodological possibility.²²¹ In response to drug safety problems one often hears calls for more premarket clinical trials, longer trials, and tighter regulatory scrutiny of trial data before new products go on the market—

als might not apply in a straightforward way to individual patients, even those within the trial.”).

218 *Id.* at 1209.

219 See FRIEDMAN ET AL., *supra* note 54, at 19–20 (discussing large simple trials in which characterization of research subjects at entry is done according to broader, less detailed criteria); Califf, *supra* note 126, at 85–86 (discussing practical clinical trials).
220 21 C.F.R. § 56.111 (2009).

221 See, e.g., Epstein, *supra* note 43, at 3 (discussing a “call for further information (and further delay) in the FDA process as a precondition for federal preemption”). Generating further information entails delay only under a presumption that the information is generated prior to approval. Postmarket evidence development would not delay consumer access to new products.

that is, to keep the same paradigm but apply it more thoroughly. Conducting more and longer clinical trials might be the solution, if the problem were simply a lack of trial data. Two cases that may fit this description are: (1) off-label use of drugs,²²² and (2) laboratory-developed tests (LDTs) that are not subject to FDA's device clearance or PMA processes.²²³ In both cases, there are genuine gaps in data and it is at least plausible to call for additional premarket trials as a way to improve patient safety (although, frankly, it is not clear off-label use injures patients at a rate disproportionate to its share of overall prescriptions written,²²⁴ nor are there data substantiating that

222 Pillans, *supra* note 62, at 700 (noting lack of trial data on off-label use). Off-label use involves prescribing FDA-approved products for novel indications (or to new patient subpopulations) that were not tested in the trials that served as the basis for product labeling and approval. See FDA, SENTINEL INITIATIVE, *supra* note 137, at 5; IOM, FUTURE OF DRUG SAFETY, *supra* note 23, at 39, 122; see also Rebecca Dresser, *The Curious Case of Off-Label Use*, HASTINGS CTR. REP., May/June 2007, at 9, 9 (providing an overview and statistics on off-label use); Barbara J. Evans, *What Will It Take to Reap the Clinical Benefits of Pharmacogenomics?*, 61 FOOD & DRUG L.J. 753, 783–85 (2006) (discussing special problems of off-label use of drugs and diagnostic tests designed for use together); Margaret Z. Johns, *Informed Consent: Requiring Doctors to Disclose Off-Label Prescriptions and Conflicts of Interest*, 58 HASTINGS L.J. 967, 969–71 (2007) (discussing risks and calling for informed consent when drugs are used off-label); David C. Radley et al., *Off-Label Prescribing Among Office-Based Physicians*, 166 ARCHIVE INTERNAL MED. 1021 (2006) (providing empirical data on off-label prescribing).

223 Naomi Aronson, *Assessing Technology for Use in Health and Medicine*, in IOM, DIFFUSION & GENOMIC INNOVATIONS, *supra* note 143, at 29, 30–31 (noting that clinical trials are not the norm in evaluating diagnostic tests). Laboratory-developed tests (formerly called “home-brew” tests) are regulated under the Clinical Laboratory Improvement Amendments of 1988 (CLIA), Pub. L. No. 100-578, 102 Stat. 2903 (codified as amended at 42 U.S.C. § 263(a) (2006)); see also 42 C.F.R. pt. 493 (1995) (CLIA regulations). Most LDTs are not subject to FDA regulation and, though they receive some internal validation by the laboratories that offer them, are not required to submit data on safety and effectiveness for prior review by an external regulator. See SECRETARY'S ADVISORY COMM. ON GENETIC TESTING, NAT'L INSTS. OF HEALTH, ENHANCING THE OVERSIGHT OF GENETIC TESTS 10–11 (2000) [hereinafter SACGT, ENHANCING OVERSIGHT], available at http://www4.od.nih.gov/oba/sacgt/reports/oversight_report.pdf; SECRETARY'S ADVISORY COMM. ON GENETICS, HEALTH, & SOC'Y, U.S. DEP'T OF HEALTH & HUMAN SERVS., U.S. SYSTEM OF OVERSIGHT OF GENETIC TESTING 65–67 (2008) [hereinafter SACGHS, U.S. OVERSIGHT], available at http://oba.od.nih.gov/oba/SACGHS/reports/SACGHS_oversight_report.pdf; TASK FORCE ON GENETIC TESTING, JOINT NAT'L INSTS. OF HEALTH-DEP'T OF ENERGY WORKING GROUP ON ETHICAL, LEGAL, & SOCIAL IMPLICATIONS OF HUMAN GENOME RESEARCH, PROMOTING SAFE AND EFFECTIVE GENETIC TESTING IN THE UNITED STATES, at ch. 2 (Neil A. Holtzman & Michael S. Watson eds., 1997) [hereinafter NIH-DOE, SAFE GENE TESTING], available at <http://www.genome.gov/10001733>.

224 Concern about off-label use ultimately rests on a presumption that on-label uses (which have been tested in clinical trials) offer a superior risk-benefit ratio to off-label uses (which have not been so tested). However, a 2003 study found that the risk-

LDTs are causing diagnostic errors or patient injuries).²²⁵ Another example was FDA's recent guidance on diabetes drug trials.²²⁶ In that case, there was a specific, identifiable gap in trial data that could be addressed through relatively small changes to the size, length, and exclusion criteria of clinical trials; these changes appear to be ethically justifiable. Beyond special cases such as these, it is problematic to require more, bigger, and longer premarket trials.

A new meningococcal conjugate vaccine, approved by FDA in 2005, suggests how big trials might have to be to detect safety problems. Given the devastating consequences of contracting meningitis, which can cause death or serious and permanent disability, vaccination of all adolescents was recommended and 5.7 million doses of the new vaccine were distributed during the first fifteen months after approval.²²⁷ In that same period, there were fifteen spontaneous reports of Guillain-Barré syndrome, which can cause paralysis or

benefit ratio for untested therapies entering phase I oncology studies was not clearly worse than the risk-benefit ratios of many FDA-approved chemotherapeutic agents. These wholly untested chemotherapies, in many instances, were just as safe and effective as FDA-approved therapies that had successfully passed through all three phases of clinical trials. Manish Agrawal & Ezekiel J. Emanuel, *Ethics of Phase I Oncology Studies: Reexamining the Arguments and Data*, 290 JAMA 1075, 1077 (2003). A recent study estimated that 21% of the 725 million annual prescriptions written in the United States were for off-label use. IOM, FUTURE OF DRUG SAFETY, *supra* note 23, at 39. A 2006 study found just over 20% of prescriptions were off-label with statistics varying significantly depending on the class of drug. See U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-08-835, PRESCRIPTION DRUGS: FDA'S OVERSIGHT OF THE PROMOTION OF DRUGS FOR OFF-LABEL USES 2 (2008) (citing Radley et al., *supra* note 222). Only 15% of off-label uses actually lack any supporting data. Radley et al., *supra* note 222, at 1023. For all uses of FDA-approved drugs (on- and off-label), it is estimated that only 60% of prescriptions written deliver their desired therapeutic benefits. Tim Peakman & Steve Arlington, *Putting the Code to Work: The Promise of Pharmacogenetics and Pharmacogenomics*, 2 DRUG DISCOVERY WORLD 35, 36 (2001) (stating a weighted average across various classes of drugs). Seven percent produce serious drug-related injuries or death. Jason Lazarou et al., *Incidence of Adverse Drug Reactions in Hospitalized Patients: A Meta-Analysis of Prospective Studies*, 279 JAMA 1200, 1202 (1998). The remaining third (that is the total drugs prescribed, minus the 60% that produce benefits, minus the 7% that produce harms) reflect non-response. There are no comparable statistics relating specifically to unsubstantiated off-label uses, so it is not known whether off-label uses account for a disproportionate share of overall drug-related injuries and deaths.

²²⁵ SACGHS, U.S. OVERSIGHT, *supra* note 223, at 32. Note, however, that CLIA does not require reporting of adverse events with LDTs. Absence of reports may imply a lack of adverse events, or it may simply mean that events are not being reported. *Id.*

²²⁶ FDA, DIABETES MELLITUS, *supra* note 72.

²²⁷ U.S. Food & Drug Admin., U.S. Dep't of Health & Human Servs., Proceedings, Sentinel Network Public Meeting 67-68 (Mar. 7, 2007) (statement of Dr. Richard

death.²²⁸ In 2007, FDA heard testimony that available postmarket databases were simply too small to answer whether there may be a causal connection between the vaccine and Guillain-Barré syndrome. Those databases allowed study of 100,000 patients who had taken the vaccine after its approval. In that 100,000-person population, not a single case of Guillain-Barré had been observed.²²⁹ Yet, obviously, 100,000 interventional subjects is a larger population than could have been included in a premarket drug trial.

A clinical trial that administered 100,000 doses presumably would have had to involve twice that many participants, some receiving the new product and some serving as controls. Enrolling 200,000 people in a clinical trial would entail major costs and delays. More problematic, it might require a whole new ethic of compulsory research participation, since this level of trial enrollment may be unachievable under current ethical norms of voluntary consent. Even if 200,000 participants could have been recruited, the trial might not have detected any instances of Guillain-Barré; had it done so, it likely would have only detected one or two events—too few to suggest whether the Guillain-Barré was vaccine-related or just a background occurrence of this very rare disorder.²³⁰ Even when investigators are “fortunate” enough to observe a serious adverse event, they may dismiss it as a fluke.²³¹ Both the scale and the ethics of today’s clinical trials leave them inherently unable to wield rare risks. “Generally, drugs cannot confidently be linked to safety problems until they have been tested in tens of thousands to hundreds of thousands of people.”²³² In practical terms, this means many risks will be detected only in the postmarket period.²³³

Extending the *length* of phase III trials also poses problems, other than the obvious one of delaying access to new therapies. FDA considered lengthening phase III trials after its 1983 survey.²³⁴ Data showed premarket studies were exaggerating adverse events associated with

Platt) [hereinafter FDA, March 7 Proceedings], available at <http://www.fda.gov/downloads/Safety/FDAsSentinelInitiative/ucm116513.pdf>.

228 *Id.*

229 *Id.*

230 IOM, FUTURE OF DRUG SAFETY, *supra* note 23, at 106 (noting that, during clinical trials, “rare events may not surface at all [and] if they do, it is at a rate so low that one cannot distinguish a drug-caused event from one expected by chance”).

231 See FRIEDMAN ET AL., *supra* note 54, at 174.

232 IOM, EMERGING SAFETY SCIENCE, *supra* note 80, at 2.

233 See IOM, UNDERSTANDING BENEFITS, *supra* note 16, at 12 (reporting presentation of Dr. Steven Galson) (noting that safety knowledge has “exponential growth” in the postmarket period when sample size increases and more data becomes available).

234 See Melmon, *supra* note 37, at 143 (discussing FDA’s 1983 survey).

short-term drug use and underestimating risks of long-term use.²³⁵ FDA concluded that phase III trials should be lengthened. Melmon has noted that “this conclusion was invalidated by their own data.”²³⁶ FDA’s data showed that phase III trials suffer high drop-out rates, above 30% in trials lasting more than twelve to eighteen months. The people most likely to drop out of a study, naturally, would tend to be those experiencing unpleasant side effects. Extending a trial could have the effect of winnowing out the subjects in whom adverse events are most likely to appear.²³⁷ Ironically, long clinical trials may give false comfort about safety by assessing long-term risks in the subgroup of subjects most resistant to adverse effects. Based on this, Melmon recommended postmarket surveillance (rather than longer premarket trials) employing a combination of methodologies that would include observational studies.²³⁸ This is the approach FDAAA embraced²³⁹ after many years’ delay.

Lengthening trials to confirm safety also raises ethical concerns. Is it ethical to continue a clinical trial to gather additional safety observations, past the point when efficacy has been shown?²⁴⁰ Even at the start of a trial, when it is not known whether the intervention or control is more effective, many clinicians regard it as ethically problematic to assign trial subjects at random to receive the control drug.²⁴¹ Once there is clear evidence that the interventional drug is more effective than the control, it would be highly questionable to continue

235 *Id.*

236 *Id.* at 144.

237 *Id.*

238 *Id.*; see also JOINT COMM’N ON PRESCRIPTION DRUG USE, *supra* note 40 (making similar recommendations in 1980).

239 See discussion *infra* Part V.A.

240 This discussion presumes that efficacy has been proved based on clinically meaningful endpoints. New drugs often are approved based on surrogate endpoints for efficacy. For example, a drug might be shown to reduce cholesterol readings (a surrogate endpoint) without showing that it actually improves clinically meaningful health outcomes (such as the number of heart attacks, strokes, or deaths). See FURBERG & FURBERG, *supra* note 60, at 61–64. When efficacy is established relative to a surrogate endpoint, there are ongoing questions about the drug’s true efficacy. See *id.* at 63–64 (discussing an osteoporosis drug that improved bone density (the surrogate endpoint), but actually caused a significantly higher risk of fractures (the true endpoint) when taken for three years). Such questions could justify extending the clinical trial to confirm efficacy, in which case it would be ethically acceptable to continue studying safety while efficacy was being confirmed.

241 FRIEDMAN ET AL., *supra* note 54, at 48.

subjecting control subjects to the less-effective control drug.²⁴² A clear finding of effectiveness typically is grounds to halt a trial or convert all subjects over to the more effective treatment. Control subjects likely would drop out of the study if informed of the efficacy difference,²⁴³ yet failing to inform them would violate informed consent. Once control subjects are switched over to the interventional drug (or if they drop out after being told about the efficacy difference), it no longer would be possible to gather comparative statistics on adverse events in the intervention versus control groups. The resulting safety data would be equivalent to a small, uncontrolled observational study of subjects taking the interventional drug. Even if informed control subjects did not drop out of the trial, subsequent safety comparisons would be of dubious quality since their adverse event reports no longer would be double-blind.

There are two main alternatives for studying the rare risks that remain undetected at the point when a phase III trial has established efficacy. The first approach, just discussed, is to conduct further clinical trials (either by delaying approval and extending phase III premarket trials, or as a phase IV clinical trial after the product is approved). The second approach is to go ahead and approve the drug and conduct postmarket surveillance (observational studies of adverse events in patients who actually take the drug). FDAAA calls for the creation of a large health data network to support this latter approach.²⁴⁴

An example suggests the power of this new approach. Prior to FDAAA, FDA had been using a very weak form of postmarket observational study.²⁴⁵ This approach, which will continue in use after FDAAA but be supplemented with additional methods, relies on largely voluntary²⁴⁶ mechanisms²⁴⁷ for reporting new safety problems

242 FURBERG & FURBERG, *supra* note 60, at 29 (discussing the view that withholding a “proven beneficial” intervention may violate the research ethical standards of the Declaration of Helsinki).

243 John E. Wennberg, *What is Outcomes Research?*, in IOM, *MODERN METHODS*, *supra* note 37, at 33, 39 (noting “the unwillingness of informed patients to accept randomization”); *see also* Weisman et al., *supra* note 95, at 132 (commenting on subjects’ unwillingness to accept randomization as causing problems with enrollment in postmarket drug trials which involve clinically available drugs).

244 *See infra* Part V.A.2.

245 *See* IOM, *EMERGING SAFETY SCIENCE*, *supra* note 80, at 74–75 (reporting presentation of Dr. June S. Almenoff); Duh et al., *supra* note 63, at 31; Pillans, *supra* note 62, at 696–98 (discussing FDA’s pre-FDAAA postmarket safety reporting systems and weaknesses of the data they provide).

246 Wood et al., *supra* note 64, at 1851.

247 21 C.F.R. § 314.80–81 (2009).

seen in patients taking approved drugs. This approach is estimated to detect only 1–10% of all adverse drug reactions²⁴⁸ and may detect them with long delays. Vioxx was approved in May 1999 and was voluntarily removed from the market by its manufacturer sixty-five months later, in September 2004.²⁴⁹ In 2007, FDA heard testimony that the Vioxx problem could have been detected much sooner if more systematic surveillance had been conducted using modern database technology.²⁵⁰ Observational studies in an insurance claims database (which included seven million persons, some of whom happened to be taking Vioxx) showed that it would have been possible to detect the drug's cardiovascular risks thirty-four months after sales began. With claims data for 100 million people, the problem could have been spotted in fewer than three months.²⁵¹ If FDA had had the necessary data networks in place to do large-scale observational studies in 1999, all of the people killed or injured by Cox-2 painkillers after August 1999 (i.e., three months after Vioxx went on sale) might have been spared. This is much faster than a phase III or phase IV clinical trial—even a very large one—could have detected the risk.

Relying on postmarket observational studies to detect rare risks may be efficient, but is it ethical? It entails approving drugs after ordinary phase III trials, knowing that patients who subsequently take the drug may be exposed to rare, as-yet-undetected risks. That seems wrong, until one considers the alternative, which is to extend phase III trials in hope of detecting these risks before products go on the market. Either way, human beings are going to be exposed to rare, undetected risks. It is merely a question of who will be exposed: will it be patients (who have voluntarily chosen to be treated with an effective drug in the expectation of receiving personal therapeutic benefits) or will it be test subjects (who altruistically expose themselves to risks to promote the safety of future patients other than themselves)? Relying on postmarket observational studies actually seems the more ethical choice since it internalizes the remaining, unknown risks to the peo-

248 Frontline Interview with Paul Seligman, Dir., Office for Drug Safety, Ctr. for Drug Evaluation & Research (PBS television broadcast Nov. 4, 2002), *transcript available at* <http://www.pbs.org/wgbh/pages/frontline/shows/prescription/interviews/seligman.html> (citing the 1–10% figure for all adverse events (serious and nonserious), but noting that the system was estimated to detect a higher percentage (between one-third and one-half) of all serious adverse events).

249 Jennifer Wolsing, *The Vioxx Litigation: Disincentivizing Patient Safety Through Misdirected Tort Rules*, 75 DEF. COUNS. J. 209, 210–11 (2008).

250 FDA, March 7 Proceedings, *supra* note 227, at 70 (statement of Dr. Richard Platt).

251 *Id.*

ple who avail themselves of a drug's benefits. Today's phase III trials can establish efficacy and some basic level of safety, but they may not detect risks that are rare, unexpected, or slow to emerge. Calls for extended premarket safety testing of drugs to detect those latter risks are endorsing an interpersonal tradeoff that subordinates the safety of research subjects to the safety of patients.

Practically and ethically, there is no way to change premarket trials that will make them as informative as the public wishes they would be. The notion that a drug can be declared "safe and effective" based on small, short-duration premarket studies was a conceit, but it was a comforting conceit to which the public willingly subscribed after 1962. It is beyond the scope here to explore broader impacts this conceit has had on popular and legal culture since it found expression in the 1962 Drug Amendments. For example, the expansion of medical product liability after 1962 owes much to this conceit.²⁵² Treating late-emerging risks as blameworthy reflects a popular expectation that risks should have been discovered before the product went on sale²⁵³—an unreasonable expectation in many cases, once the limitations of premarket studies are understood. Conversely, the notion that an FDA approval should preempt state tort actions²⁵⁴ seems dubious once the frailty of FDA's 1962 evidentiary paradigm is taken into account. Drug safety problems earlier this decade laid bare that the conceit was a conceit. According to current thinking, "[t]he benefit-risk profiles of pharmaceuticals are constantly evolving as new data are collected throughout the life cycle of a drug."²⁵⁵ How this view will affect our broader legal culture is still unknown, but its impact on medical product regulation is clear: FDA's 1962 evidentiary paradigm,

252 It should be noted that the vast expansion of drug-injury liability after 1960 reflects other factors as well, such as publication of the *Second Restatement of Torts* in 1965. Epstein, *supra* note 43, at 5. See *id.* at 5–7 for discussion of additional contributing factors.

253 See, e.g., Rabin, *supra* note 5, at 304 ("In some cases, drugs have side effects that were unanticipated by the agency when they were approved for sale, even if they should have been anticipated by the drug manufacturer.").

254 For discussions favoring preemptive effect of FDA's drug approval and labeling decisions, see *Wyeth v. Levine*, 129 S. Ct. 1187, 1219–21 (2009) (Alito, J., dissenting) and Requirements on Content and Format of Labeling for Human Prescription Drug and Biological Products, 71 Fed. Reg. 3922, 3933–36 (Jan. 24, 2006) (discussing, in the preamble to a proposed regulation, whether drug labeling decisions by FDA preempt state failure-to-warn claims); see also *Riegel v. Medtronic, Inc.*, 128 S. Ct. 999, 1006–07 (2008) (holding, in a medical device case, that state personal injury action is preempted for devices subject to FDA's premarket approval process, but not for devices subject to 510(k) clearance).

255 IOM, UNDERSTANDING BENEFITS, *supra* note 16, at 1.

with its heavy reliance on data from premarket trials, is broken beyond repair. It needs to be replaced, and that is what FDAAA sets out to do.

IV. THE CHALLENGES OF GENOMIC MEDICINE

The recent drug safety problems, though serious, might not have given impetus to fundamental reforms had there not been an additional factor at work. By the middle of this decade, the Human Genome Project was starting to bear fruit. Genomic discoveries were being translated into clinically useful products at a mere trickle²⁵⁶ but these early examples made clear that future therapies may look quite different from those of the past.²⁵⁷ Without a new evidentiary paradigm, FDA soon would find itself “using 20th-century tools to evaluate 21st-century advances.”²⁵⁸ Moreover, even if the products stayed the same, the science for evaluating them is changing.²⁵⁹ Conceived in the 1960s, FDA’s premarket study requirements fail to take advantage of later-developed technologies for assessing risks and benefits.²⁶⁰ The relevant technologies are in two main areas²⁶¹: (1) basic scientific approaches that can help predict safety²⁶² and effectiveness²⁶³ earlier in the drug-development process, perhaps even before drugs are first taken by human beings in clinical trials²⁶⁴ and (2) new analytical tools to glean risk-benefit information from large databases during the postmarket period.²⁶⁵ FDA’s premarket study requirements ignore evidentiary opportunities at either end, “both before and after an

256 Evans, *supra* note 222, at 753 (discussing the slow pace of clinical translation of genomic discoveries).

257 See *infra* Part IV.A (discussing post-genomic predictive and preventive products and how they differ from traditional medical products of the twentieth century).

258 IOM, BIOMARKER-BASED TOOLS, *supra* note 174, at 30 (reporting presentation of Dr. Janet Woodcock); see also FDA, INNOVATION OR STAGNATION, *supra* note 143, at ii (noting a need to develop new tools for evaluating the safety and effectiveness of medical products).

259 See FDA, INNOVATION OR STAGNATION, *supra* note 143, at 16–20 (discussing new scientific techniques for assessing safety); *id.* at 20–25 (discussing new scientific techniques for assessing efficacy). See generally IOM, EMERGING SAFETY SCIENCE, *supra* note 80 (discussing emerging techniques for assessing drug safety).

260 Evans & Flockhart, *supra* note 81, at 47, 53–54; see also FDA, March 7 Proceedings, *supra* note 227, at 4 (statement of Dr. Andrew von Eschenbach) (discussing the availability of new tools and technology for clinical trials and the need for further progress).

261 IOM, EMERGING SAFETY SCIENCE, *supra* note 80, at 2.

262 *Id.* at 3, 13.

263 FDA, INNOVATION OR STAGNATION, *supra* note 143, at 20–25.

264 IOM, EMERGING SAFETY SCIENCE, *supra* note 80, at 13.

265 *Id.* at 2.

RCT.”²⁶⁶ Taking advantage of these opportunities requires a whole new paradigm. This section discusses two things: the impact of new products and the impact of new analytical capabilities. A common theme emerges from both discussions. FDA’s old evidentiary paradigm not only has problems with safety; it has problems evaluating efficacy as well.

A. *The Challenge of Predictive and Preventive Medical Technologies*

FDAAA grants FDA important new powers which, if skillfully exercised, can reshape regulation to unlock the promise of genomic medicine. What, precisely, is that promise? Many people grope for words when asked to state, succinctly, what genomic medicine is and how it is likely to change the field of medicine. The question sometimes elicits a cryptic, one-word answer: Herceptin. Herceptin (trastuzumab) is a biologic therapy that has a favorable risk-benefit ratio in a subgroup of breast cancer patients whose tumors, when tested, display certain genetic characteristics.²⁶⁷ It exemplifies the potential for pharmacogenomics to predict variations in drug response before a patient takes the drug. However, it grossly understates the promise of genomic medicine to equate genomic medicine with pharmacogenomics.

Genomic medicine has the potential to usher in a whole new model of medicine. If it succeeds, this will mark the second time in the history of western medicine that a new model has emerged. The eventual impact of this shift will be to restore prognosis to a place of legitimacy among the three traditional branches of medicine (diagnosis, treatment, and prognosis). Prognosis occupied a central role in clinical medicine from the time of Hippocrates until late in the nineteenth century.²⁶⁸ With treatment stymied by lack of effective therapies, patients looked to doctors to diagnose illness, supply information about the likely course of their ineptly treated diseases, and recom-

266 Sheldon Greenfield & Richard L. Kravitz, *Heterogeneity of Treatment Effects: Subgroup Analysis*, in IOM, *LEARNING HEALTHCARE*, *supra* note 95, at 113, 119; *see also* IOM, *BIOMARKER-BASED TOOLS*, *supra* note 174, at 60 (reporting calls for better integration of basic sciences and clinical research).

267 *See* IOM, *BIOMARKER-BASED TOOLS*, *supra* note 174, at 21–24 (reporting presentation of Dr. Paul Waring) (discussing Herceptin’s therapeutic action in patients with various test results); Annetine Gelijns, *Lessons for Genomics from Other Technologies*, in IOM, *DIFFUSION & GENOMIC INNOVATIONS*, *supra* note 143, at 16, 19 (citing Herceptin as a successful example of drug-test codevelopment).

268 *See* Ben A. Rich, *Prognostication in Clinical Medicine: Prophecy or Professional Responsibility?*, 23 J. LEGAL MED. 297, 299–302 (2002).

mend measures to ward off future illness.²⁶⁹ These last two services—prediction and prevention—involved prognostic judgment.

The twentieth century brought the first new model of medicine, spurred by rapid advances in diagnosis and treatment.²⁷⁰ Medicine embraced a curative model²⁷¹ in which clinicians' primary task is to diagnose and treat illness after it occurs. The twentieth century did bring notable successes in disease prevention, but these tended to reflect interventions at the population level (for example, better sewerage systems and mass vaccination programs) that can be characterized as triumphs of public health rather than of patient-focused medicine.²⁷² Prognosis—predicting health outcomes at the level of individual patients—came to be seen as unscientific and ethically dubious in light of the uncertainties involved.²⁷³ Prognosis withered as a branch of medicine and persisted in twentieth century medicine mainly as an adjective modifying the diagnosis, as in “terminal” cancer and “end-stage” renal disease.²⁷⁴

Genomic medicine promises further advances in diagnosis and treatment, continuing a century-old trend. Herceptin is one such advance. However, the signature feature of genomic medicine is not just that it will deliver more and better therapies, but that it will place prediction and prognosis on the scientific footing that they largely lacked in the twentieth century.²⁷⁵ Predicting how patients will respond to a given therapy, such as Herceptin, is only a start. The eventual goal is to use genetic information to predict who is susceptible to future disease;²⁷⁶ to predict which of those individuals actually may manifest symptoms when exposed to specific, known pathogens

269 See, e.g., Nicholas A. Christakis, *The Ellipsis of Prognosis in Modern Medical Thought*, 44 SOC. SCI. & MED. 301, 302–11 (1997) (discussing evolution of care for pneumonia as an example of how the role of prognosis has been reduced over time in clinical medicine).

270 Rich, *supra* note 268, at 300.

271 Christakis, *supra* note 269, at 301, 312–13; Rich, *supra* note 268, at 300–01.

272 Califf, *supra* note 143, at 5.

273 Nicholas A. Christakis & Theodore J. Iwashyna, *Attitude and Self-Reported Practice Regarding Prognostication in a National Sample of Internists*, 158 ARCHIVE INTERNAL MED. 2389, 2391–92 (1998).

274 Rich, *supra* note 268, at 302.

275 Gelijns, *supra* note 267, at 18 (“Genomic interventions may produce diagnostic technologies that enhance prognostic abilities.”); see also Christakis & Iwashyna, *supra* note 273, at 2389 (citing genetic testing as a factor giving prognosis renewed vitality as a branch of medicine); Francis S. Collins et al., *A Vision for the Future of Genomics Research*, 422 NATURE 835 (2003) (discussing future research priorities to translate achievements of the Human Genome Project into clinical use).

276 See IOM, *DIFFUSION & GENOMIC INNOVATIONS*, *supra* note 143, at 1 (noting the availability of genetic markers of increased risk for various chronic diseases).

or environmental stimuli;²⁷⁷ to develop preventive interventions that arrest development of disease before it becomes manifest²⁷⁸ or, failing that, to explain disease progression and recurrence in ways that let sickness and the burdens of treatment be minimized.²⁷⁹ These latter capabilities—all of which involve predictive judgments—are the promise of genomic medicine.

In the past, genetic information often was seen as an end in itself—“an endpoint that did not improve health care outcome.”²⁸⁰ The goal of genetic testing was to inform patients whether they had a genetic condition²⁸¹ rather than to guide interventions aimed at changing the condition. Historically, medical genetics dealt with conditions where little could be done and, often, the point was simply to aid decisions on reproduction or abortion.²⁸² Genetic tests could diagnose rare genetic conditions like cystic fibrosis and they could reveal predisposition to certain chronic diseases,²⁸³ but few interventions were available.²⁸⁴ A transformation is now underway.²⁸⁵ Genomic medicine aims to harness genetic knowledge to improve health outcomes.²⁸⁶ The goal is not simply diagnosis, prediction, and prognosis; the goal is to intervene in ways that alter the prognosis.

277 See Gelijns, *supra* note 267, at 18 (commenting that positive tests for disease susceptibility do not always imply disease will occur because ultimate outcomes are affected by other factors such as environment); see also Wylie Burke & Bruce M. Psaty, *Personalized Medicine in the Era of Genomics*, 298 JAMA 1682 (2007) (discussing how improved understanding of interactions between genetic and environmental factors will lead to better prognostic capability).

278 See, e.g., Wylie Burke, *Integrating Genetic Technology into a Health Care System*, in IOM, DIFFUSION & GENOMIC INNOVATIONS, *supra* note 143, at 33, 33–34 (noting increasing potential to intervene to improve health outcomes after susceptibilities have been identified).

279 See, e.g., Gelijns, *supra* note 267, at 18 (discussing possibility of using genetic tests to identify breast cancer patients at high risk of recurrence who need adjuvant chemotherapy); Sean Tunis, *Translating Medical Innovations with Appropriate Evidence*, in IOM, DIFFUSION & GENOMIC INNOVATIONS, *supra* note 143, at 25, 27 (discussing evidentiary review of *Onco type Dx* gene-expression profiling for prediction of risk of breast cancer recurrence).

280 Burke, *supra* note 278, at 33.

281 *Id.* at 33–34.

282 *Id.* at 34.

283 See, e.g., IOM, DIFFUSION & GENOMIC INNOVATIONS, *supra* note 143, at 1 (noting such tests' ability to discover conditions like cystic fibrosis and Huntington's disease and to reveal predisposition to diabetes and coronary artery disease).

284 Burke, *supra* note 278, at 33–34; Gelijns, *supra* note 267, at 20 (noting that identifiable diseases often have “no cures, only treatments with limited effectiveness and major side effects”).

285 See IOM, DIFFUSION & GENOMIC INNOVATIONS, *supra* note 143, at 1.

286 *Id.*

“When new technologies are introduced into health care they may be relatively primitive”²⁸⁷ Presently we are in the Stone Age of genomic medicine. For example, genetic tests can identify women whose genes predispose them to breast cancer, but today’s preventive interventions have the technological sophistication of a stone axe: “[T]here are no known measures to substantially reduce the risk of breast cancer in women who test positive for these mutations short of prophylactic mastectomies”²⁸⁸ Hope is for the day when people who test positive for genes associated with cancer susceptibility can receive medicines that supply the tumor-suppressing proteins their genes are unable to manufacture, or that cancel the tumor-promoting proteins with which defective genes are flooding their bodies. That is the promise of genomic medicine.

This interventional focus of genomic medicine will, in some respects, make genetic information less exceptional than it once was. Like any other medical information, it will become a regular part of day-to-day clinical decisions.²⁸⁹ The Department of Health and Human Services’ Personalized Health Care Initiative²⁹⁰ aims to foster wider integration of genetic data into routine clinical practice. Yet this trend presents challenges distinct from those encountered in regulating twentieth century medical products. The sheer futurity of benefits from predictive and preventive products will strain FDA’s existing systems for evaluating effectiveness. Such products, by design, will deliver benefits and risks over prolonged time frames (ten to sixty years), far beyond the duration of clinical trials as we now know them.

Another problem is that genomic medicine combines technologies in ways that cut across FDA’s traditional regulatory categories (device, drug, biologic).²⁹¹ In particular, drugs will be used in con-

287 Gelijns, *supra* note 267, at 18.

288 IOM, BIOMARKER-BASED TOOLS, *supra* note 174, at 69 (reporting presentation of Dr. Scott Ramsey).

289 Burke, *supra* note 278, at 34.

290 IOM, DIFFUSION & GENOMIC INNOVATIONS, *supra* note 143, at 1; *see also* U.S. Dep’t of Health & Human Servs., Personalized Health Care, <http://www.hhs.gov/myhealthcare> (last visited Nov. 9, 2009) (detailing the Initiative’s goal of using genetic information in personalized healthcare).

291 *See* 21 C.F.R. § 3.2(e) (2009) (defining “[c]ombination product”). Future products that combine drugs, devices, and/or biologics may or may not fit within this formal definition which requires cross-referencing of specifically named products. *See id.* Either way, combined use of products poses a number of special regulatory problems. *See* Evans, *supra* note 222, at 786–87. A special office has been established within FDA to address issues raised by such products. *See* U.S. Food & Drug Admin., Office of Combination Products, <http://www.fda.gov/AboutFDA/CentersOffices/>

junction with *in vitro* diagnostic devices (genetic tests and tests for other biomarkers²⁹²). Tests will predict patient response to specific drugs. Drugs may be taken preventively by people who have nothing presently wrong with them other than a bad predictive test result. In these situations, the safety and efficacy of the test and drug may not be cleanly separable.²⁹³ If a test predicts a person safely can take a drug, but the person nevertheless sustains a drug-related injury, which was unsafe: the drug, the test, or both?²⁹⁴ Suppose a drug is designed to reduce incidence of a certain type of cancer in patients with positive tests for susceptibility. If, after many years, a test subject does not get the cancer, does that mean the preventive intervention worked, that the cancer-susceptibility test was inaccurate, or that the test correctly identified a susceptibility that failed to express itself as fully-blown disease for reasons that had nothing to do with the preventive intervention? What will efficacy mean in this context and how—and over how long—must it be assessed?

There is a risk that preventive interventions may enter clinical use with no FDA review of their safety and effectiveness. This will be possible when a preventive intervention is an off-label use of a drug already approved for treating manifest disease. Thus, cholesterol-lowering statins are a candidate to become one of the first widely used preventive interventions.²⁹⁵ Statins already are approved for use in treating high cholesterol. It is thought by some that they may have potential for reducing long-term health risks in patients who would not, based on their current status, be seen as needing treatment for high cholesterol.²⁹⁶ This latter use would amount to a preventive intervention. Its risks and benefits are unknown. The American Society of Clinical Oncology and the American Urological Association stated that off-

OC/OfficeofScientificandMedicalPrograms/OfficeofCombinationProducts/default.htm (last visited Nov. 9, 2009).

292 A biological marker (biomarker) is a measurable property—such as a gene, a protein circulating in the blood, or a metabolite (a chemical formed as the body metabolizes a drug)—that, by its presence, reveals information about clinically relevant conditions such as disease, disease susceptibility, or a person's potential to benefit from or be harmed by a particular therapy. FURBERG & FURBERG, *supra* note 60, at 61.

293 B.J. Evans, *Distinguishing Product and Practice Regulation in Personalized Medicine*, 81 CLINICAL PHARMACOLOGY & THERAPEUTICS 288, 290 (2007).

294 Evans, *supra* note 222, at 771.

295 C. Thomas Caskey, Dir. & C.E.O., Brown Found. Inst. of Molecular Med., Presentation at Cambridge Healthtech Inst. R&D Risk Mitigation Conference, at slide 14 (Mar. 26, 2008) (on file with author) (showing presymptomatic use of statins as part of a new paradigm for coronary artery disease).

296 *Id.*

label use of finasteride, a drug used to treat male pattern baldness and urological problems, may reduce the risk of developing prostate cancer.²⁹⁷ In the case of finasteride, there is some clinical trial evidence supporting this use although the preventive use has not been approved by FDA. Since off-label prescribing of drugs is lawful,²⁹⁸ approved drugs can be prescribed for preventive purposes even without prior clinical trials. Even if FDA had authority to require such trials, they might be difficult to carry out. There are well-known problems in accruing subjects for clinical trials of drugs that already are on the market, since patients who want the drug can gain access to it without enrolling in a trial.²⁹⁹ Subjects' unwillingness to accept randomization in this situation contributes to low completion rates for postmarket clinical trials.³⁰⁰

Preventive interventions pose regulatory challenges far more complex than the drug safety problems FDA faced earlier this decade. With drug safety, the central problem is that safety is uncertain at the point of FDA approval. However, FDA traditionally has been able to assess the efficacy of drugs before they are approved (at least with respect to the drugs' indicated uses and, admittedly, sometimes through the use of surrogate efficacy endpoints³⁰¹). For future medical products that have long-term disease prevention as the endpoint, the uncertainties will go to both efficacy and safety. To establish true efficacy of these products in premarket trials would entail long-delayed access for patients and long-deferred cash flow for product sponsors. The economics, in many cases, would be fatal to product development. Predictive and preventive technologies seemingly will have to be approved based on surrogate endpoints of efficacy or on a showing that there is a plausible scientific basis to expect long-term efficacy. There may be decades of uncertainty before efficacy can be confirmed. This implies a need for ongoing postmarket surveillance of efficacy as well as of safety.³⁰² Yet FDA's traditional systems for

297 Shankar Vedantam, *More Men Are Urged to Take Drug Against Prostate Cancer*, WASH. POST, Feb. 25, 2009, at A14.

298 See *supra* notes 222, 224 and accompanying text (discussing off-label use of drugs).

299 See IOM, BIOMARKER-BASED TOOLS, *supra* note 174, at 71 (reporting presentation of Dr. Scott Ramsey); see also FRIEDMAN ET AL., *supra* note 54, at 7–10 (discussing logistical and ethical problems of recruiting subjects for trials of drugs that already are clinically available).

300 Weisman et al., *supra* note 95, at 132.

301 See *supra* note 240 (defining surrogate endpoints); *infra* Part V.B.1.

302 See IOM, DIFFUSION & GENOMIC INNOVATIONS, *supra* note 143, at 24 (discussing the need for evidence with respect to unexpected benefits as well as unexpected risks).

postmarket surveillance have focused on safety, not efficacy. Predictive and preventive products of the future will amplify strains on FDA's 1962 evidentiary paradigm, which already were apparent as applied to twentieth century medical products.³⁰³

Another challenge is that preventive interventions imply giving drugs to healthy people—people who are susceptible to future illness but are currently asymptomatic. Risks that are acceptable in treating manifest disease may be unacceptable when drugs are used in healthy people who will receive only remote, future benefits. Moreover, if disease susceptibility is misjudged, exposure to *any* risk is needless. Safety and effectiveness of predictive tests are major concerns. There have been calls for prospective RCTs to assess risks and benefits of tests that will be used to guide decisions about treatment and prevention.³⁰⁴

At present, most such tests do not go through clinical trials before they enter clinical use. Over 90% of genetic tests currently on the market are lab-developed tests (LDTs) regulated under CLIA³⁰⁵ and not regulated by FDA.³⁰⁶ Even when FDA does regulate a test, prospective clinical trials of safety and effectiveness are not necessarily required.³⁰⁷ CLIA does not require a data-driven regulatory review to confirm safety and efficacy before a test moves into clinical use.³⁰⁸ Yet people rely on LDTs when making important medical decisions. For example, CLIA-regulated LDTs include the widely used BRCA tests for breast cancer susceptibility³⁰⁹ and the *Onco type Dx* test which is sometimes used to aid decisions whether women with node-negative Stage 1 breast cancer can forego chemotherapy.³¹⁰ Off-label use of tests will

303 See discussion *supra* Parts II–III.

304 See, e.g., IOM, *BIOMARKER-BASED TOOLS*, *supra* note 174, at 70–71 (reporting Dr. Scott Ramsey's concern that "[s]tudies should clearly define the risk/benefit ratio of a biomarker test prior to its use in the clinic").

305 Clinical Laboratory Improvement Amendments of 1988, Pub. L. No. 100-578, 102 Stat. 2903 (codified as amended at 42 U.S.C. § 263a (2006)); see also 42 C.F.R. pt. 493 (2009) (outlining laboratory requirements).

306 SACGHS, U.S. OVERSIGHT, *supra* note 223, at 39 (noting that over 1500 genetic tests are currently offered in the United States, of which only several dozen have been cleared or approved by FDA).

307 See *supra* notes 100–01 and accompanying text.

308 Evans, *supra* note 222, at 768.

309 INST. OF MED. & NAT'L RESEARCH COUNCIL, *SAVING WOMEN'S LIVES* 225 (2005), available at http://books.nap.edu/openbook.php?record_id=11016.

310 See sources cited *supra* note 279 (discussing use of prognostic biomarkers for breast cancer recurrence); see also Steven Shak, *Introducing a Genomic Innovation to Clinical Practice*, in IOM, *DIFFUSION & GENOMIC INNOVATIONS*, *supra* note 143, at 53, 55–60 (discussing CLIA-regulated status of the *Onco type Dx* test).

also be an issue. It is not uncommon for doctors to press old tests into new uses for which they were not intended. Thus, the prostate-specific antigen (PSA) test was intended for surveillance and prognostic use in men already diagnosed with prostate cancer, but has been widely used off-label to screen healthy men for possible prostate cancer—a use for which its safety and effectiveness are unconfirmed.³¹¹ In the same way, existing tests could be diverted into (unvalidated) use as tests for disease susceptibility.³¹²

It is appealing to demand prospective RCTs of all such tests, but such a policy appears unrealistic. There are fundamental barriers to imposing drug-like premarket study requirements on tests. Unlike drugs, many tests have short product life cycles.³¹³ While a drug may remain on the market for twenty to fifty years, a test may be replaced by a new-generation test in as little as twelve to eighteen months.³¹⁴ Development costs and timelines also differ: 1–2 years and \$1–2 million to develop a device, versus 10–13 years and \$1 billion for a drug.³¹⁵ When medical technology is rapidly evolving, RCTs may be inappropriate;³¹⁶ results would come too late to be of relevance. There are also logistical problems. The needed clinical trials would be “extremely large, lengthy, and costly. In some cases they may not be feasible because of difficulties in accruing enough patients”³¹⁷ How to fund such trials is also a major question.³¹⁸ The genetic testing industry is structurally different from the drug industry. Tests often serve niche markets; there has not yet been a “blockbuster” genetic test. Test developers may be small, thinly capitalized entities³¹⁹ that lack the financing capacity that has enabled large drug manufacturers to endure the cash outlays and deferred sales revenues

311 IOM, *BIOMARKER-BASED TOOLS*, *supra* note 174, at 70 (reporting presentation of Dr. Scott Ramsey).

312 *Id.*

313 *Id.* at 77 (“Because the technologies for genomics and proteomics assays are rapidly evolving, . . . standards have to be adaptable to the changes in technology that are continually occurring.”).

314 Evans, *supra* note 222, at 781.

315 Rebecca Henderson & Cate Reavis, *Eli Lilly: Recreating Drug Discovery for the 21st Century* 16 (MIT Sloan Sch. of Mgmt., Doc. No. 07-043, Mar. 13, 2008), available at <https://mitsloan.mit.edu/MSTIR/IndustryEvolution/RecreatingDrugDiscovery/Documents/07-043-Recreating-Drug-Discovery.pdf>.

316 FRIEDMAN ET AL., *supra* note 54, at 8.

317 IOM, *BIOMARKER-BASED TOOLS*, *supra* note 174, at 71 (reporting presentation of Dr. Scott Ramsey).

318 *Id.* at 71–72.

319 Evans, *supra* note 222, at 789–90 (discussing role of academic scientists and small-test developers in discovery and development of new tests).

associated with long premarket RCTs. There have been calls for public funding of trials to validate tests,³²⁰ but the cautious assumption is that it may not be forthcoming.

Two governmental advisory panels, in 1997 and 2000, examined the question of regulatory oversight of tests; both called for rigorous, data-driven review of tests before they move into clinical use.³²¹ In 2007, a successor advisory panel³²² revisited this issue and, in light of the many issues surrounding premarket validation of tests, edged away from that recommendation. This last panel instead suggested a risk-stratified approach that would focus a more rigorous review on higher-risk tests.³²³ Criteria for risk stratification and regulatory mechanics have not, as yet, been worked out.³²⁴

To be fair, FDA's existing evidentiary paradigms (for drugs and devices) may be adequate for many of the products that will flow out of the Human Genome Project. This will be true when risks and benefits can be ascertained, with reasonable accuracy, using FDA's traditional premarket study mechanisms. This was the case with Herceptin and its related first-generation test, which were approved by FDA in separate, coordinated biologics licensing and device approvals on the same day.³²⁵

The problem will lie with products intended for long-term predictive or preventive uses. Here, premarket gathering of data may be

320 IOM, BIOMARKER-BASED TOOLS, *supra* note 174, at 73 (reporting presentation of Dr. Charles L. Sawyers) (noting that some observers believe that because biomarkers that predict adverse drug reactions protect the public, they should therefore be publicly funded).

321 See NIH-DOE, SAFE GENE TESTING, *supra* note 223, at ch. 2; SACGT, ENHANCING OVERSIGHT, *supra* note 223, at ix-x.

322 See SACGHS, U.S. OVERSIGHT, *supra* note 223, at i.

323 *Id.* at 10-11 (providing recommendations concerning oversight of clinical utility of tests without calling for data-driven review of clinical utility before tests go into clinical use, as had earlier reports).

324 *Id.* at 136 (listing elements that should be included in a risk-stratification classification algorithm).

325 See Genentech, Herceptin Development Timeline, <http://www.gene.com/gene/products/information/oncology/herceptin/timeline.html> (last visited Nov. 7, 2009); see also Letter from Jay P. Siegel, Dir., Office of Therapeutics Research & Review, Ctr. for Drug Evaluation, to Robert L. Garnick, Genentech, Inc., Sept. 25, 1998, available at <http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/ApprovalApplications/TherapeuticBiologicApplications/ucm091360.pdf> (granting FDA's approval of the biologics license application for Herceptin); U.S. Food & Drug Admin., September 1998 PMA Approvals, <http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/DeviceApprovalsandClearances/PMAApprovals/ucm115106.htm> (last visited Nov. 26, 2009) (showing approval of Dako HercepTest on the same day Herceptin was approved).

infeasible or less-than-fully informative. The likely reality is that many predictive and preventive medical products will be moving into clinical use with substantial uncertainty about their risks and benefits. Off-label use of drugs and tests, and CLIA regulation of LDTs, afford wide pathways for this to occur. These lingering uncertainties imply a need for ongoing postmarket evidence development (focusing on efficacy as well as safety) and expanded regulatory powers in the period after products are in clinical use. The 1962 evidentiary paradigm, with its heavy concentration of research and regulatory effort in the premarket period, is weak in both these areas and therefore poorly suited to the task ahead.

B. *The Challenge of Heterogeneous Drug Response*

Other recent developments cast doubt on FDA's assessments of efficacy and safety, even as these apply to traditional medical products. "As scientific knowledge advances, ability to classify is improved. Today's homogeneous group may be considered heterogeneous tomorrow."³²⁶ The Human Genome Project has had this effect. It sparked advances in basic sciences such as genetics, proteomics,³²⁷ and metabolomics.³²⁸ These sciences reveal that people who look alike in terms of traditional trial selection criteria may differ greatly at the genetic, molecular, cellular, and tissue levels, and these differences affect how people respond to drugs.³²⁹ Subjects in a phase III drug trial are characterized in terms of a set of trial selection criteria

326 FRIEDMAN ET AL., *supra* note 54, at 34.

327 See Mayo Clinic, Understanding Genomics and Proteomics, <http://www.mayoclinic.org/mayo-magazine/2006-spr-glossary.html> (last visited Oct. 27, 2009) (defining proteomics as the study of the structure and function of the various proteins that are manufactured by people's genes); Princeton Univ., WordNet Search—Proteomics, <http://wordnetweb.princeton.edu/perl/webwn?s=proteomics> (last visited Oct. 27, 2009) (defining proteomics as a branch of genetics that studies "the full set of proteins encoded by a genome").

328 See The Human Metabolome Project, Metabolomics Overview, <http://www.metabolomics.ca/about/overview.htm> (last visited Oct. 27, 2009) (defining metabolomics as the study of metabolites found in the human body). Metabolites include chemicals and molecules that the body makes as it metabolizes a drug, and these can differ from one person to the next.

329 See Greenfield & Kravitz, *supra* note 266, at 113 (discussing genotypic and phenotypic characteristics that affect drug response and risk of adverse reactions); see also Tommy Andersson et al., *Drug-Metabolizing Enzymes: Evidence for Clinical Utility of Pharmacogenomic Tests*, 78 CLINICAL PHARMACOLOGY & THERAPEUTICS 559 (2005) (discussing use of genetic markers to predict drug metabolism); Lawrence J. Lesko et al., *Pharmacogenetics and Pharmacogenomics in Drug Development and Regulatory Decision Making: Report of the First FDA-PWG-PhRMA-DruSafe Workshop*, 43 J. CLINICAL PHARMACOLOGY 342 (2003) (discussing regulatory use of pharmacogenetic information).

such as age, gender, severity or duration of underlying disease, presence or absence of other medical conditions, and whether they use drugs other than the one being tested. Subjects who are alike in terms of those variables are presumed to be “comparable”³³⁰ to one another. Unfortunately, no compact set of demographic and clinical variables captures the true extent of individual variability. Phase III drug trials misperceive their human research subjects, presuming dissimilar individuals to be alike.³³¹

By analogy, we all are familiar with pixilated facial photos such as the ones used by investigative reporters to conceal the identities of whistleblowers. At a very rough level of pixilation, people’s facial photos are reduced to assemblages of colored squares and everybody looks more or less alike. A human is a creature with shoulders and a head; some inferences about gender can be drawn based on hairstyle and facial hair. Suppose a sample of six photos is drawn from a group of sixty such photos. At a rough level of pixilation, photos in this sample may seem comparable to one another and representative of the larger group. That may no longer be true once the pictures are brought into full, sharp focus. Suppose three of the six selected photos depict people with melanomas on their noses, while all fifty-seven of the other photos display people with clear complexions. The six selected photos are neither comparable to one another nor representative of the larger group, at least for purposes of skin-cancer research.

If blurry, pixilated photos were the only means available for observing human beings, it would be reasonable to rely on them and try to do the best research one could do at that level of pixilation. As clearer photos became available, it would be time to augment or replace those old research techniques. That is precisely what is happening to FDA’s clinical trial process. The traditional phase III drug trial “pixilated” its participants in terms of a rough set of trial selection criteria (such as age, gender, comorbidities) and baseline measurements (such as blood pressure readings at the start of the study).³³² People who shared these characteristics were presumed to be alike, in the sense of having an equal *ex ante* probability of being helped or

330 See 21 C.F.R. § 314.126(b)(4) (2009) (discussing selection of human participants and variables pertinent in ensuring comparability of test subjects and control groups).

331 FDA, March 7 Proceedings, *supra* note 227, at 10 (statement of Dr. Janet Woodcock) (discussing failure of trials to differentiate subpopulations using today’s available science).

332 FRIEDMAN ET AL., *supra* note 54, at 130–39 (discussing baseline assessment in clinical trials).

harmful by the test drug. Now, basic scientific advances are making it possible to bring research subjects into a sharper focus. What will be the impacts of this change?

Here, it is helpful to distinguish the concept of generalizability and two different aspects of comparability. Generalizability, discussed earlier,³³³ concerns whether trial subjects are representative of people *not involved in the trial*. For example, are the subjects (on average) sufficiently like real patients (on average) to allow the trial results to be extrapolated to the routine clinical setting? Comparability concerns whether people in the trial are similar to *one another*. There are two aspects to this. First, is the control group comparable to the intervention group? If not, then observed differences in their health outcomes may reflect differences between the two groups, rather than the effects of taking versus not taking the drug. Second, are people *within* the interventional group comparable to one another? If not, then the *average* effects of the drug within that group may be quite different from the effects experienced by particular individuals. This last phenomenon is the one that is causing concern.

Heterogeneity of Treatment Effects (HTE), defined as variation of results produced by the same treatment in different patients,³³⁴ has always been observed in clinical drug trials.³³⁵ Figure 1³³⁶ is a schematic of variable responses that might be observed in the interventional group of a phase III drug trial—that is, among the subjects who were exposed to the hypothetical drug.

The horizontal axis shows the frequency of various responses within the interventional group. The vertical axis records the magnitude of benefits and harms to various people. In this hypothetical trial, 10% of the interventional subjects, seen at left, reacted very badly. They experienced harms (shown in black) and got no benefits at all. Many of the subjects (the 40% in the middle) were “non-responders,” experiencing neither benefits nor harms from the drug. In the remaining half of the intervention group, subjects experienced varying degrees of benefit (shown in gray) and harm. Moving left to right, some people got a mix of harms and benefits; others got a small benefit, while some benefited greatly with no adverse effects.

These variations occur for many reasons including mundane ones (for example, failure by some subjects to follow the instructions).

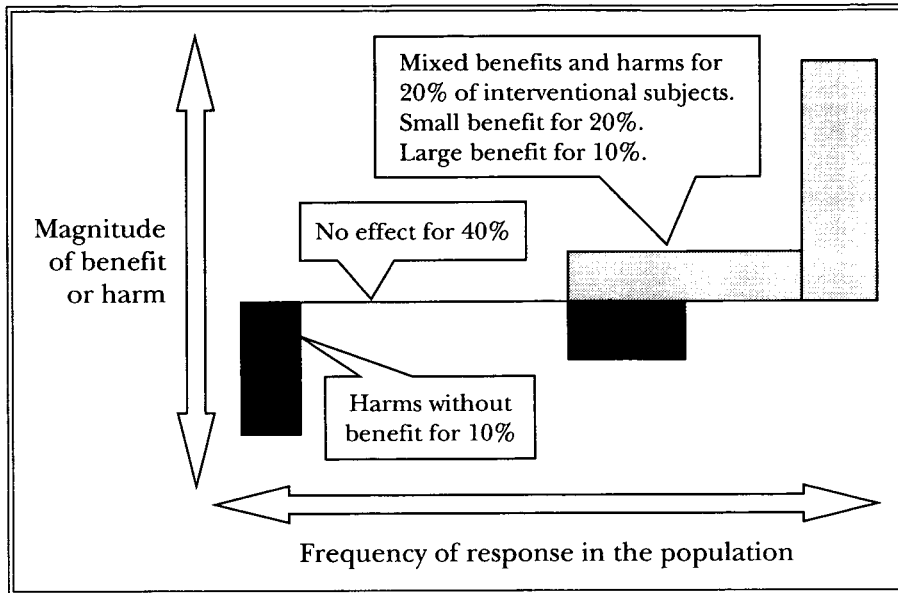
333 See discussion *supra* Part II.B.3.

334 Greenfield & Kravitz, *supra* note 266, at 114.

335 *Id.*

336 Figure 1 is used with permission of the *Food and Drug Law Journal*. See also Evans, *supra* note 222, at 763 fig.2 (discussing a conceptually similar diagram).

FIGURE 1. HETEROGENEOUS RESPONSE WITHIN THE INTERVENTIONAL GROUP IN A CLINICAL TRIAL



As noted earlier, subjects may get bigger benefits from a drug if they are more severely ill or at higher risk for the condition the drug is designed to treat (for example, the risk of heart attack).³³⁷ Recent scientific advances have shed light on a host of additional factors—genetic, biological, and metabolic—that also help explain the variations in response.³³⁸ For example, people have different genes that cause their livers to make different enzymes, and those enzymes affect how patients metabolize drugs.³³⁹ Tumors express different genes that may make a tumor more or less susceptible to destruction by a particular drug.³⁴⁰ Two people with the same illness may have been infected by genetically different germs, which are more or less suscep-

337 See *supra* notes 217–18 and accompanying text.

338 Greenfield & Kravitz, *supra* note 266, at 113.

339 Letter from Steven I. Gutman, Dir., Office of *In Vitro* Diagnostic Device Evaluation & Safety, Ctr. for Devices & Radiological Health, to James F. Kelly, Regulatory Affairs, Roche Molecular Sys., Inc. (Dec. 23, 2004), available at http://www.accessdata.fda.gov/cdrh_docs/pdf4/k042259.pdf (approving a drug-metabolizing enzyme genotyping system to detect variations in liver enzymes that have been shown, in peer-reviewed literature, to have an impact on drug response).

340 See, e.g., Genentech, Full Prescribing Information for Trastuzumab (Herceptin) (2009), available at <http://www.gene.com/gene/products/information/oncology/herceptin/insert.jsp> (describing a breast cancer therapy targeted to tumors expressing a particular gene).

tible to a particular course of treatment. “Within clinical trials, these markers have the potential for identifying a patient’s potential for responsiveness”³⁴¹

This new understanding of individual variability does not cast doubt on the basic validity of comparisons between the interventional and control groups in a randomized clinical trial. The advantage of randomization, always, has been that it helps ensure comparability of these two groups.³⁴² If subjects are randomly assigned to be in either the intervention or the control group, then each group will receive a mix of people with various characteristics. Variations among the subjects (both the known variations that have been measured, and unknown ones such as hidden genetic or metabolic differences) will tend to be spread evenly between the two groups.³⁴³ Recent advances highlight how many factors were unknown and thus ignored in the past. However, this was not “news.” When conducting trials, scientists always knew it was impossible to identify and measure all the factors that affect drug response.³⁴⁴ By randomly assigning subjects to the two groups, it could safely be assumed that the unknown factors would exert similar impacts on both groups.³⁴⁵ When this is true, intergroup comparisons still are meaningful. Unknown factors tend to exert the same impact on health outcomes observed in both groups of a large clinical trial. One does not need to know what the unknown factors are, so long they affect both groups similarly. Thus, individual variability does not undermine the basic validity of large, randomized clinical trials.

What these discoveries have undercut is FDA’s reliance on average statistics to characterize the safety and effectiveness of drugs. As noted by a senior FDA official, “[t]he clinical trial process has been highly observational in its conduct, primarily because we don’t have the tools to look at the basis for individual response so we look at population responses.”³⁴⁶ In other words, clinical trials simply observe how individual subjects respond and compile average statistics based

341 Greenfield & Kravitz, *supra* note 266, at 113.

342 FRIEDMAN ET AL., *supra* note 54, at 43.

343 *Id.* at 37, 43; *see also id.* at 38 (noting that the “investigator can only describe to a limited extent the kinds of participants in whom an intervention was evaluated”).

344 *Id.* at 118 (noting that, at the start of a trial, “all important prognostic factors have probably not been identified, nor can all of them be measured”).

345 *Id.* at 43, 138 (noting that, while it is possible by chance to end up with intervention and control groups that differ with respect to some unknown prognostic factor, this is unlikely in a large, randomized trial).

346 IOM, BIOMARKER-BASED TOOLS, *supra* note 174, at 31 (internal quotation marks omitted) (quoting Dr. Janet Woodcock).

on those observations, but do not inquire why the subjects responded the way they did. It is assumed that “despite such diversity, the effect of the intervention is more similar among the various types of participants than not.”³⁴⁷ This assumes that average responses are meaningful indicators of a drug’s safety and effectiveness for individuals.

FDA’s 1962 drug approval criteria only require that a drug be safe and effective at the level of the average person who took it during premarket trials.³⁴⁸ If the drug delivered any therapeutic benefit (that is, if the gray area in Figure 1 is larger than the corresponding gray area observed in the control group), the average benefits are positive and the drug can be deemed effective. A few patients who have dramatic effects may account for much of the average result of a trial.³⁴⁹ Trial results can be overwhelmed by few patients who get a big benefit.³⁵⁰ A drug is “safe” if it has a favorable ratio of benefits to risks.³⁵¹ If the collective benefits in the interventional group (the total gray area) are judged to outweigh the collective harms (the black area) in that same group, then that ratio is favorable. The drug can be deemed safe. Both of these determinations, safety and efficacy, are based on average effects within the group.

Obviously, for many of the trial subjects in Figure 1, the drug was neither safe nor effective. After approval, this drug is destined to have ongoing problems with safety and efficacy due to variations in individual response. Labeling must disclose the safety risks that were observed among subjects who took the drug. What is not disclosed is the maldistribution of risks and benefits. Labeling does not disclose the percentage of subjects who received nothing but harm, nor does it disclose the non-response rate, nor the varying degrees of benefit and variable severity of harms. In short, drug labeling does not disclose most of the information depicted in Figure 1. In analyzing and presenting trial data, it is assumed that subjects who took the drug were comparable to one another, such that each could expect a

347 FRIEDMAN ET AL., *supra* note 54, at 34.

348 See Evans, *supra* note 222, at 764.

349 Kent & Hayward, *supra* note 217, at 1210 (citing John P.A. Ioannidis & Joseph Lau, *The Impact of High-Risk Patients on the Results of Clinical Trials*, 50 J. CLINICAL EPIDEMIOLOGY 1089 (1997) and discussing the fact that, frequently, trial outcomes depend on results seen in relatively small number of subjects who were in a position to gain large benefits from the drug).

350 See *id.* at 1209.

351 See *supra* notes 215–16 and accompanying text (discussing FDA’s concept of safety).

response similar to the population average.³⁵² The fact that their responses differed is attributed to chance.

This last point has become highly contestable in the years after the Human Genome Project. Among carefully screened test subjects who look alike in terms of measured characteristics, individual drug response is heterogeneous. Some of this heterogeneity really does reflect random chance, but scientists now know that much of it does not. If there are a priori reasons to expect differences, “analyses of means would give the wrong answer.”³⁵³ These a priori reasons could include genetic and metabolic differences that affect drug response³⁵⁴ or unmeasured differences in patients’ true health risks and hence their potential to benefit from a drug.³⁵⁵ Averaging responses across dissimilar individuals may produce average data that are relevant to no one.³⁵⁶ “When [average] statistics dominate the entire drug regulatory approval process . . . the end result is distorted because it does not account for individual variability.”³⁵⁷ In this situation, “the effort to assess risks and benefits breaks down.”³⁵⁸ “[E]rror infects the entire process. At the outset the FDA uses the wrong standards insofar as it evaluates the usefulness and risk of drugs on the assumption that all parties are like the median user.”³⁵⁹ These errors affect both efficacy and safety. Already by 1979, Congress was hearing testimony that most FDA-approved drugs only work in 35% to 70% of patients who take them.³⁶⁰ The science of that day could not explain or predict the observed variations in individual response. Thus, there was no scientific basis to do anything other than continue framing safety and efficacy in average terms. Now, there is.

By ignoring important inter-individual variations, FDA’s evidentiary requirements after 1962 cast safety and efficacy as attributes of

352 Greenfield & Kravitz, *supra* note 266, at 115.

353 IOM, UNDERSTANDING BENEFITS, *supra* note 16, at 50 (reporting presentation of Dr. Brian L. Strom).

354 See *supra* notes 338–41 and accompanying text.

355 See Kent & Hayward, *supra* note 217, at 1209.

356 See *id.* at 1210 (noting that average risk and average treatment effect can differ from that in a typical patient if baseline risk is skewed).

357 IOM, UNDERSTANDING BENEFITS, *supra* note 16, at 50 (reporting presentation of Peter Barton Hutt).

358 *Id.* at 16.

359 Epstein, *supra* note 43, at 32.

360 *The Food and Drug Administration’s Process for Approving New Drugs: Hearings Before the Subcomm. on Sci., Research, & Tech. of the H. Comm. on Sci. & Tech.*, 96th Cong. 211 (1979) (statement of Dr. Gilbert McMahon).

the medical product itself.³⁶¹ In reality, a product's safety and efficacy have meaning only at the level of particular individuals or subgroups for which response-determining factors are understood.³⁶² Absent such knowledge, product-level assurances of safety and effectiveness are not meaningful and may be dangerously misleading to the public. Science is far from having a complete understanding of the many factors that affect drug response; only a few such factors are known.³⁶³ Still, individual variations in treatment response have gone from being an "unknown unknown," as they were in the 1960s, to a "known unknown." Scientists know enough to know that the hypothetical average trial participant—that pixilated patient for whom FDA declares products to be safe and effective—is a fiction.

What is the right policy response to this knowledge? In theory it would be possible, at great expense, to expand trial populations to include multiple subgroups stratified according to every known factor thought to be capable of driving variations in treatment response. These subtrials would assess which subgroups safely and effectively can consume a drug, and that information could be reflected in drug labeling. This obviously is impractical. The stratified subgroups would offer, at best, a partial "depixilation" of the overall trial population. The trial would assess the impact of known factors, but there still would be unknown factors and these would continue to generate unexpected variability in response to approved drugs. The required trial populations would be vast, raising the same practical and ethical issues discussed above in connection with rare risks. Trials for some subgroups might be overtly unethical since it is a fair inference, even before a trial is conducted, that people with certain genotypes may face higher-than-average risk of adverse events. While this cannot be known for certain before doing the trial, is it ethical even to include these subgroups in clinical trials? Yet if these subgroups are routinely excluded from trials, they ultimately will lack an adequate supply of drugs approved for their use.

361 More precisely, safety and efficacy were framed as attributes of the product and the indication for which it is labeled, for example, drug *X* is safe and effective for use in treating headache pain. See 21 U.S.C. § 355(d)(1) (2006) (requiring drugs to be "safe for use under the conditions prescribed, recommended, or suggested in the proposed labeling thereof"). However, FDA's permissive policy on off-label use has contributed to widespread public perception that safety and effectiveness are attributes of the product, irrespective of the indication for which it ultimately is used. See *supra* notes 222, 224.

362 See IOM, UNDERSTANDING BENEFITS, *supra* note 16, at 16.

363 See *id.* at 25 (citing the need to "[a]cknowledge that genetic polymorphisms exist and that most have not been characterized" and that they are responsible for multiple dose-response curves for some drugs).

This new understanding about variable treatment response needs to be woven into FDA's evidentiary base. It "does not fit into the traditional models for running clinical trials and developing therapeutics or diagnostics."³⁶⁴ A year before Congress passed FDAAA, the case for a new paradigm was clear:

The post-genomic revolution will increase the pressure on the system to produce irrefutable evidence about which technologies are effective and which are ineffective or even dangerous. Knowledge of genes, proteins, and metabolites, along with functional imaging, will divide populations into multiple "phenotypes," creating both an unprecedented opportunity to improve clinical outcomes as well as an enormous subgroup issue that will require many more trials in larger numbers of human subjects. As the market becomes flooded with inadequately studied tests used to stratify treatment, the demand for adequate comparative studies will understandably increase. With current regulations, however, the cost of the required increase in clinical trials enrollment will be prohibitive.³⁶⁵

V. SEVEN PILLARS OF THE NEW EVIDENTIARY PARADIGM

FDAAA sets seven pillars of a new evidentiary paradigm. The first is to expand postmarket evidence development and diversify its sources. The second is to take a flexible view of the best evidence, making case-by-case determinations of the best methodology for assessing benefits and risks in the postmarket period. The third is to regard the risk-benefit ratio as dynamic and subject to successive improvement throughout a drug's commercial life. The fourth is to recognize that many of these improvements will be made by parties other than the drug manufacturer and to address entry barriers that block wider participation in research and development. The fifth is to regard efficacy failure as a safety problem in its own right. The sixth is to direct attention to safety and efficacy at the individual/subgroup level as well as at the average level. The seventh pillar is to expand postmarket decisional authority, so that emerging risk-benefit information is not idle knowledge but drives actions to improve the public's health.

A. *Postmarket Evidence Development*

FDAAA rejects the notion that drugs can be made safe and effective by directing ever-increasing research and regulatory effort into

³⁶⁴ IOM, *BIOMARKER-BASED TOOLS*, *supra* note 174, at 60 (reporting presentation of Dr. Stephen Friend).

³⁶⁵ Califf, *supra* note 25, at 498 (writing in 2006).

the premarket period. Under FDAAA, FDA's premarket evidence requirements continue in force but they are not statutorily expanded. FDAAA responds to calls for FDA to expand its evidentiary base and harness multiple, new sources of data, including observational studies³⁶⁶ and genetic and other biomarkers that help explain safety and effectiveness.³⁶⁷

The 1962 Amendments conceived drug regulation as a gatekeeping function. The regulator's key decision was whether to open the gate and let new products onto the market. FDAAA retains the gatekeeper but adds capability to detect and manage risks after products pass through the gate. This approach may strike some as compromising safety, since it implies an acceptance that the gatekeeper will let some unsafe products slip through the gate. This is not compromise; it is realism.³⁶⁸ The gate is intrinsically porous, and safety cannot be achieved by fighting that fact but rather by responding to it. In Melmon's view, key constituencies such as the medical profession and academics overestimated the power of premarket testing and consequently showed "little, if any, leadership" in developing and using postmarket risk-benefit data.³⁶⁹ In FDAAA, Congress has supplied the missing leadership.

1. Power to Require Evidence Generation

FDAAA expands FDA's authority to require postmarket clinical trials.³⁷⁰ Even as it does so, it diminishes the relative dependence on clinical trial data (both pre- and postmarket) by supplementing them with additional forms of evidence during the postmarket period.

366 See, e.g., IOM, *FUTURE OF DRUG SAFETY*, *supra* note 23, at 105–06; Greenfield & Kravitz, *supra* note 266, at 120.

367 See generally FDA, *INNOVATION OR STAGNATION*, *supra* note 143, at 16–25 (calling for increased use of safety and efficacy biomarkers); U.S. Food & Drug Admin., Critical Path Initiative, <http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/default.htm> (last visited Oct. 26, 2009) [hereinafter FDA, Critical Path Initiative] (describing FDA's Critical Path Initiative, one of the goals of which is to increase the use of biomarkers in ascertaining the safety and effectiveness of drugs).

368 See discussion *supra* Parts II–III (discussing the methodological, practical, and ethical limitations of premarket drug trials).

369 Melmon, *supra* note 37, at 144.

370 FDAAA § 901(a), 21 U.S.C.A. § 355(o)(3) (West Supp. 2009) (letting FDA require postmarket studies or clinical trials of drugs with known or suspected safety problems); *id.* § 902(a), 21 U.S.C.A. § 352(z) (West Supp. 2009) (letting drugs be regarded as misbranded if the manufacturer fails to carry out required postmarket studies).

Observational studies appear set to become the main workhorse of postmarket evidence development after FDAAA.³⁷¹

In the past, FDA claimed that it already possessed statutory authority to order postmarket studies, including clinical trials.³⁷² This authority was not expressly stated, so FDA exercised it gingerly.³⁷³ FDA's accelerated approval regulations,³⁷⁴ issued in 1992, let the agency require postmarket studies³⁷⁵ in one special circumstance: when surrogate endpoints³⁷⁶ of effectiveness are used to speed the approval of drugs that address unmet needs for treating serious and life-threatening conditions. For example, a cancer drug might be approved based on proof that it shrinks tumor size (a surrogate endpoint), without proof that it actually improves patients' overall health and survival (which are the clinically meaningful measures of effectiveness).³⁷⁷ For drugs so approved, FDA could require postmarket studies to confirm efficacy but not safety.³⁷⁸ FDA declined to set specific timetables for completion of these studies³⁷⁹ and many

371 See *id.* § 901(a), 21 U.S.C.A. § 355(o)(3)(D) (requiring alternatives such as observational studies and studies with Sentinel System data to be rejected before a postmarket clinical trial can be ordered).

372 See New Drug, Antibiotic, and Biological Drug Product Regulations; Accelerated Approval, 57 Fed. Reg. 58,942, 58,953–54 (Dec. 11, 1992) (discussing, in preambles to FDA's proposed and final rulemaking on accelerated approval, statutory provisions that imply a power for FDA to order postmarket studies); Expanded Availability of Investigational New Drugs Through a Parallel Track Mechanism for People with AIDS and Other HIV-Related Disease, 57 Fed. Reg. 13,234, 13,236–58 (Apr. 15, 1992); see also Levitt et al., *supra* note 20, at 179 (“Although the FDA historically had no specific authority to require postmarketing studies as a condition of approval, the agency cited sections 505(k) and 701(a) of the FDCA in support of this practice. Section 505(k) grants the FDA authority over the establishment and maintenance of records necessary to determine whether there are grounds for withdrawing a [previously granted approval], while section 701(a) gives the agency the authority to promulgate regulations for the efficient enforcement of the FDCA.”).

373 See *id.*; see also New Drug, Antibiotic, and Biological Drug Product Regulations; Accelerated Approval, 57 Fed. Reg. at 58,954 (noting past cases in which FDA conditioned new drug approvals on postmarketing studies).

374 21 C.F.R. §§ 314.500–.560 (2009) (for drugs); *id.* §§ 601.40–.46 (for biologics).

375 *Id.* §§ 314.510, 601.41.

376 See *supra* note 240 and accompanying text.

377 See CTR. FOR DRUG EVALUATION & RESEARCH, U.S. FOOD & DRUG ADMIN., GUIDANCE FOR INDUSTRY: CLINICAL TRIAL ENDPOINTS FOR THE APPROVAL OF CANCER DRUGS AND BIOLOGICS 3–5 (2007), available at <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071590.pdf>.

378 See 21 C.F.R. §§ 314.510, 601.41.

379 New Drug, Antibiotic, and Biological Drug Product Regulations; Accelerated Approval, 57 Fed. Reg. at 58,954 (describing FDA's rationale for declining to impose sanctions for failure to complete postmarket studies).

of them were never completed.³⁸⁰ The Food and Drug Administration Modernization Act of 1997³⁸¹ (FDAMA) contained Fast Track provisions confirming authorities similar to those FDA had claimed in its 1992 accelerated approval regulations.³⁸²

FDAAA expands FDA's authority to require postmarket safety studies irrespective of whether a drug was approved on an expedited basis (accelerated approval/Fast Track) or under FDA's regular procedures.³⁸³ Such studies can be ordered after approval if new safety information comes to light.³⁸⁴ These studies can include clinical trials or alternatives including observational studies; the alternatives must be considered first and found insufficient before a clinical trial can be ordered.³⁸⁵ FDA can require studies to be completed on a definite timetable³⁸⁶ although there is a "good cause" defense if the timetable is not met.³⁸⁷ These provisions remove one barrier—vague statutory authority—that previously blocked efforts to improve the quality of evidence available in the postmarket period. However, there is a second major barrier: lack of infrastructure.

2. Shared Infrastructure for Evidence Development

FDAAA treats evidence generation as a vast, shared enterprise. FDAAA rejects the idea that it is enough simply to order individual drug manufacturers to conduct discrete, isolated studies of the risks and benefits of their drugs. Making drugs safe and effective requires shared informational infrastructure, cooperative efforts, and inputs from the public and from diverse stakeholders throughout the health-care industry.

In the past, large-scale observational studies have tended to be infeasible because there was little infrastructure in place to support them.³⁸⁸ Systems to support one postmarket drug study often would

380 See Susan Okie, *What Ails the FDA?*, 352 *NEW ENG. J. MED.* 1063, 1065 (2005).

381 Pub. L. No. 105-115, 111 Stat. 2296 (codified at scattered sections of 21 U.S.C.).

382 21 U.S.C. § 356(b)(2) (2006).

383 See FDAAA § 901(a), 21 U.S.C.A. § 355(o)(3) (West Supp. 2009) (allowing FDA to require studies to assess known risks, unexpected risks, and risk signals—that is, potential new risks that are suspected based on observed clinical outcomes in patients using the drug).

384 *Id.*, 21 U.S.C.A. § 355(o)(3)(C).

385 *Id.*, 21 U.S.C.A. § 355(o)(3)(D)(ii).

386 *Id.*, 21 U.S.C.A. § 355(o)(3)(E)(ii).

387 *Id.*; see also Kessler & Vladeck, *supra* note 11, at 489 (noting that the requirements for a showing of good cause are not yet clear).

388 U.S. Food & Drug Admin., U.S. Dep't of Health & Human Servs., Proceedings, Sentinel Network Public Meeting 85 (Mar. 8, 2007) [hereinafter FDA, March 8 Pro-

have potential use in other studies. For example, software capable of aggregating Vioxx adverse events across ten large insurance databases also could be used to study adverse events with other drugs. As with many other types of networked infrastructure (pipelines, telecommunication systems, power grids), the efficient solution is to develop a common system that reduces duplicative capital investment.³⁸⁹ Coordinating such an effort requires regulatory powers that FDA did not have in the past.

To the extent FDA had the power to order a postmarket study under the 1962 paradigm, this was a power to order a particular manufacturer to conduct a study of a specific drug. There was no statutory basis to require cooperative efforts or sharing of infrastructure among manufacturers, and FDA lacked jurisdiction over many of the entities (physicians, pharmacists, insurers, health-care providers) that hold relevant data. Drug manufacturers were responsible for conducting and financing postmarket studies.³⁹⁰ To do an observational study, a manufacturer first would have to construct suitable information systems. By analogy, it was as if each time a merchant wished to ship a product from New York to California, she had to factor the cost of constructing the interstate highway network into the cost of the shipment. Observational studies, when used at all, tended to be small in scale because of this problem.³⁹¹ Proposals to offer reductions in tort liability as an inducement for manufacturers to provide better information about postmarket drug safety³⁹² generally have not addressed the feasibility of their doing so, in the absence of appropriate infrastructure. Even if the spirit were made willing (through offers of preemption or other new defenses), the infrastructure was weak.

The centerpiece of FDAAA is a program to create a nationally scaled health data network³⁹³ to support postmarket observational studies.³⁹⁴ This network, known as the Sentinel System,³⁹⁵ has been

ceedings], *available at* <http://www.fda.gov/ohrms/dockets/dockets/07n0016/07n0016-tr00002.pdf> (statement of Dr. Tom Gross).

389 CHARLES F. PHILLIPS, JR., *REGULATING PUBLIC UTILITIES* 51–54 (3d ed. 1993).

390 FDA, March 8 Proceedings, *supra* note 388, at 85 (statement of Dr. Tom Gross).

391 *Id.*

392 *See, e.g.,* Nagareda, *supra* note 3, at 45 (proposing to induce drug manufacturers to provide information about drug safety risks by offering them the prospect of preemption from tort lawsuits as they do so).

393 FDAAA § 905(a), 21 U.S.C.A. §§ 355(k)(3)–(4) (West Supp. 2009).

394 Though ostensibly a safety network, safety is defined in FDAAA in a way that will let efficacy failure be treated as a safety issue. *See id.* §§ 901(b), 905(a), 21 U.S.C.A. 355-1(b)(1) (West Supp. 2009), 355(k)(3)(C)(i)(II). Thus, Sentinel System data can be used for both safety and efficacy studies. Evans, *supra* note 6, at 601.

described elsewhere³⁹⁶ and only summary information is provided here. The network is to be populated with patients' Medicare, military, and private insurance claims data, health records, and pharmaceutical purchase data.³⁹⁷ Congress's goal is to include data for twenty-five million people by July 2010 and one hundred million people by July 2012.³⁹⁸ Steps already have been taken to facilitate FDA's access to Medicare data³⁹⁹ and data from the Veterans' Health Administration,⁴⁰⁰ which will suffice to reach the twenty-five-million-person level.⁴⁰¹ FDA is pursuing voluntary arrangements with private health insurers and could reach the one-hundred-million-person level with data from about ten large insurers.⁴⁰² An even larger system is technically feasible over the longer term⁴⁰³ and would produce better evidence of variable drug response among population subgroups.⁴⁰⁴

Crucially, FDA's access to these data does not require individual consent of the persons whose data are involved.⁴⁰⁵ Congress rejected a consent-based model and instead drew on legal traditions allowing unconsented use of data for public health purposes such as tracking epidemics. FDAAA orders FDA to carry out certain public health responsibilities, including specific, enumerated types of postmarket studies.⁴⁰⁶ As a governmental body carrying out a statutory public health mandate, FDA can obtain data without individual privacy

395 See U.S. Food & Drug Admin., FDA's Sentinel Initiative, <http://www.fda.gov/Safety/FDASentinelInitiative/default.htm> (last visited Nov. 3, 2009); see also FDA, SENTINEL INITIATIVE, *supra* note 137 (describing the purpose and structure of the Sentinel Initiative).

396 FDA, SENTINEL INITIATIVE, *supra* note 137; see also Evans, *supra* note 6, at 588, 601–04 (discussing various features of the proposed Sentinel network).

397 FDAAA § 905(a), 21 U.S.C.A. §§ 355(k)(3)(C)(i)(III)(aa)–(cc).

398 *Id.*, 21 U.S.C.A. § 355(k)(3)(B)(ii).

399 Medicare Program; Medicare Part D Claims Data, 73 Fed. Reg. 30,664, 30,664 (May 28, 2008) (to be codified at 42 C.F.R. § 423).

400 FDA, SENTINEL INITIATIVE, *supra* note 137, at 18.

401 HHS, New Efforts, *supra* note 7.

402 See FDA, March 7 Proceedings, *supra* note 227, at 73 (statement of Dr. Richard Platt).

403 *Id.*

404 See *id.* at 72–73 (statement of Dr. Miles Braun).

405 See Evans, *supra* note 6, at 602–03, 610–26 (discussing privacy authorization requirements for release of Sentinel data); *id.* at 626–31 (discussing informed consent requirements); see also KRISTEN ROSATI, AN ANALYSIS OF LEGAL ISSUES RELATED TO STRUCTURING FDA SENTINEL INITIATIVE ACTIVITIES 86–87 (2009), available at <http://www.regulations.gov/search/Regs/contentStreamer?objectId=090000648098bad2&disposition=attachment&contentType=pdf> (concluding also that individualized informed consent and privacy authorizations are not required).

406 See *infra* notes 414–22 and accompanying text.

authorizations, invoking an exception to the HIPAA Privacy Rule's authorization requirements.⁴⁰⁷ Moreover, FDA's informed consent regulation applies only to clinical trial subjects and would not apply, unless amended, to persons whose data are used in observational studies.⁴⁰⁸ Even if FDA were to place the system under the ethical norms of the Federal Policy for the Protection of Human Subjects of Biomedical Research (Common Rule),⁴⁰⁹ these norms afford various avenues to use Sentinel System data in studies without individual consent.⁴¹⁰

The complex ethical and privacy issues this presents have been explored at length elsewhere.⁴¹¹ In effect, Congress determined that the public health benefits of having this system are weighty enough to override individuals' interest in keeping their data out of it. A consent-based model would have rendered a system of this scale infeasible and would have undercut its evidentiary value. For example, a consent-based model invites selection biases, since it produces a data set of self-selected consenters. It fails to produce the broad inclusiveness required to support timely, valid conclusions about how patients are reacting to a particular drug. It lacks the scale to detect rare reactions in population subgroups and may lack the continuity to detect long-term risks and benefits. Under FDAAA and federal privacy law, it will be lawful for FDA to use people's data in postmarket drug safety surveillance without individual authorization or consent provided the data are used in ways that fit within the scope of uses Congress authorized in section 905 of FDAAA.⁴¹²

These allowed uses are broad but not unlimited.⁴¹³ They include identifying and reporting drug safety risks;⁴¹⁴ conducting safety surveillance⁴¹⁵ (that is, monitoring how products are performing in actual clinical use to detect signals that may suggest emerging safety problems); identifying trends and preparing regular reports on adverse event trends;⁴¹⁶ and exporting data for further analysis.⁴¹⁷

407 See 45 C.F.R. pts. 160, 164 (2009); *id.* § 164.512(b)(1)(i) (providing an exception to privacy authorization requirements for governmental public health authorities collecting data pursuant to statutory mandates); *see also* Evans, *supra* note 6, at 597 n.73 (summarizing other potentially applicable exceptions).

408 See Evans, *supra* note 6, at 626–27.

409 45 C.F.R. §§ 46.101–124 (2009).

410 Evans, *supra* note 6, at 629–31.

411 See *id.* at 622–31.

412 See *id.* at 601–02.

413 *Id.* at 612–13.

414 FDAAA § 905(a), 21 U.S.C.A. §§ 355(k)(3)(C)(i)(I)–(II) (West Supp. 2009).

415 *Id.*, 21 U.S.C.A. § 355(k)(3)(C)(i)(III).

416 *Id.*, 21 U.S.C.A. §§ 355(k)(3)(C)(i)(IV)–(V).

FDA is authorized to collaborate with public, private, and academic entities to conduct advanced analyses of drug safety data⁴¹⁸ which could include methodological studies (for example, to develop better analytical tools) and substantive studies of “advanced safety questions.”⁴¹⁹ This last term is broadly defined and would encompass studies of drug efficacy as well as of safety.⁴²⁰ Advanced drug safety studies potentially include research uses of Sentinel System data as well as studies that fit within the concept of public health studies.⁴²¹ Congress delegated wide discretion for FDA to determine whether proposed studies do or do not fit within the authorized categories of use.⁴²²

The Sentinel System’s heavy reliance on administrative data (such as insurance and Medicare claims) has advantages and disadvantages. Administrative data increasingly are used in observational studies because they are widely available and relatively inexpensive to analyze.⁴²³ They cover large, diverse groups of real people, for example, everybody who is enrolled with a particular insurer.⁴²⁴ This minimizes selection biases, since the effects of taking a particular drug can be observed in a wide variety of old, young, high- and low-risk patients.⁴²⁵ Administrative data reflect actual clinical effectiveness rather than theoretical efficacy in the artificial conditions of clinical trial.⁴²⁶ However, these data have known weaknesses. Entities, such as healthcare providers and insurers, that maintain these databases generally “lack motivation to record information that does not have an immediate impact on reimbursement.”⁴²⁷ Thus, insurance claims data may record that a person had a laboratory test performed on a certain date but fail to include the test result.⁴²⁸ There are coding inconsistencies (the same medical condition may be recorded differently in different databases)⁴²⁹ and biases (such as upcoding, where the severity of a patient’s illness is overstated in order to qualify for higher reimburse-

417 *Id.*, 21 U.S.C.A. § 355(k)(3)(C)(i)(VI).

418 *Id.*, 21 U.S.C.A. § 355(k)(4)(A).

419 *Id.*, 21 U.S.C.A. § 355(k)(4)(A)(ii).

420 Evans, *supra* note 6, at 601.

421 *Id.* at 622–23.

422 *Id.* at 601, 612.

423 Roos et al., *supra* note 107, at 47.

424 *Id.* at 48.

425 *Id.* at 52.

426 *Id.* at 50–51.

427 *Id.* at 48.

428 *Id.* at 56.

429 FDA, March 7 Proceedings, *supra* note 227, at 31 (statement of Dr. Jeffrey Hill).

ment).⁴³⁰ Insured people generally tend to be healthier than uninsured people, and this can cause administrative data to understate poor health outcomes.⁴³¹

Observational studies of drug safety will require linking data across separate administrative databases (for example, linking insurance claims data with clinical data from various sources to follow the entire course of a patient's illness, treatment, and outcome; or linking data across insurers to follow patients through changes in their health plans). Linkage is technically difficult⁴³² and greater degrees of linkage increase privacy concerns.⁴³³ In response, FDA has stated that, at least initially, it will not adopt a centralized architecture for the system and will leave raw data where they currently reside (for example, with an insurer). "Questions would be sent to appropriate, participating data sources, who in turn would, in accordance with existing privacy and security safeguards, evaluate their data and send results for Agency review."⁴³⁴ This approach aids privacy protection but may produce lower-quality evidence than could be achieved at a higher level of data integration.⁴³⁵ To achieve Congress's objectives, FDA will need to embrace high levels of data linkage at some point and, when it does so, the privacy and data security arrangements will become a topic of crucial public interest.

Administrative data alone cannot resolve questions about risks and benefits of drugs. Administrative data can reveal, for example, that a patient was treated for a stroke after having purchased a certain drug. On average, about 80% of these coincidences turn out to be something other than a drug-related adverse event.⁴³⁶ Establishing whether the drug may have caused the stroke requires follow-up analysis, often including a review of the patient's full medical record.⁴³⁷ Thus, FDAAA envisions that FDA may obtain other types of data, if the

430 Roos et al., *supra* note 107, at 57.

431 *Id.*

432 Evans, *supra* note 6, at 594–95.

433 See Mark A. Rothstein, *Health Privacy in the Electronic Age*, 28 J. LEGAL MED. 487, 489 (2007); Nicholas P. Terry & Leslie P. Francis, *Ensuring the Privacy and Confidentiality of Electronic Health Records*, 2007 U. ILL. L. REV. 681, 700.

434 U.S. Food & Drug Admin., U.S. Dep't of Health & Human Servs., Solicitation No. HHS-FDA-RFI-09-002, FDA Sentinel Initiative: Request for Information (Feb. 20, 2009), [hereinafter FDA, Request for Information], available at <https://www.fbo.gov/index?id=5182af215e556265fde72a8e969fa46e>.

435 Evans, *supra* note 6, at 606.

436 FDA, March 7 Proceedings, *supra* note 227, at 57 (statement of Dr. Marc Overhage).

437 See *id.* at 57–58; *id.* at 66–68 (statement of Dr. Richard Platt).

Secretary of HHS deems the data necessary.⁴³⁸ In theory this could entail access to patients' whole medical records or previously stored tissue specimens.⁴³⁹ FDA has indicated it does not plan to invoke this power on a routine basis. At least during the system's early years, the plan seems to be to perform observational studies using available administrative data, with access to full health records only occasionally and selectively, to clarify the causes of specific observed events.

Large-scale observational studies have enormous potential as an evidentiary tool. In 2008, the Institute of Medicine noted the need, in an age of genomic medicine, for follow-up studies to detect unknown benefits as well as unknown risks.⁴⁴⁰ The Sentinel System can help detect both. In addition to speeding the detection of drug safety problems, it will offer, for the first time, a workable tool to resolve nagging questions about efficacy. Today, these questions include the uncertain efficacy of drugs prescribed off-label and drugs approved based on surrogate endpoints—questions that, in practice, often are deemed not worth the cost of a clinical trial. In the future, these questions increasingly will include uncertainties about the long-term efficacy of predictive and preventive medicines which deliver benefits too remote in time to be observed in traditional clinical trials.

B. The Flexible Best Evidence

The second pillar of FDA's new evidentiary paradigm is that there is no best evidence—at least, there is no single source of evidence or methodological “gold standard” that is best for all seasons and all medical products. The Sentinel System will make it possible to do observational studies that were not feasible in the past. This means that, for many questions, more than one study methodology will now be available, necessitating choices about when to answer questions and which methodology to use. The routine choices FDA will be making for specific drugs in specific contexts will have important ramifications for the public's health and also may have major commercial impacts on drug manufacturers. These respective interests may be in conflict. Appropriate administrative procedures for making these choices are not yet resolved. This is a matter with which legal scholars need to engage to ensure that FDA's decisional processes adequately protect the public's interests.

438 FDAAA § 905(a), 21 U.S.C.A. §§ 355(k)(3)(C)(i)(III)(aa)–(cc) (West Supp. 2009).

439 Evans, *supra* note 6, at 588.

440 IOM, *DIFFUSION & GENOMIC INNOVATIONS*, *supra* note 143, at 24.

1. Deciding When to Generate Evidence

One of FDA's most crucial determinations under FDAAA will be when to resolve uncertainty about safety and effectiveness. The key decision here is whether to answer questions premarket or postmarket.⁴⁴¹ FDA already makes these types of determinations today, whenever it approves a premarket clinical trial design that relies on surrogate endpoints⁴⁴² or intermediate clinical endpoints⁴⁴³ which, while reflecting some degree of meaningful health improvement, do not ensure a good ultimate outcome. Even before 1992, when FDA's regulations first addressed the use of surrogate endpoints, FDA had been using them to approve new drugs.⁴⁴⁴ "Reliance on a surrogate endpoint almost always introduces some uncertainty into the risk/benefit assessment"⁴⁴⁵ and defers questions for later resolution. Clinical trial endpoints set the bar of what must be known before a product moves into clinical use. For example, will the manufacturer only have to prove that the drug reduces patients' blood cholesterol levels, or must they show that the drug reduces actual accumulations of plaque in their arteries, or that it reduces the number of strokes and heart attacks they suffer, or that it actually increases longevity and health?

FDA has insisted that "this judgment does not represent either a 'lower standard' or one inconsistent with section 505(d) of the act [that is, with the statutory requirement for proof of safety and effi-

441 *Id.* at 62.

442 See Janet Woodcock, Acting Deputy Comm'r for Operations, Food & Drug Admin., Presentation, *A Framework for Biomarker and Surrogate Endpoint Use in Drug Development*, at slide 6 (Nov. 4, 2004), available at http://www.fda.gov/ohrms/dockets/ac/04/slides/2004-4079S2_03_Woodcock.ppt (defining surrogate endpoint as a "biomarker intended to substitute for a clinical endpoint" and noting that a surrogate endpoint "is expected to predict clinical benefit (or harm, or lack of benefit) based on epidemiologic, therapeutic, pathophysiological or other scientific evidence"); *supra* note 240 (discussing surrogate endpoints).

443 See Woodcock, *supra* note 442, at slide 5 (defining a clinical endpoint as a "characteristic or variable that reflects how a patient feels, functions, or survives").

444 New Drug, Antibiotic, and Biological Drug Product Regulations; Accelerated Approval, 57 Fed. Reg. 58,942, 58,944 (Dec. 11, 1992) (discussing, in the preamble to the accelerated approval rulemaking, the history of FDA's use of surrogate biomarkers). Under its traditional approval regulations, FDA can approve a product based on surrogate endpoints if the surrogate is already known to be a good predictor of favorable clinical outcomes (morbidity and mortality). Under accelerated approval, FDA can use surrogate endpoints even if the surrogate is only "reasonably likely" to predict good clinical outcomes. *Id.*; see also 21 C.F.R. §§ 314.510, 601.41 (1992) (detailing criteria for approval based on surrogate endpoints for efficacy).

445 New Drug, Antibiotic, and Biological Drug Product Regulations; Accelerated Approval, 57 Fed. Reg. at 58,944.

cacy].”⁴⁴⁶ Even if one accepts that assertion, it is undeniable that using surrogate endpoints punts uncertainties into the postmarket period. FDA’s Critical Path Initiative envisions that, in the future, FDA’s approval process will make greater use of surrogate endpoints for both safety and efficacy.⁴⁴⁷

There is much to be said on both sides of this issue. Some of the potential concerns, after FDAAA, are these: As postmarket evidence development improves, will the sheer availability of postmarket study options create temptation to defer more and more questions to be resolved after drug approval? What is the irreducible minimum of information that must be known about a drug before it goes on the market? How much residual risk is too much to defer for postmarket evaluation? Congress delegated these questions to FDA’s discretion. It is not clear how that discretion will be constrained and insulated from commercial influences.

On the plus side, the use of surrogate endpoints may make it possible for FDA to address a longstanding problem that Congress was unwilling to confront. For many years, the 1962 Drug Amendments have been criticized⁴⁴⁸ as a costly (over \$800 million per drug) way to generate dubious evidence of safety and efficacy (see Parts II and III above) while delaying patients’ access to desired therapeutic possibilities (as in *Abigail Alliance v. von Eschenbach*⁴⁴⁹) and deterring innovation to develop new therapies (as reflected in the declining number of new drugs put forward for approval⁴⁵⁰). FDAAA does not directly address whether the 1962 Amendments were fundamentally ill conceived. For those who believe they were, FDAAA merely spreads an additional layer of regulation—postmarket evidence development—over the failed edifice of the 1962 evidentiary paradigm (a remediation concept akin to the Chernobyl sarcophagus), as if more regulation can solve the problems of defective regulation. The right approach, all along, may have been to allow drugs onto the market based on evidence of basic safety plus plausible grounds to expect effi-

446 *Id.*

447 See FDA, INNOVATION OR STAGNATION, *supra* note 143, at 16–25.

448 See, e.g., Roberts & Bodenheimer, *supra* note 14; Lechter, *supra* note 90; Califf, *supra* note 25 (criticizing impacts of the 1962 Amendments on drug access, drug development costs, and pharmaceutical innovation).

449 495 F.3d 695 (D.C. Cir. 2007), *cert. denied*, 128 S. Ct. 1069 (2008); see also *supra* note 201 and related text (discussing *Abigail Alliance*).

450 FDA, INNOVATION OR STAGNATION, *supra* note 143, at 2 fig. 2 (showing a declining ten-year trend in innovative therapies submitted for FDA approval); see also Epstein, *supra* note 43, at 32 (“The pace of innovation was without question greater in the era of less regulation.”).

cacy, subject to rigorous postmarket studies to assess—quickly—whether the expectations at approval are borne out, with decisive actions taken when they are not. FDAAA creates infrastructure that would support rapid postmarket assessments of this type. However, Congress did not make corresponding adjustments to reduce the cost, access delays, and other problems of FDA’s existing premarket evidentiary requirements.

Statutory reform of these requirements likely was seen as politically infeasible, particularly in the post-Vioxx environment. The public likes the idea of FDA’s premarket clinical trial process, whatever its methodological limitations may be (and these are poorly understood). Reforming the threshold requirements for drug approval, overtly via statute, would have been a recipe for public outrage. Congress avoided the issue. This implicitly delegated to FDA the task of determining how FDA’s premarket evidence requirements might be adapted in the post-FDAAA world where postmarket studies are a realistic option. FDA already has the power to make this determination through its routine, day-to-day decisions concerning suitable endpoints for premarket clinical trials. As a science-based, discretionary determination, these decisions will be virtually unreviewable. Yet they will be a fulcrum of public protection after FDAAA. The process through which they are made and their impacts need to be transparent.

2. Choices Among Study Methodologies

Once a drug has been approved, the critical issue will be which study methodology to use to resolve lingering uncertainties and, in particular, when to insist on postmarket RCTs.⁴⁵¹ Pharmacogenomics often is described as “the right drug, at the right dose, for the right patient.”⁴⁵² FDAAA takes a similar view of regulatory evidence during the postmarket period: getting the best evidence of benefits and risks requires the right methodology, at the right time, for the right product. FDA’s traditional premarket study requirements continue in effect and will provide a certain “floor” of risk-benefit information that will be available for all approved drugs (although, as has been the case before FDAAA, this floor will be somewhat moveable based on the choice of clinical trial endpoints). After approval, the situation grows even more fluid. FDAAA does not establish any generally applicable requirements for postmarket study methodologies.

⁴⁵¹ IOM, *DIFFUSION & GENOMIC INNOVATIONS*, *supra* note 143, at 62.

⁴⁵² This phrase has been so endlessly repeated that it is difficult to determine who first said it.

FDAAA recognizes that different study methodologies may supply the best evidence at different points in a drug's life cycle. After approval, the preferred methodology for further studies may vary from drug to drug and may depend on the type of risk being studied (for example, how rare the risk is, or how serious). Data from prospective RCTs are not always the best evidence.⁴⁵³ Clinical trials study a proxy for reality—a small trial population—rather than the large, diverse population that is actually consuming a drug. Clinical trials make sense in the premarket period, when no other clinical data exist. However, large observational studies of people who actually took the drug may be the best way to study safety after approval.⁴⁵⁴ Postmarket observational studies may reveal a risk faster than a phase IV clinical trial could have done.⁴⁵⁵ Clinical trials and observational studies both may be needed during the postmarket period to get a full picture of a drug's risks and benefits.⁴⁵⁶ Observational studies offer advantages for studying long-term outcomes in the real-world clinical setting⁴⁵⁷ and they may be superior if the risk being studied would make an RCT unethical.⁴⁵⁸ Postmarket RCTs do offer advantages for certain questions,⁴⁵⁹ but they are not always feasible.⁴⁶⁰ In 2008, an Institute of Medicine workshop on genomic innovation noted the need for flexible evidentiary standards that employ a combination of methodologies.⁴⁶¹ FDAAA embraces this view.

FDA has discretion to decide the best way to clarify risks and benefits after drugs are approved. FDAAA calls for FDA's Drug Safety and Risk Management Advisory Committee (or its successor) to provide guidance to the agency on priority safety questions and the best ways

453 IOM, *DIFFUSION & GENOMIC INNOVATIONS*, *supra* note 143, at 64.

454 FDA, March 7 Proceedings, *supra* note 227, at 4 (statement of Dr. Andrew von Eschenbach).

455 See *supra* notes 250–51 and related text (discussing speed with which Vioxx risks could have been detected using large-scale observational techniques).

456 See IOM, *FUTURE OF DRUG SAFETY*, *supra* note 23, at 105–06; FURBERG & FURBERG, *supra* note 60, at 17.

457 Weisman et al., *supra* note 95, at 129.

458 FURBERG & FURBERG, *supra* note 60, at 29.

459 IOM, *FUTURE OF DRUG SAFETY*, *supra* note 23, at 105–06 (discussing advantages of postmarket trials to confirm efficacy of drugs approved using surrogate efficacy biomarkers and to assess causation of common adverse events (such as a heart attack) which might have been caused either by the drug or by the patients' underlying condition).

460 Wylie Burke, *General Observations*, in IOM, *DIFFUSION & GENOMIC INNOVATIONS*, *supra* note 143, at 81, 82; see also IOM, *DIFFUSION & GENOMIC INNOVATIONS*, *supra* note 143, at 64 (“Not everything will be tested.”).

461 Burke, *supra* note 460, at 82.

to study them.⁴⁶² The options may include studies using Sentinel System data, other post-approval studies, or postmarket clinical trials.⁴⁶³ FDAAA calls for these decisions to be made in a “public process.”⁴⁶⁴ What this means is unclear. These decisions will involve complex scientific judgments that few members of the public can wield. It may be hard to enlist public involvement when, as noted earlier, many members of the public are under the impression that FDA-approved drugs have no risks.⁴⁶⁵

Before FDAAA, the public has shown very little interest in the related decisions FDA makes about the choice of endpoints in premarket trials. Is it plausible that the public will engage with decisions about the timing and methodology for postmarket studies? Even if the public does engage, its involvement may be ill-informed and unconstructive. Caskey has noted a tendency of patient advocacy groups to view evidence generation with suspicion, seeing evidence as a tool for denial of insurance reimbursement rather than as a tool for making treatments safer and more effective.⁴⁶⁶ This phenomenon was seen again recently when patient advocacy groups protested provisions of the American Recovery and Reinvestment Act of 2009⁴⁶⁷ (ARRA) that would fund studies of the comparative effectiveness of drugs.⁴⁶⁸ Such studies aim to protect patients by assessing which drugs work and which do not work, but there was alarm that these data might justify denial of insurance coverage for the underperforming drugs. Why the protestors desired to protect their access to badly performing drugs was not explained.⁴⁶⁹

If the public will not or cannot engage, then what is an appropriate regulatory response? Should the government appoint—and pay

462 FDAAA § 905(a), 21 U.S.C.A. § 355(k)(4)(C) (West Supp. 2009).

463 *Id.*

464 *Id.*

465 See IOM, UNDERSTANDING BENEFITS, *supra* note 16, at 1, 7.

466 C. Thomas Caskey, *The Drug Development Crisis: Efficiency and Safety*, 58 ANN. REV. MED. 1, 10 (2006).

467 Pub. L. No. 111-5, 123 Stat. 115.

468 See, e.g., Alicia Mundy, *Drug Makers Fight Stimulus Provision*, WALL ST. J., Feb. 10, 2009, at A4 (describing concerns that comparative effectiveness research may lead to rationing of health care); Betsy McCaughey, *Ruin Your Health With the Obama Stimulus Plan*, BLOOMBERG, Feb. 9, 2009, http://www.bloomberg.com/apps/news?pid=20670001&refer=columnist_mccaughey&sid=aLzFDxfbwhzs (expressing concern that comparative effectiveness provisions of ARRA will enforce excessive uniformity and erode patient care).

469 *But see* Mundy, *supra* note 468 (pointing out that some advocacy groups whose names may suggest they are patient advocacy organizations actually are lobbying arms of the drug, device, and biotechnology industries).

for—a scientific panel with a specific duty to represent the public's interests in these discussions? The pharmaceutical industry will be represented by well qualified scientists of its own. FDA, as regulator, ordinarily would be responsible for protecting the public's interest; however, it may be some years before the agency regains sufficient public trust to serve credibly in this role.

If meaningful public representation is not possible, can the impact of FDA's choices at least be made more transparent? For example, FDA could disclose more clearly what is not known about the safety and effectiveness of drugs when they are approved, and what FDA has decided to leave unknown by not ordering particular types of study. Today's drug labeling tells only what is known about a drug's risks and benefits but does not give a sense of all that is still unknown. This has contributed to a culture of mass drug marketing and consumption in which people are eager to get the latest drug, often believing that it must be better and safer than older drugs when in fact, such comparative data rarely exist.⁴⁷⁰ FDA's regulatory process, to date, has failed to convey a proper sense of its own uncertainty. "Humility about the limits of science is critical to enjoy the trust and respect of the public."⁴⁷¹ There are still large gaps in knowledge.⁴⁷² FDAAA aims to reduce those gaps through postmarket evidence development. As it does so, the effort needs to embrace procedures that improve public awareness of the gaps still there.

C. *Successive Improvement of the Risk-Benefit Ratio*

FDAAA rejects the notion that the risk-benefit ratio is a fixed attribute of the drug product itself. The risk-benefit ratio is dynamic and can be successively improved throughout a drug's commercial life.⁴⁷³ Safety and efficacy are no longer attributes of the drug itself, but of the drug and the decisional process through which it is prescribed. In principle, almost any drug can be made safer and more effective by using biomarker tests to screen patients and predict which people will respond favorably to that drug.⁴⁷⁴ Obviously, some drugs are more promising candidates for these strategies than other

470 CONG. BUDGET OFFICE, RESEARCH ON THE COMPARATIVE EFFECTIVENESS OF MEDICAL TREATMENTS 4 (2007), available at <http://www.cbo.gov/ftpdocs/88xx/doc8891/12-18-ComparativeEffectiveness.pdf>.

471 IOM, UNDERSTANDING BENEFITS, *supra* note 16, at 24 (reporting presentation of Dr. Dennis J. Paustenbach).

472 *Id.* at 25.

473 Evans, *supra* note 222, at 791–92.

474 George W. Sledge, *What is Targeted Therapy?*, 23 J. CLINICAL ONCOLOGY 1614 (2005).

drugs.⁴⁷⁵ When a drug is teamed with a screening test that accurately predicts safety and/or effectiveness, the result is a successor product (the drug-test hybrid) that offers a new risk-benefit ratio superior to that of the stand-alone drug.

FDAAA recognizes that successive improvement requires contributions from parties other than the drug manufacturer. FDA's 1962 evidentiary paradigm relied on drug manufacturers to conduct research to establish drugs as safe and effective.⁴⁷⁶ Their work was seen as substantially complete at the point of FDA approval⁴⁷⁷ when they had produced a drug with a passable average risk-benefit ratio. Unfortunately, manufacturers' incentives to make successive improvements are rather weak after drug approval.⁴⁷⁸ Sales of drugs to non-responders are highly remunerative, accounting for an estimated 40% of overall pharmaceutical industry revenues at present.⁴⁷⁹ A rational manufacturer would tend not to invest in research to cut its own revenues. Such research bestows positive externalities on other parties, such as sellers of screening tests and doctors whose drug-injury liabilities might be reduced if adverse responders were screened out.⁴⁸⁰ FDA's regulations have not had—and even after FDAAA will not have—an effective cost-spreading mechanism to let drug manufacturers recover the costs of postmarket research from all the other parties who stand to gain from an improved risk-benefit ratio.⁴⁸¹

475 See Lawrence J. Lesko & Janet Woodcock, *Translation of Pharmacogenomics and Pharmacogenetics: A Regulatory Perspective*, 3 NATURE REV. 763, 767 (2004) (noting that screening strategies are most promising for drugs that have a narrow therapeutic range, large individual variability of response, and serious safety problems).

476 FURBERG & FURBERG, *supra* note 60, at 7.

477 Evans, *supra* note 222, at 791–92.

478 For discussion of incentives for pharmacogenomic research by drug manufacturers, see generally, Peakman & Arlington, *supra* note 224, at 35–40 (discussing future marketing challenges and legal liability issues that may confront pharmaceutical companies that do not provide tailored medicines); Lembit Rāgo, *Pharmacogenomics and Existing Therapies*, 17 WHO DRUG INFO. 84 (2003) (noting a trend by regulatory authorities to require pharmaceutical companies to identify which patients are most likely to benefit from a drug); Michael Dickson et al., *Survey of Pharmacoeconomic Assessment Activity in Eleven Countries* (Org. for Econ. Co-operation & Dev., Working Paper No. 4, 2003), available at <http://www.oecd.org/dataoecd/27/25/2955828.pdf> (discussing the use of pharmacoeconomic assessment by OECD in determining which drugs are reimbursed through public health programs).

479 Barbara J. Evans et al., *Creating Incentives for Genomic Research to Improve Targeting of Therapies*, 10 NATURE MED. 1289, 1289 (2004) (basing weighted average estimate over all classes of drugs on Peakman & Arlington statistics for non-response rates and HHS estimates for annual prescription drug expenditures).

480 See Evans & Flockhart, *supra* note 81, at 56.

481 *Id.*

FDAAA seeks to engage parties other than the drug manufacturer in the process of improving the safety and effectiveness of FDA-approved drugs. Such parties include, for example, academic researchers; clinical laboratories and test manufacturers; developers of screening strategies, bioinformatics algorithms, and software; and health care providers.⁴⁸² In the past decade, academic researchers and other third parties (such as genetic test developers) have played key roles in identifying genetic and molecular screening strategies to use with preexisting drugs.⁴⁸³ Discovery of biomarkers and tests to predict drug response “is often undertaken outside of the company developing the drug.”⁴⁸⁴ This pattern of third-party discovery is expected to continue.⁴⁸⁵ “Networked discovery”⁴⁸⁶ also is common and involves cooperative work among test developers, drug companies, academic and clinical laboratories, health care providers and payers, and other entities.⁴⁸⁷

Third-party discovery requires policies to “bridge the gap between the basic and clinical world, as well as to connect academic and industrial realms.”⁴⁸⁸ The Institute of Medicine has noted a need for greater coordination to accelerate progress in biomarker discovery⁴⁸⁹ and has called for new models of cooperation.⁴⁹⁰ FDA’s ability to foster the required cooperation has been a concern. Many of the parties just discussed are not subject to FDA regulation. FDAAA does not change FDA’s jurisdiction or extend it to new entities not already regulated by FDA. However, without regulating the parties themselves, FDA still can take steps to remove barriers that stand in their way. FDAAA contains provisions aimed at encouraging cooperative activities, including the access provisions discussed in Part V.D and the risk management provisions discussed in Part V.G.

482 Evans, *supra* note 222, at 787–88.

483 See Lesko & Woodcock, *supra* note 475, at 767 (citing example); Sledge, *supra* note 474, at 1614 (discussing the discovery of new strategies for targeting older breast cancer therapies).

484 IOM, BIOMARKER-BASED TOOLS, *supra* note 174, at 21 (reporting presentation of Dr. Paul Waring).

485 Evans, *supra* note 222, at 757.

486 *Id.* at 788 fig. 5.

487 *Id.*; see also IOM, BIOMARKER-BASED TOOLS, *supra* note 174, at 24–26 (reporting presentation of Dr. Paul Waring) (discussing challenges of cooperation between drug and test developers).

488 IOM, BIOMARKER-BASED TOOLS, *supra* note 174, at 60 (reporting remarks of Dr. Michael E. Phelps).

489 *Id.* at 57–58 (reporting presentation of Dr. Stephen Friend).

490 IOM, DIFFUSION & GENOMIC INNOVATIONS, *supra* note 143, at 21.

D. Putting Evidence Into the Hands of Innovators

Problems with data access have been a major barrier to successive improvement of drugs by third parties (parties other than the drug manufacturer).⁴⁹¹ Data on clinical experience with drugs are necessary to detect variations in response and to explore reasons for the variations. FDA's old evidentiary paradigm relied heavily on drug manufacturers to collect data and conduct postmarket studies.⁴⁹² As a result, much of what is known about the clinical performance of drugs is in proprietary data sets not available to other researchers. Even when data theoretically are available (for example, by collecting adverse event statistics directly from health care providers) the cost of assembling them can be prohibitive, especially for thinly capitalized test developers and academic researchers. Investment in duplicative data sets wastes funds that otherwise could be spent validating new screening strategies and developing new tests.

A crucial feature of the Sentinel network is its access provisions. FDAAA lets FDA engage outside entities to help establish the Sentinel network.⁴⁹³ FDAAA also lets FDA involve public, academic, and private entities in various functions that use Sentinel data, such as classifying, analyzing, or aggregating Sentinel data; investigating priority drug safety questions; and performing advanced studies and analysis of safety risk.⁴⁹⁴ The entities involved in establishing the network can be, but need not be, the same entities engaged to analyze the data. Thus, FDA could collaborate with insurers or other holders of large datasets to establish the system and engage those same entities to perform studies and analytical functions with the data. However, FDA also could engage entirely different entities to conduct studies and analysis. FDA has wide discretion to determine which entities are appropriate partners for which functions.

FDA has expressed its intent to adopt a decentralized phase I network architecture in which organizations that possess data would respond to queries from FDA and return responses to the agency in

491 *Id.* (noting that “the system prevents open exchange of information and creates many barriers to communication among affected parties”); *see also* IOM, *BIOMARKER-BASED TOOLS*, *supra* note 174, at 57 (reporting presentation of Dr. Stephen Friend) (calling for more access to patient materials and data); Evans, *supra* note 222, at 788 (discussing various legal and regulatory barriers to cooperation).

492 FDA, March 8 Proceedings, *supra* note 388, at 85 (statement of Dr. Tom Gross).

493 FDAAA § 905(a), 21 U.S.C.A. § 355(k)(3)(C)(v) (West Supp. 2009).

494 *Id.*, 21 U.S.C.A. § 355(k)(4)(A), 355(k)(4)(D)(i).

aggregated, de-identified form.⁴⁹⁵ Under this design, FDA would be engaging these organizations to provide both data access and analytical services to respond to FDA's queries. However, FDA also has legal authority to approve use of the Sentinel network to respond to queries submitted by outside entities, so long as their queries fit within the scope of authorized uses of Sentinel data described in FDAAA.⁴⁹⁶ From a legal standpoint, it would be lawful for FDA to approve release of Sentinel data either to the agency itself or to outside users that FDA has engaged to do analytical work.⁴⁹⁷ These users could be parties other than the organizations that possess the data. However, the agency is keenly aware of the privacy issues this presents and has indicated it does not plan to move data away from their current locations.⁴⁹⁸ The more likely scenario would be for outside users to work with FDA to define queries which, if approved by the agency, would be submitted for analysis behind the privacy firewalls of organizations that possess the data, with answers returned in an aggregated, anonymized form.⁴⁹⁹ FDA is still in the process of developing Sentinel System governance arrangements, including the framework for approving outside data uses (or queries) and the terms and conditions of such uses.⁵⁰⁰ These arrangements obviously are critical to the tasks of protecting privacy, ensuring that data are used ethically, and maintaining public acceptability of all data uses.⁵⁰¹

The effect of these provisions is to empower FDA to engage entities other than those that now possess data to perform drug safety studies and research that use the data. It is fairly easy to criticize these

495 See, e.g., JANET M. MARCHIBRODA, EHEALTH INITIATIVE FOUND., DEVELOPING A GOVERNANCE AND OPERATIONS STRUCTURE FOR THE SENTINEL INITIATIVE 4, 8 (2009), available at <http://www.regulations.gov/search/Regs/home.html#documentDetail?R=09000064809a82f0> (discussing FDA's phase I network architecture in which organizations that now process data would keep the data behind their respective privacy firewalls and respond to queries sent by FDA).

496 FDAAA § 905(a), 21 U.S.C.A. §§ 355(k)(4)(A)(i)–(iii); see also FDA, SENTINEL INITIATIVE, *supra* note 137, at 16 (showing a research component as part of the system's organizational structure).

497 FDAAA § 905(a), 21 U.S.C.A. § 355(k)(3)(C)(i)(VI) (authorizing "export" of data for further aggregation, statistical analysis, or reporting).

498 Evans, *supra* note 6, at 621, 626.

499 See FDA, Request for Information, *supra* note 434.

500 See U.S. Food & Drug Admin. et al., Sentinel Initiative: Structure, Function, and Scope (Dec. 16, 2008) (transcript available at <http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/CriticalPathInitiative/PastEventsonCPI/UCM113469.pdf>) (exploring Sentinel System governance and privacy issues at a day-long public workshop in Washington, D.C.).

501 Evans, *supra* note 6, at 639–53 (discussing Sentinel System governance arrangements).

access provisions on ethical or privacy grounds. It is perhaps harder to understand why Congress saw them as a necessary part of the new evidentiary paradigm. Making drugs safer and more effective requires making evidence accessible to the parties most likely to pursue that goal. FDAAA breaks FDA's sole dependence on the drug manufacturer as a source of discovery and innovation in response to emerging evidence. It does not compel third parties to innovate, but enables them to do so. If drug manufacturers lack incentives to improve the risk-benefit ratio of their approved drugs, then FDA can arrange third-party access to Sentinel data for uses that serve that goal. Under FDA's phase I architecture, this "access" likely would take the form of approving queries from qualified third parties with which FDA is collaborating. The possibility of engaging outside parties to help improve drug safety may, in itself, help improve incentives for drug manufacturers to do so, knowing that others may enter the field if they fail.

Frankly, the concern is not whether FDA will over-share access to Sentinel data. The concern is that FDA may fail to ensure the level of access Congress envisioned. In recent years FDA has tended to favor a consensual model of regulation, a sort of regulation by consent of the regulated, and has shown a certain reluctance to flex compulsory power.⁵⁰² In line with this preference, FDA has been pursuing voluntary arrangements to get private holders of data (such as large insurers, healthcare providers, and pharmaceutical companies) to participate in the Sentinel network.⁵⁰³ This approach may place FDA under pressure to agree to terms that restrict third-party queries. For example, data holders might agree to run queries only for other data holders that have contributed data; they may wish to exclude queries by entities that have not contributed data of their own. Such restrictions would perpetuate access barriers for test developers and other

502 See, e.g., *id.* at 631–33 (discussing FDA's traditional reliance on voluntary regulatory compliance); *infra* Part V.G.1 (discussing FDA's voluntary approach to cross-labeling issues).

503 See, e.g., MARCHIBRODA, *supra* note 495, at 6–7, 13 (asserting that FDA lacks authority to compel data access); FDA, Request for Information, *supra* note 434 ("Data will continue to be managed by its owners, and only data or organizations who agree to participate in this system will be included."). *But see* Barbara J. Evans, *Authority of the Food and Drug Administration to Require Data Access and Control Rights in the Sentinel Data Network*, 65 *FOOD & DRUG L.J.* (forthcoming Feb. 2010) (manuscript at 7–46, available at <http://ssrn.com/abstract=1508672>) (analyzing FDA's statutory power to require data access and concluding that FDAAA gives FDA authority to require access to data, even if the agency may prefer, for policy reasons, to pursue consensual approaches).

third parties that may have little data to contribute, but who could make productive use of drug-related data if granted access.

FDA must resist pressure to accept terms that thwart Congress's purpose, which was to provide data access to those best able to improve drug safety. FDA may need to flex the full array of compulsory powers available to it. The quest for regulation by consensus is a laudable one unless it thwarts achievement of Congress's public health objectives. Appropriate access to Sentinel data was an essential feature of Congress's overall postmarket regulatory design.

Sentinel data access appears set to emerge as a fruitful subject of scholarship and litigation once the system is up and running. Given the breadth of FDA's discretion to approve or disapprove proposed uses of Sentinel data (i.e., to approve which queries can be run), will access be nondiscriminatory and will the access procedures be clearly enunciated? What recourse will be available if the agency fails to implement appropriate third-party access? In the tort litigation context, there will be questions about the discoverability of the underlying Sentinel System data themselves, in situations where FDA has used these data to generate risk-benefit information that is publicly reported in summary form.⁵⁰⁴ These and many other access-related issues are ripe for study as is the problem of protecting individual privacy in various contexts of access and disclosure.

E. Making Efficacy Regulation Effective

FDAAA's handling of efficacy is revolutionary. The revolution is in a definitional section where it is easy to overlook. For the first time in history,⁵⁰⁵ Congress has instructed FDA to concern itself with individual variations in a drug's effectiveness and to treat these variations as a potential safety issue in their own right. This aspect of FDAAA has significant potential to influence broader legal norms surrounding the sale and consumption of drugs. For example, if FDA regards efficacy failure as a safety problem, will courts eventually recognize it as a basis for tort liability?

504 See FDAAA § 915, 21 U.S.C.A. § 355(r)(2)(C) (West Supp. 2009) (calling for summaries of Sentinel System findings to be published on FDA's Internet-based communication system described *infra* Part V.G.1); see also Mark A. Hall et al., *Measuring Medical Practice Patterns: Sources of Evidence from Health Services Research*, 37 WAKE FOREST L. REV. 779 (2002) (discussing the related problem of tapping large databases used in health services research for use as evidence of the standard of care in malpractice cases).

505 See discussion *infra* Part V.F.

FDAAA achieves this change through a subtle, three-step sleight of definition.⁵⁰⁶ The change is almost invisible in the text. First, Congress defines a new term, “adverse drug experience,”⁵⁰⁷ which includes the usual things FDA traditionally counted as drug-related adverse events plus “any failure of expected pharmacological action of the drug.”⁵⁰⁸ This last prong would let non-responders⁵⁰⁹ be counted as having had an adverse drug experience. Non-responders are neither hurt nor helped by a drug. Because they are unhurt, they traditionally were not seen as having suffered an adverse event. Yet, obviously, it is an adverse experience to take a drug and still be sick. Moreover, it is a common experience: efficacy failure is estimated to occur with 30–60% of the prescriptions written in the United States.⁵¹⁰ Second, Congress defined a “serious” adverse drug experience as one that kills or places the patient at immediate risk of death, results in persistent or significant incapacity or substantial disruption of ability to conduct normal life functions, or which may require medical or surgical intervention to prevent such outcomes.⁵¹¹ Third, Congress closed the definitional loop by equating “serious risk” to risk of a serious adverse drug experience.⁵¹²

These little changes give FDA big new powers. The agency can tackle the problem of non-response with the same tools it has available for managing drug safety problems. These tools include, for example, the power to order postmarket studies, labeling changes, and risk mitigation steps described *infra*.⁵¹³ In situations where patients’ non-response to a drug entails potentially serious consequences, such as death or significant incapacity, FDA can address the

506 FDAAA §§ 901(b), 905(a), 21 U.S.C.A. §§ 355-1(b)(1), 355(k)(3)(C)(i)(II).

507 *Id.* § 901(b), 21 U.S.C.A. § 355-1(b)(1).

508 *Id.*, 21 U.S.C.A. § 355-1(b)(1)(E).

509 *See supra* fig. 1 (schematically depicting non-responders as the 40% of patients in the center of the chart who received neither benefits nor harms from the drug).

510 *See* Peakman & Arlington, *supra* note 224; *see also* Kenneth R. Evans, Ontario Cancer Biomarker Network, Presentation, *Challenges and Potential in Biomarker Discovery and Development*, at slide 3 (on file with author) (reporting a non-response rate of 40–60% for depression drugs; a 4–75% non-response rate for asthma drugs; a 50–75% non-response rate for certain diabetes drugs (sulfonurea, biguanides, and glitazones), and a 70–90% non-response rate for various drugs used in treating breast, lung, and brain cancers); Henderson & Reavis, *supra* note 315, at 2 fig. 1 (showing non-response rates for fourteen major classes of drugs and reporting that non-response rates vary from 20% to around 75%, with most drugs having non-response rates in the 40–60% range).

511 FDAAA § 901(b), 21 U.S.C.A. § 355-1(b)(4)(A).

512 *Id.*, 21 U.S.C.A. § 355-1(b)(5).

513 *See* discussion *infra* Part V.G.2.

problem of non-response just as it would address any other safety issue. This empowers FDA to conduct postmarket surveillance of efficacy—a crucial power in the age of genomic medicine, when many products will be entering the market with unresolved questions about their long-term efficacy.

Historically, FDA has viewed non-response as an efficacy issue but not a safety issue in its own right. This view is imbedded in the risk-benefit methodology FDA uses to assess safety when approving new drugs. This methodology (which still is in effect as of today) requires reviewers to distinguish two types of adverse events: (1) those that are attributable to progression of the underlying disease (that is, to non-response), and (2) those that are caused by the drug itself.⁵¹⁴ Only the latter events are counted in the risk-benefit ratio FDA uses to assess whether the drug is safe enough to approve.⁵¹⁵ Injuries caused by non-response are deliberately ignored. Yet these harms are potentially serious. Though not directly poisoned by the drug, non-responders may suffer “lost-chance”⁵¹⁶ injuries (opportunity costs), including profound ones. Cancer patients may die. Diabetes patients may suffer kidney failure, loss of eyesight, or amputations. The potential for lost-chance injuries exists whenever a patient has a progressive disease and multiple treatment options are available. Taking an ineffective drug implies foregoing other therapies that might have halted or slowed the disease progression.

FDAAA may or may not lead to changes in the risk-benefit methodology FDA applies during drug approval, but it will change the way non-response is handled after approval. This approach (waiting until the postmarket period to address the safety impacts of non-response) makes sense. These impacts arise when a drug is effective on average (and hence approvable) but fails to work in some patients. In other words, the potential for harm reflects individual variability in response to the drug. Data on individual variability often are of poor quality at the point when drugs are first approved. Clinical trials, with their small sample sizes and short duration, may spot some instances of non-response and may provide some insight into the health consequences. However, much larger sample sizes are needed to under-

514 See CDER, MAPP 6010.3, *supra* note 216, §§ 7.1, 7.1.5.5; FDA, CONDUCTING A CLINICAL SAFETY REVIEW, *supra* note 142, at 5–6, 8, 12–14.

515 See *id.*; see also Evans, *supra* note 222, at 762–64 (discussing this problem).

516 “Lost-chance” doctrine, which is recognized in some states, lets patients bring tort suits when they have suffered irreversible disease progression as a result of a negligent error or delay in treating or diagnosing their disease. See, e.g., Martin J. McMahon, Annotation, *Medical Malpractice: Measure and Elements of Damages in Actions Based on Loss of Chance*, 81 A.L.R. 4th 485 (1990).

stand the full range of variability, and longer follow-up is needed to assess how non-response ultimately affects health outcomes. This problem lends itself to large postmarket observational studies. In all likelihood, FDA will continue to approve drugs with significant rates of non-response, just as it has done in the past. However, FDAAA adds non-response to the list of issues to be monitored and managed in the postmarket period. That is the right time to address this problem, since that is when high-quality regulatory evidence first becomes available.

F. *Piercing the Veil of Average Safety and Efficacy*

In the twentieth century, FDA restricted its focus to average safety and efficacy.⁵¹⁷ This was driven partly by scientific limitations of the day, but another factor was at work: There was legislative intent for FDA not to concern itself with individual variations in drug response. Congress expressed this intent clearly when enacting the 1962 amendments. By authorizing FDA to concern itself with variability of treatment response,⁵¹⁸ FDAAA redraws a forty-five-year-old jurisdictional boundary.

The scope of FDA's power to regulate medical practice, a traditional area of state regulation, was a "hot-button" issue as Congress debated passage of the 1938 Act.⁵¹⁹ Congress disclaimed intent for FDA's regulation of medical products to entail broad regulation of medical practice.⁵²⁰ As a matter of policy, FDA subsequently sought to avoid regulating physicians' activities.⁵²¹ This policy may account for FDA's failure to address the safety impacts of non-response between 1938 and 1962. The 1938 Act authorized FDA to oversee drug safety but not drug efficacy.⁵²² This would have let FDA address non-response, but only to the extent that it was viewed as a safety problem.

517 See discussion *supra* Part IV.B.

518 See discussion *supra* Part V.E.

519 See Joel E. Hoffman, *Administrative Procedures of the Food and Drug Administration*, in 2 FUNDAMENTALS OF LAW AND REGULATION, *supra* note 20, at 13, 17–24; Evans & Flockhart, *supra* note 81, at 50–51 (discussing legislative debate in the late 1930s).

520 Legal Status of Approved Labeling of Prescription Drugs; Prescribing for Uses Unapproved by the Food and Drug Administration, 37 Fed. Reg. 16,503 (Aug. 15, 1972) (discussing, in the preamble to a proposed rulemaking, Congress's legislative intent in passing the FDCA).

521 *Id.*; see also David G. Adams, *The Food and Drug Administration's Regulation of Health Care Professionals*, in 2 FUNDAMENTALS OF LAW AND REGULATION, *supra* note 20, at 423 ("FDA has traditionally taken the position that it does not regulate the practice of medicine or pharmacy and has generally avoided regulatory actions that would directly restrict or interfere with professional service to patients.").

522 See discussion *supra* Introduction.

It was indeed a safety problem then as now. However, non-response was (and still is) primarily an efficacy issue. From the early twentieth century up until the 1962 amendments went into effect, questions of efficacy were regarded as the province of physicians and the medical profession.⁵²³ Nongovernmental bodies such as the American Medical Association operated programs to assess drug efficacy and disseminate this information to their members.⁵²⁴ To have addressed safety impacts of non-response, FDA would have had to intrude on “efficacy space” already occupied by physicians. FDA declined to do so.

The 1962 amendments at last authorized FDA to address questions of efficacy. Before these amendments were passed, however, there was a legislative struggle over what “efficacy” means. During this struggle, Congress parsed efficacy into two aspects, only one of which would be subject to federal oversight. In 1962, legislators understood that efficacy assessments involve subjective judgments⁵²⁵ and that unambiguous proof of efficacy rarely exists. Senators Dirksen and Hruska sought assurance that FDA would not require unanimous or even preponderant evidence of efficacy before approving a drug.⁵²⁶ In response, the 1962 amendments required only “substantial” evidence of efficacy⁵²⁷—a position advocated by the pharmaceutical industry.⁵²⁸ Congress expressed its intent that: (1) this standard should not imply “identical results for different patients”⁵²⁹ and (2) physicians, rather than FDA, should be the ultimate arbiters of what is effective for a particular patient.⁵³⁰ Thus, Congress was fully aware of individual variability of drug response but steered FDA away from engaging with this problem.

In effect, Congress deemed average efficacy to be a product regulatory issue subject to federal regulation, but left individual efficacy in the realm of state medical practice regulation. Two factors explain

523 See Lechter, *supra* note 90, at 144 (citing PAUL STARR, *THE SOCIAL TRANSFORMATION OF AMERICAN MEDICINE* 131–33 (1982)).

524 *Id.*

525 *Id.* at 156.

526 *Id.*

527 S. REP. NO. 1744, *supra* note 14, at 2921.

528 See Julie A. Grow, *The Legislative History of the 1962 Drug Amendments* 32 (May 1, 1997) (unpublished manuscript), available at <http://leda.law.harvard.edu/leda/data/189/jgrow.pdf> (citing *Hearings before the Subcomm. on Antitrust and Monopoly of the Senate Comm. on the Judiciary, pursuant to S. Res. 52 and S. 1552, Pt. 4, 87th Cong. 1997* (1961) (statement of Eugene N. Beesley, President, Eli Lilly & Co.), reprinted in 19 U.S. FOOD & DRUG ADMIN., *A LEGISLATIVE HISTORY OF THE FEDERAL FOOD, DRUG, AND COSMETIC ACT AND ITS AMENDMENTS* 810–11 (1979)).

529 S. REP. NO. 1744, *supra* note 14, at 2921.

530 *Id.* at 2920–21.

Congress's decision. The first was the primitive state of mid-twentieth-century science, which saw variable response as a chance event, beyond human control and, therefore, not something that could be improved through regulatory oversight.⁵³¹ Second, the scope of federal power to regulate medical practice was a concern in 1962 just as it had been in 1938.⁵³² The American Medical Association opposed the 1962 amendments.⁵³³ Making drugs effective at the level of individual patients requires physician involvement.⁵³⁴ Even if FDA could describe specific types of people likely to respond badly or favorably to a drug, physicians still would need to assess whether patients met those criteria and they would need to prescribe drugs appropriately in light of that information.⁵³⁵ To ensure individual efficacy, the agency seemingly would need a mechanism for enforcing physicians' compliance with its approved product labeling. In 1962, Congress was unwilling to assert that federal jurisdiction extended that far. In fairness, Congress's decision did not significantly diminish public health or safety: The science of that day would not have supported meaningful regulation of individual effects, even if Congress had been comfortable with the jurisdictional issues it presented.⁵³⁶

Twentieth-century law meekly accepted science's incapacity to predict whether drugs will or will not be effective for particular patients. This resignation was seen in Congress's parsing of efficacy in 1962, and it is reflected in the risk-benefit methodology FDA still uses today to approve new drugs. It has influenced broader legal doctrine (such as tort liability standards that presently impose few duties to ensure the efficacy of health care) and commercial norms in the health care industry (such as payment and insurance reimbursement

531 Evans & Flockhart, *supra* note 81, at 47.

532 S. REP. NO. 1744, *supra* note 14, at 2920–21.

533 Lechter, *supra* note 90, at 148; *see also* Grow, *supra* note 528, at 15–17 (recounting testimony given on behalf of AMA during hearings before the Subcommittee on Antitrust and Monopoly of the Senate Committee on the Judiciary as the 1962 amendments were considered).

534 *See Hearings before the Subcomm. on Antitrust and Monopoly of the Senate Comm. on the Judiciary, pursuant to S. Res. 52 and 5. 1552, pt. 1, 87th Cong. 37 (1961)* (statement of Dr. Hugh H. Hussey, Chairman, Bd. of Trustees, Am. Med. Ass'n), *reprinted in* 17 U.S. FOOD & DRUG ADMIN., A LEGISLATIVE HISTORY OF THE FEDERAL FOOD, DRUG, AND COSMETIC ACT AND ITS AMENDMENTS 608 (1979).

535 *See IOM, UNDERSTANDING BENEFITS, supra* note 16, at 16–17 (“Accurately identifying such populations involves significant physician involvement, over which . . . FDA has little control. While . . . FDA can define benefits and risks for different populations, it cannot prevent the inappropriate prescribing by physicians once a drug is on the market.”).

536 *See* Evans & Flockhart, *supra* note 81, at 47.

standards that require no refunds when treatments fail). FDAAA asserts that it is time to do something about treatment failure. This is revolutionary; the revolution was timely if not overdue. It remains to be seen whether, how, and when FDAAA's new view of therapeutic non-response may rub off on other areas of health law. The immediate impact is that efficacy failure is now, at least for postmarket regulatory purposes, a legitimate drug safety issue and Congress, for the first time, sees individual/subgroup variability as a legitimate concern of federal regulation.

G. Making Evidence Consequential

The seventh pillar of the new paradigm is expanded regulatory decisionmaking authority in the postmarket period. The postmarket evidence FDA was gathering before FDAAA was "seriously flawed"⁵³⁷ and has been criticized as giving an incomplete picture of the safety of approved drugs.⁵³⁸ It deserves an even harsher criticism: it was largely inconsequential. Congress had given FDA very little power to order action in response to emerging postmarket evidence. Without such power, postmarket evidence is idle knowledge. FDAAA grants FDA a bundle of new powers to apply postmarket evidence to improve the public's health.

1. Communicating Evidence to the Public and Physicians

FDAAA lets FDA require safety-related labeling changes at any point in a drug's life cycle.⁵³⁹ Before FDAAA, FDA had control over initial product labeling, but if new risks emerged after product

537 IOM, UNDERSTANDING BENEFITS, *supra* note 16, at 52.

538 See IOM EMERGING SAFETY SCIENCE, *supra* note 80, at 79 (reporting presentation of Dr. William DuMouchel) (pointing out that FDA's MedWatch postmarket drug safety database relies on voluntary reports by patients and physicians and thus includes only an inaccurate "numerator" (how many incidents were voluntarily reported) and no "denominator" (the total number of people who took the drug)); Duh et al., *supra* note 63, at 31 (pointing out that when FDA tried to supplement physician-reported data through voluntary agreements to access data from Health Maintenance Organization databases, the resulting data had the customary limitations seen with administrative datasets); see also Pillans, *supra* note 62, at 697 (pointing out that the "numerator" was susceptible to underreporting); *id.* at 698 (discussing voluntary adverse-drug-reaction reporting systems in various nations).

539 See FDAAA § 901(a), 21 U.S.C.A. § 355(o)(4) (West Supp. 2009) (letting FDA notify manufacturers of safety information it believes should be included in drug labeling and order the change following a period for response and discussions with the manufacturer).

approval, FDA could only encourage voluntary⁵⁴⁰ labeling changes by threatening to withdraw its previously-granted approval.⁵⁴¹ This threat rang hollow for risks that were not serious enough to justify removing a drug from the market. Even for serious risks, this approach was problematic if the drug was providing therapeutic benefits for at least one patient subgroup.⁵⁴² The situation shown schematically in Figure 1 is actually quite common. A drug may be delivering benefits to some patient subgroups even while endangering others. If a late-discovered risk affects only a subset of patients taking the drug, withdrawing the approval would protect some patients at the cost of denying therapeutic benefits to others who have come to rely on the drug.⁵⁴³ Withdrawal would not necessarily improve the public's health, so FDA's threats had little credibility if manufacturers declined to make requested labeling changes.

FDA now has authority to order safety-related labeling changes. Since FDAAA views efficacy failure as a potential safety problem, this seemingly could include labeling changes to address emerging problems with individual or average efficacy. During the first six months after these provisions went into effect, FDA ordered four safety-related labeling changes.⁵⁴⁴ While helpful, this new power is only a partial solution.

Traditional labeling is not an ideal medium for communicating the information doctors will need in the twenty-first century.⁵⁴⁵ For

540 Levitt et al., *supra* note 20, at 178; see also I. Scott Bass, *Enforcement Powers of the Food and Drug Administration: Drugs and Devices*, in 2 FUNDAMENTALS OF LAW AND REGULATION, *supra* note 20, at 55, 70–74 (discussing the scope and limits of FDA's recall authority).

541 21 U.S.C. § 355(e) (2006). A complete list of grounds for withdrawal is set forth in 21 C.F.R. § 314.150 (2004).

542 See Epstein, *supra* note 43, at 25 (calling it “utterly indefensible” to remove a drug from the market in situations where there is a way to distinguish in advance the portion of the total population that is at risk for an adverse reaction, and noting that in these situations the “key question is to develop some protocol that allows for the identification of these high risk users, usually a minority”).

543 See, e.g., John Leland, *Pain Pills Withdrawn, Many Renew Search for Relief*, N.Y. TIMES, Mar. 6, 2005, at A30 (describing hardships for patients for whom Cox-2 painkillers had been effective, when some of these products were removed from the market due to cardiovascular risks).

544 U.S. FOOD & DRUG ADMIN., U.S. DEP'T OF HEALTH & HUMAN SERVS., FDAAA IMPLEMENTATION—HIGHLIGHTS ONE YEAR AFTER ENACTMENT 5 (2008) [hereinafter FDA, FDAAA IMPLEMENTATION], available at <http://www.fda.gov/downloads/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDCAAct/SignificantAmendmentstotheFDCAAct/FoodandDrugAdministrationAmendmentsActof2007/UCM083181.pdf>.

545 Evans, *supra* note 222, at 780–82.

example, doctors increasingly are being asked to apply complex genetic and biomarker screening strategies to predict which patients are suitable candidates for particular drugs. To succeed, they need detailed, application-specific, up-to-date information about the best screening strategies and tests to use with particular drugs.⁵⁴⁶ In recent years, this led to calls for cross labeling of tests and drugs. Cross labeling exists when a drug's labeling identifies specific screening tests and provides information on how to vary doses in response to test results, or when a test's labeling describes how to use the test to screen patients for a specific drug.⁵⁴⁷ Cross labeling is what most of us would like to have, when a drug is being given to us or to our loved ones, yet very few cross-labeled products exist.⁵⁴⁸ It is more common for drug labeling—if it includes any information about screening strategies at all—simply to note that patient response may vary based on certain genetic or other factors, without recommending specific tests to use or explaining how to vary dosing in response to the test results.⁵⁴⁹

As of 2005, FDA seemed unsure of its authority to force drug and test manufacturers to cross label their products, if either party objected to doing so.⁵⁵⁰ Segments of the industry voiced strong opposition to mandatory cross labeling, citing problems with apportion-

546 FDA, March 7 Proceedings, *supra* note 227, at 7–8 (statement of Dr. Janet Woodcock) (discussing the need for the full weight of scientific advances to be brought to bear on treatment decisions and the need for up-to-date, accurate information at the point of patient care).

547 Evans, *supra* note 222, at 785–87.

548 *Id.* at 780. *But see, e.g.*, Genentech, Full Prescribing Information for Herceptin (2009), available at <http://www.gene.com/gene/products/information/oncology/herceptin/insert.jsp> (cross-referencing specific, named tests to use with the drug).

549 *See, e.g.*, Eli Lilly & Co., Full Prescribing Information for Strattera, §§ 12.3, 17.7 (2009), available at <http://pi.lilly.com/us/strattera-pi.pdf> (noting that the drug is metabolized primarily through the cytochrome P450 2D6 enzymatic pathway and commenting that dose may need to be adjusted when the drug is used along with certain other drugs that inhibit this pathway, but not naming a specific test); *see also* Lesko & Woodcock, *supra* note 475, at 766 (discussing factors FDA considered in deciding how to address gene-drug interactions in atomoxetine (Strattera) labeling).

550 *See* Evans, *supra* note 222, at 785–86; *see also* Drug Info. Ass'n & U.S. Food & Drug Admin., Proceedings, Combination Products and Mutually Conforming Labeling Workshop (May 10, 2005) [hereinafter DIA & FDA, May 10 Proceedings], *transcript available at* <http://www.fda.gov/CombinationProducts/MeetingsConferences/Workshops/ucm135152.htm> (statements of Nancy Stade, Office of the Chief Counsel, U.S. Food & Drug Admin., and Suzanne O'Shea, Office of Combination Products, U.S. Food & Drug Admin.) (discussing whether FDA can approve one company's product without voluntary conformity by another company).

ment of product liability and various commercial concerns.⁵⁵¹ Even if FDA did have such authority (or if the parties voluntarily agreed to work together), there are evidentiary barriers to cross labeling. Most of the available screening tests are CLIA-regulated LDTs.⁵⁵² Because these tests do not pass through a rigorous, data-driven regulatory review before they go on the market, many of them are unable to meet FDA's evidentiary standards for including information in drug labeling. Another concern is whether drug labeling can be revised quickly enough to keep pace with the rapid evolution of new, better screening tests.⁵⁵³ Continual revisions would be required, and labeling changes are not free.⁵⁵⁴

FDA indicated in 2005 that it would encourage voluntary cooperation among drug and test manufacturers⁵⁵⁵ but would stop short of forcing them to cross label their products. Whatever the merits of this policy, it poses obstacles to commercialization of third-party screening tests for use with preexisting drugs⁵⁵⁶ and has been a barrier to third-party participation in efforts to make drugs safer and more effective.⁵⁵⁷ It also leaves doctors groping for information about which test is best to use with which drug. In FDAAA, Congress could have granted FDA authority to require compulsory cross labeling. Congress did not do so. It instead pursued a multifaceted solution that acknowledges: (1) the underlying gaps in evidence about screening strategies and tests for use with drugs, (2) the need for timely communication of rapidly evolving advice.

By improving postmarket evidence development generally,⁵⁵⁸ FDAAA creates infrastructure that can be used to develop new screening strategies and to validate specific tests (including CLIA-regulated LDTs) for use with particular drugs. By doing postmarket observational studies, FDA will spot drugs that have high rates of adverse- or non-responders. People in these subgroups could be approached to obtain their consent for follow-up studies to identify genetic and other

551 See DIA & FDA, May 10 Proceedings, *supra* note 550 (statements of Drs. Anna Longwell, Combination Prods. Coalition, & David Eveleth, Exec. Dir. of Med. & Developmental Scis., Pfizer, Inc.).

552 SACGHS, U.S. OVERSIGHT, *supra* note 223, at 39.

553 Evans, *supra* note 222, at 781.

554 Prescription Drug User Fee Rates For Fiscal Year 2009, 73 Fed. Reg. 45,017, 45,022 (Aug. 1, 2008) (quoting a fee of \$623,600 in fiscal year 2009 for a supplemental application to amend an existing new drug application, if the supplemental application requires review of clinical data).

555 Evans, *supra* note 222, at 781.

556 *Id.* at 786.

557 See discussion *supra* Part V.C.

558 See discussion *supra* Part V.A.

factors that may explain their responses to the drug. Such studies would fit under the rubric of “advanced drug safety studies” envisioned by FDAAA. Thus FDA has power to authorize the use of Sentinel System data in these studies, although individual consents still would be required to collect new specimens for genetic analysis. When there is insufficient evidence about the clinical utility of existing screening tests, the Sentinel System could be used to conduct observational studies comparing outcomes in patients who were and were not screened with particular tests. These and other follow-up studies could help develop hard evidence to support better risk management techniques and, in some cases, labeling changes.

To make communication more timely, FDAAA orders FDA to establish an Internet-based system⁵⁵⁹ for disseminating risk information to patients and health care providers.⁵⁶⁰ The system is to provide an easily searchable,⁵⁶¹ one-stop source of information on drug-related risks and how best to manage them. It already includes information that previously has been available in paper form, such as labeling, package inserts, Medication Guides,⁵⁶² and safety alerts (product recalls and warning letters).⁵⁶³ However, a key purpose of this system is to provide continuous, close-to-real-time feedback of emerging safety data, including early signals of potential problems that are not yet well enough verified to warrant labeling changes or warnings.⁵⁶⁴ Already, the system has begun reporting potential new risks identified using FDA’s pre-FDAAA adverse event reporting system.⁵⁶⁵ Congress has authorized public access to anonymized summaries of Sentinel System safety findings as these become available.⁵⁶⁶ For all newly approved drugs, risk information will be updated within eighteen months or at such time as the drug has been used by 10,000 people.⁵⁶⁷ The system also will provide information on how to manage a drug’s risks, including pre-FDAAA risk management plans that already were

559 The system’s early implementation can be viewed at U.S. Food & Drug Admin., Postmarket Drug Safety Information for Patients and Providers, <http://www.fda.gov/cder/drugSafety.htm> (last visited Dec. 6, 2009).

560 FDAAA § 915, 21 U.S.C.A. § 355(r) (West Supp. 2009).

561 *Id.*, 21 U.S.C.A. § 355(r)(2)(A).

562 *Id.*, 21 U.S.C.A. §§ 355(r)(2)(B)(i)–(ii), 355(r)(3).

563 *Id.*, 21 U.S.C.A. § 355(r)(2)(B)(iv).

564 *Id.*

565 U.S. Food & Drug Admin., Potential Signals of Serious Risks/New Safety Information Identified from the Adverse Event Reporting System (AERS), http://www.fda.gov/cder/aers/potential_signals (last visited Oct. 29, 2009).

566 FDAAA § 915, 21 U.S.C.A. § 355(r)(2)(C).

567 *Id.*, 21 U.S.C.A. § 355(r)(2)(D).

in place for some drugs and the new Risk Evaluation and Mitigation Strategies⁵⁶⁸ envisioned by FDAAA.⁵⁶⁹

FDAAA has not changed FDA's traditional labeling requirements. They will continue in effect and the new Internet-based communication system will exist alongside them. FDA has met its statutory deadline (one year from enactment of FDAAA⁵⁷⁰) to get the system started, but its functionality will be evolving over a period of years. It is still too early to tell how it may evolve. When fully developed, it eventually could supplant traditional labeling as the primary medium patients and physicians rely on for risk-benefit information about drugs.

2. Applying the Evidence

Communicating risk-benefit information will not improve public health, unless the information actually is applied at the point when physicians prescribe drugs. Labeling changes repeatedly have been shown, in empirical studies, to have little impact on physicians' prescribing behavior.⁵⁷¹ Changing FDA's communication medium is unlikely to change that fact, without a workable mechanism to promote physician compliance with warnings, instructions, and risk management strategies. The physician assistant who administered the drug that cost the patient an arm in *Wyeth v. Levine* was prescribing it directly at odds with warnings stated in labeling⁵⁷² and presumably would have been equally unimpressed by a warning delivered over the Internet.

568 See discussion *infra* Part V.G.2; see also U.S. Food & Drug Admin., Approved Risk Evaluation and Mitigation Strategies (REMS), <http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm111350.htm> (last visited Nov. 22, 2009) [hereinafter FDA, REMS] (listing approved REMS plans).

569 FDAAA § 915, 21 U.S.C.A. § 355(r)(2)(B)(v).

570 *Id.*, 21 U.S.C.A. § 355(r)(1).

571 See, e.g., Walter Smalley et al., *Contraindicated Use of Cisapride: Impact of Food and Drug Administration Regulatory Action*, 284 JAMA 3036, 3038 (2000) (finding that labeling revisions and efforts to communicate contraindications of the drug cisapride (Propulsid) had little impact on prescribing behavior); Raymond L. Woosley & Glenn Rice, *A New System for Moving Drugs to the Market*, ISSUES SCI. & TECH. ONLINE, Winter 2005, <http://www.issues.org/21.2/woosley.html> (relating how physician noncompliance with warnings and contraindications in drug labeling ultimately forced manufacturers to remove cisapride (Propulsid), terfenadine (Seldane), astemizole (Hismanal), troglitazone (Rezulin), bromfenac (Duract), and trovafloxacin (Trovan) from the market, even though these drugs would have been safe if instructions in labeling had been heeded).

572 *Wyeth v. Levine*, 129 S. Ct. 1187, 1217–18 (2008) (Alito, J., dissenting).

During the twentieth century, FDA pursued a policy of not regulating physicians.⁵⁷³ This was embodied in the agency's permissive policy on off-label use.⁵⁷⁴ This policy let physicians choose to disregard instructions and warnings in drug labeling. FDA took the position that "labeling is not intended either to preclude the physician from using his best judgment in the interest of his patient, or to impose liability if he does not follow the package insert."⁵⁷⁵ This policy made a certain amount of sense under the 1962 regulatory paradigm, which focused FDA's attention on average safety and efficacy. Unable to provide meaningful guidance about individual safety and efficacy, FDA left this determination to physicians. FDA understood the limitations of its average statistics and this policy, arguably, was as good a solution as any.

FDA's permissive policy on off-label use loses this rationale in a world where FDA is able to provide more nuanced information about benefits and risks. In coming years, FDA will be providing information at the individual/subgroup level for some (not all) drugs. For such drugs, a permissive policy on off-label use is hard to justify. Some off-label uses may be beneficial or at least innocuous (for example, trying out a drug for a new indicated use in patients who, based on screening tests, are at low risk of adverse reactions). Other off-label uses may be overtly dangerous or wasteful (for example, giving it to patients who, based on screening tests, are known to be at high risk of adverse reactions or non-response). Off-label policy itself needs to be nuanced, varying with the specific drug and use.⁵⁷⁶ Certain "bad" off-label uses may need to be banned.⁵⁷⁷

One of the unanticipated outcomes of the Human Genome Project has been to blur the line between medical product and medical practice regulation.⁵⁷⁸ If safety and efficacy are conceived in individual, rather than average, terms, product regulation necessarily touches matters traditionally seen as medical-practice issues.⁵⁷⁹ Prod-

573 See discussion *supra* Parts V.D–V.E.

574 Legal Status of Approved Labeling for Prescription Drugs; Prescribing for Uses Unapproved by the Food and Drug Administration, 37 Fed. Reg. 16,503, 16,504 (Aug. 15, 1972) (stating that labeling is not intended to impede the physician's exercise of judgment concerning what is best for the patient or to impose liability for prescribing decisions that are at odds with drug labeling).

575 *Id.*

576 See Evans, *supra* note 222, at 784.

577 *Id.* at 785.

578 See Evans, *supra* note 293, at 288–93 (discussing how personalized medicine blurs the line between product regulation and practice regulation); see also IOM, UNDERSTANDING BENEFITS, *supra* note 16, at 16–17.

579 Evans, *supra* note 293, at 290, 292.

uct and practice regulation cease to be as cleanly separable as the 1938 Act and the 1962 amendments conceived them to be. Personalized medicine—the use of screening tests to inform individual prescribing decisions—makes FDA a purveyor of information that bears on clinical decisions. The problem is how to ensure this information is put to use.

Despite Congress's delicacy about the matter, courts have never found constitutional limits on FDA's power to regulate physicians.⁵⁸⁰ "There is little doubt under modern law that Congress has ample power to regulate the manufacture, distribution, and use of drugs and medical devices."⁵⁸¹ The jurisdictional lines Congress posited in 1938 and 1962 were the product of intense lobbying by the medical profession⁵⁸² and may have been unduly timid. Nonetheless, FDA actions that influence medical practice remain controversial. The 1976 Medical Device Amendments expressly authorized FDA to approve medical devices subject to restrictions on their use and distribution.⁵⁸³ The older provisions of FDCA, dealing with drugs and biologics, lacked similar provisions. Since 1990, FDA has struggled with how to ensure safe prescribing of drugs for which there are known strategies that could reduce risks.⁵⁸⁴

In 1992, FDA interpreted the FDCA as allowing use restrictions for drugs and biologics.⁵⁸⁵ FDA issued regulations⁵⁸⁶ letting new drugs and biologics be approved subject to "restrictions to ensure safe use"⁵⁸⁷ if FDA found that the product was effective but not safe without restrictions.⁵⁸⁸ These regulations were part of FDA's accelerated approval program (in 21 C.F.R. part 314, subpart H). Unlike FDA's

580 See Adams, *supra* note 521, at 424–25.

581 Epstein, *supra* note 43, at 7.

582 See Adams, *supra* note 521, at 423; Grow, *supra* note 528, at 15–17.

583 Medical Device Amendments of 1976, Pub. L. No. 94-295, § 2, 90 Stat. 539, 565 (adding Section 520(e) of the Food, Drug, and Cosmetic Act) (codified as amended at 21 U.S.C. § 360j(e) (2006)).

584 Scott Gottlieb, *Drug Safety Proposals and The Intrusion of Federal Regulation Into Patient Freedom and Medical Practice*, 26 HEALTH AFFAIRS 664, 665–68 (2007).

585 See New Drug, Antibiotic, and Biological Drug Product Regulations; Accelerated Approval 57 Fed. Reg. 58,942, 58,951 (Dec. 11, 1992) (interpreting existing statutory authority and judicial rulings as allowing FDA to impose postmarketing restrictions on drugs); New Drug, Antibiotic, and Biological Drug Product Regulations; Accelerated Approval, 57 Fed. Reg. 13,234, 13,237 (Apr. 15, 1992) (same).

586 21 C.F.R. §§ 314.520, 601.42 (2009).

587 *Id.* §§ 314.520(a), 601.42(a) (allowing use restrictions, such as limiting distribution to certain facilities or to physicians with special training, or conditioning distribution on the performance of specific medical procedures).

588 *Id.*

authority to require postmarket studies under subpart H, FDA could impose use restrictions regardless whether a drug was approved based on surrogate or clinical endpoints.⁵⁸⁹ FDA expressed its intention to invoke this authority only rarely⁵⁹⁰ and, in subsequent years, has stayed true to this promise. Outside its subpart H authority, FDA also worked with drug sponsors to implement use restrictions as part of voluntary risk management plans for some drugs.⁵⁹¹ In 2005, FDA issued three guidance documents clarifying the correct design of these plans and the circumstances when sponsors should consider putting a plan in place.⁵⁹² These plans have tended to be confined to relatively few products with serious, known side effects.⁵⁹³

FDA's authority to restrict the use of drugs under subpart H was questioned and criticized as an intrusion on the practice of medicine.⁵⁹⁴ Nevertheless, drug sponsors tended not to challenge restrictions FDA imposed when approving new drugs. Faced with the alternative of not having their product approved at all, sponsors generally were willing to agree to proposed restrictions. FDA has tremendous bargaining power at that moment in a drug's life. What was missing, however, was a mechanism to impose restrictions later in a drug's life to manage new risks discovered in the postmarket period.

FDAAA gives FDA clear statutory authority to condition the sale of drugs on specific measures to manage their risks.⁵⁹⁵ The vehicle

589 Under the accelerated approval program, FDA could order postmarket studies only when it approved a product based on surrogate endpoints of effectiveness. See *supra* Part V.B.1. However, the authority to restrict distribution and use was not so restricted. Thus, for example, FDA used this regulation to restrict distribution of the abortion drug mifepristone (RU-486) even though the approval was based on a clinical endpoint (effectiveness in producing an abortion). See Memorandum from Ctr. for Drug Evaluation & Research to NDA 20-687 MIFEPREX (mifepristone) Population Council at 1 (September 28, 2000), available at <http://www.fda.gov/downloads/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm111366.pdf>.

590 New Drug Antibiotic, and Biological Drug Product Regulations; Accelerated Approval, 57 Fed. Reg. 58,942-43 (Dec. 11, 1992).

591 Gottlieb, *supra* note 584, at 666.

592 See Press Release, U.S. Food & Drug Admin., FDA Issues Final Risk Minimization Guidances (March 24, 2005), available at <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2005/ucm108425.htm>.

593 Gottlieb, *supra* note 584, at 669-70.

594 See New Drug, Antibiotic, and Biological Drug Product Regulations; Accelerated Approval, 57 Fed. Reg. at 58,951 (discussing public comments disputing FDA's claimed authority to impose restrictions).

595 See FDAAA § 901(b), 21 U.S.C.A. §§ 355-1(a)(1)-(2) (West Supp. 2009) (letting FDA require risk evaluation and mitigation strategies for newly- and already-approved drugs). FDA may require such strategies to include conditions for market-

for imposing restrictions is called a Risk Evaluation and Mitigation Strategy (REMS).⁵⁹⁶ FDA can require a REMS either when a drug initially is approved⁵⁹⁷ or subsequently if FDA determines, based on emerging evidence, that a REMS is needed to ensure that the drug's benefits outweigh its risks.⁵⁹⁸ If FDA requires a drug to have a REMS and the drug sponsor fails to maintain compliance with its terms, sale of the drug will be unlawful.⁵⁹⁹ All REMS must include a program for ongoing evaluation of the risk.⁶⁰⁰ Additional elements may be included.⁶⁰¹ Some of these are simple, familiar measures FDA already has been using, such as requiring Medication Guides, patient package inserts, or warning letters to advise healthcare providers of a risk.⁶⁰² However, more draconian measures can be imposed on drugs that are effective but have known risks so serious that the drug otherwise would be unavailable.⁶⁰³ That is, the risk would block approval of a new drug or cause approval of an existing drug to be withdrawn. For such drugs, FDA can condition sale on specific restrictions known as "elements to ensure safe use."⁶⁰⁴

The REMS use restrictions are not a novel concept, since they resemble measures already in effect under subpart H and voluntary risk management programs.⁶⁰⁵ FDAAA elaborates six potential elements of risk management programs⁶⁰⁶ and, importantly, gives spe-

ing and sales, such as restrictions on how, where, and by what type of professional the drug may be prescribed or requirements that patients receive monitoring or testing to manage risks. *Id.*, 21 U.S.C.A. § 355-1(f)(3); *see also id.* § 901(a), 21 U.S.C.A. §355(p) (West Supp. 2009) (making it unlawful for a person to introduce a new drug into interstate commerce if the drug requires a REMS and the person does not maintain compliance with the REMS); *id.* § 902(a), 21 U.S.C.A. § 352(y) (West Supp. 2009) (letting drugs be regarded as misbranded if manufacturers fail to comply with such restrictions).

596 *Id.* §§ 901(a)-(b), 21 U.S.C.A. §§ 355(p), 355-1.

597 *Id.* § 901(b), 21 U.S.C.A. § 355-1(a)(1).

598 *Id.*, 21 U.S.C.A. § 355-1(a)(2)(A).

599 *Id.* § 901(a), 21 U.S.C.A. § 355(p)(1)(B).

600 *Id.* § 901(b), 21 U.S.C.A. § 355-1(d).

601 *Id.*, 21 U.S.C.A. § 355-1(e).

602 *Id.*

603 *Id.*, 21 U.S.C.A. § 355-1(f).

604 *Id.*, 21 U.S.C.A. §§ 355-1(f)(1)(B), (3).

605 *See Kessler & Vladeck, supra* note 11, at 491.

606 *See* FDAAA § 901(b), 21 U.S.C.A. § 355-1(f)(3) (describing six elements of safe use: (1) FDA can limit *who* can prescribe a drug, for example, limiting it to health care providers with special training or experience; (2) FDA can require special certification of entities that dispense the drug; (3) FDA can restrict a drug for use in particular health-care settings, for example, requiring drugs that can have sudden, life-threatening reactions to be administered in hospitals where immediate emergency care would be available; (4) FDA can require that patients have complied with safe-

cific authorization to require the use of modern biomarker screening strategies and pharmacogenetic tests. A REMS can require that patients meet safe-use conditions, which can include laboratory testing before the drug is prescribed.⁶⁰⁷ For example, pregnancy testing could be required to make sure drugs that cause birth defects are not given to pregnant women, or a pharmacogenetic test could be required to assess whether a patient is a suitable candidate for the drug. This provision, in effect, lets a REMS be used to achieve functional cross labeling of a drug and a test, without actually altering the labeling of either product. This may help address the evidentiary problems that have blocked cross labeling of drugs with CLIA-regulated LDTs. Evidentiary standards for a REMS are presently less formalized, and thus may be more flexible, than the standards for including information in labeling. This requirement could be buttressed by another allowed element—inclusion of patients taking a drug in a registry,⁶⁰⁸ so that their outcomes can be followed to confirm the test's utility and develop evidence to support a later change in the drug's labeling.

Use restrictions offer a much-needed alternative to pulling a drug from the market when it is beneficial to some patients but poses serious risks for others. When patient populations are segmented by predictable variations in response, “it [is] utterly indefensible to remove the drug from the market altogether so long as warnings can enforce the proper rules for market segmentation.”⁶⁰⁹ Unfortunately, warnings alone cannot enforce the needed segmentation. A number of beneficial drugs have had to be withdrawn from the market when doctors failed to “do the right thing” in response to the FDA’s warnings.⁶¹⁰ Use restrictions aim to save drugs that are benefiting some patient subgroups by imposing binding restrictions to protect other subgroups (adverse or non-responders).

To date, the REMS that have been approved did not involve the use of pharmacogenetic testing to address variable response. After the REMS provisions of FDAAA took effect in March, 2008, FDA

use conditions such as testing prior to administration of the drug; (5) FDA can require specific monitoring of patients to detect adverse events quickly while they still can be mitigated; (6) FDA can require patients taking the drug to be enrolled in a registry so that their outcomes can be followed).

607 *Id.*, 21 U.S.C.A. § 355-1(f)(3)(D).

608 *Id.*, 21 U.S.C.A. § 355-1(f)(3)(F).

609 Epstein, *supra* note 43, at 25.

610 Gottlieb, *supra* note 584, at 666 (commenting on, rather than arguing, the position that doctors are not doing “the right thing” when they ignore drug labeling).

immediately deemed⁶¹¹ a list of sixteen existing drugs and biologics to be subject to REMS.⁶¹² These products already had been subject to use restrictions either under the subpart H regulations⁶¹³ or by voluntary agreement of their manufacturers. This initial list of REMS generally included drugs with abuse or addiction potential or having severe risks (such as causing birth defects, sudden life-threatening reactions such as anaphylaxis or cardiac death, or irreversible organ damage).⁶¹⁴ The list also included mifepristone (RU-486),⁶¹⁵ an abortion drug approved in a politically charged environment; for RU-486, subpart H was used as an instrument to set abortion policy as well as to set safety restrictions.⁶¹⁶ Beyond this initial group of drugs for which FDA “deemed” REMS restrictions to be in effect, FDA has been requiring REMS on a case-by-case basis. FDA approved thirteen REMS between March 2008 (when REMS authority took effect) and the end of September 2008 (the first anniversary of FDAAA’s enactment) and the agency had approved eighty-nine REMS as of late November 2009.⁶¹⁷ In eleven of the thirteen cases reported for FDAAA’s first year, the REMS involved only a Medication Guide or risk communication plan.⁶¹⁸ Early in 2009, FDA announced plans to use REMS to restrict distribution of twenty-four popular narcotics that account for a disproportionate share of drug-related deaths in the United States.⁶¹⁹

611 See *id.* § 909(b)(1) (allowing FDA to require REMS for products under pre-existing use restrictions).

612 Identification of Drugs and Biological Products Deemed to Have Risk Evaluation and Mitigation Strategies for Purposes of the Food and Drug Administration Amendments Act of 2007, 73 Fed. Reg. 16,313, 16,314 (Mar. 27, 2008).

613 21 C.F.R. §§ 314.520, 601.42 (2009).

614 Robert Gliklich & Michelle Bertagna Leavy, *Changes in FDA’s Approach to Risk*, APPLIED CLINICAL TRIALS, Oct. 2008, at 44, 46 tbl.2.

615 Identification of Drugs and Biological Products Deemed to Have Risk Evaluation and Mitigation Strategies for Purposes of the Food and Drug Administration Amendments Act of 2007, 73 Fed. Reg. at 16,314.

616 See discussion *infra* Part VI.

617 See FDA, FDAAA IMPLEMENTATION, *supra* note 544, at 6 (reporting thirteen REMS approved as of September 25, 2008); FDA, REMS, *supra* note 568 (reporting eighty-nine REMS approved as of November 22, 2009).

618 FDA, FDAAA IMPLEMENTATION, *supra* note 544, at 6.

619 U.S. Food & Drug Admin., Opioid Drugs and Risk Evaluation Mitigation Strategies, <http://www.fda.gov/Drugs/DrugSafety/InformationbyDrugClass/ucm163647.htm> (last visited Nov. 2, 2009); see also Gardiner Harris, *F.D.A. to Place New Limits on Prescriptions of Narcotics*, N.Y. TIMES, Feb. 9, 2009, at A13 (describing a proposed program to address risks associated with opioid medications); John Gever, *FDA to Step Up Regulation of Extended-Release Opioids*, MEDPAGE TODAY, Feb. 9, 2009, <http://www.medpagetoday.com/ProductAlert/Prescriptions/12810> (describing proposed measures to address problems with opioid medications).

Overprescribing of these drugs and failure of physicians to heed warnings already in labeling were cited as reasons for invoking the REMS mechanism.⁶²⁰ Of the eighty-nine REMS approved by late November 2009, only seven contained any use restrictions.⁶²¹

Already, there are complaints that REMS will “make drug lifecycle management more complex, more costly, and resource intensive.”⁶²² Safety advocates fear, and industry representatives hope, that REMS may amount to little more than a requirement for Medication Guides and risk communication plans, with the more cumbersome use restrictions rarely invoked. To those who say REMS are too cumbersome to get off the ground, a word of caution is in order: It is now just under two years since the effective date of this statute. Assessing its impact now is like attempting, in 1964, to assess the ultimate impact of FDA’s clinical trial requirements. Few in 1964 would have imagined the massive, resource-intensive clinical trial enterprise that grew out of the 1962 amendments.⁶²³ The fact that clinical trials were cumbersome did not ultimately deter FDA from requiring them. The same is likely true of REMS, and the right question is how REMS may evolve over the next five to forty years.

VI. UNRESOLVED DOCTRINAL ISSUES

In FDAAA, Congress declined to speak about two important matters, leaving them to be considered on a case-by-case basis by courts. Congress resisted entreaties from the pharmaceutical industry to insert an express preemption clause into the drug provisions of FDCA,⁶²⁴ and Congress left the line between medical product and medical practice regulation blurrier than it has been at any time since FDCA was enacted in 1938. Congress’s refusal to make general pronouncements about these matters was wholly consistent with the spirit of FDAAA, which eschews “one-size-fits-all” approaches to drug regulation and embraces nuanced, fact-specific treatment both of evidence development and of regulatory action in response to the evidence.

A. *Preemption of Suits Against Manufacturers After FDAAA and Wyeth v. Levine*

FDAAA explodes once-simple categories of regulatory action into a multitude of subtly different forms. For example, “drug approval”

620 Harris, *supra* note 619.

621 FDA, REMS, *supra* note 568.

622 Gliklich & Leavy, *supra* note 614, at 48.

623 See discussion *supra* Introduction.

624 Kessler & Vladeck, *supra* note 11, at 468 & n.27.

may mean approval subject only to FDAAA's general requirement to reevaluate risk after eighteen months or 10,000 prescriptions; approval subject to postmarket study requirements; approval subject to a simple REMS that lacks use restrictions; approval subject to REMS use restrictions; and approval based on surrogate endpoints—a category that has existed since 1992 but which is set to become more salient as FDA incorporates modern biomarker technologies into its drug approval pathways. Moreover, FDA will be acting in an environment of continuously accruing postmarket evidence. Suppose FDA determines, at $T(n)$, that evidence, for the first time, supports posting an early warning of risk on its Internet-based communication system. Will this determination preempt state lawsuits alleging that, at the earlier time $T(n-1)$, a manufacturer failed to warn of this risk?⁶²⁵ Whether any of FDA's new forms of action have preemptive effect is a fact-specific inquiry best left to courts, which is where Congress left the question.

The regulatory actions at issue in *Wyeth v. Levine* were a drug approval decision FDA made in 1955, decisions on changes to the drug's labeling in 1973, 1976, and 1981, and intermittent correspondence in which FDA did not pressure the manufacturer for a labeling change in subsequent years.⁶²⁶ The Court rejected Wyeth's argument that these types of action by FDA establish "both a floor and a ceiling for drug regulation"⁶²⁷ such that a state-law verdict that the labeling was inadequate would thwart Congress's purposes and objectives.⁶²⁸ The Court saw FDA's approval and labeling decisions, such as those at issue in *Wyeth v. Levine*, as a scheme of minimal regulation that could coexist with state tort law.⁶²⁹ These decisions did not amount to optimal regulation in which a regulator makes delicate risk-risk trade-offs that might be disrupted if states established higher de facto requirements through their tort damage awards.⁶³⁰

625 See Nagareda, *supra* note 3, at 30–31 (discussing the impact, in litigation over suicide risks with selective serotonin reuptake inhibitors (SSRIs), of determinations FDA made at various points in time that there was not yet any credible evidence to support a labeling change); *id.* at 28 (noting that "matters of temporal perspective loom large in the preemption analysis").

626 See *Wyeth v. Levine*, 129 S. Ct. 1187, 1192 (2009).

627 *Id.* at 1199.

628 See *id.* at 1193–94, 1199.

629 See *id.* at 1200–03.

630. See Nagareda, *supra* note 3, at 39 ("When rules of any sort concerning a given risk have the potential to increase other kinds of risks, regulators necessarily must start thinking less in terms of minimal rules and more in terms of optimal ones. But that move itself undermines the premise of minimal regulation that underlies the supposed coexistence of tort and administration.")

Wyeth v. Levine does not foreclose the possibility that some forms of post-FDAAA regulatory action may support findings of implied pre-emption. “In such cases, the Court has performed its own conflict determination, relying on the substance of state and federal law and not on agency proclamations of pre-emption.”⁶³¹ The Court has “given ‘some weight’ to an agency’s views about the impact of tort law on federal objectives when ‘the subject matter is technical [and] the relevant history and background are complex and extensive.’”⁶³² Some of the decisions FDA will be making after FDAAA may have the character of optimal regulation. For example, FDA’s decisions not to impose REMS use restrictions may, in some cases, reflect a studied policy of letting a variety of competing risk management strategies be tried in clinical practice, to see which one proves best. This decision very much would have the character of the Department of Transportation’s decision that was at issue in *Geier v. American Honda Motor Co.*⁶³³ Discussing this case in *Wyeth v. Levine*, the Court noted that “state tort claims premised on Honda’s failure to install airbags” were preempted because they “presented an obstacle to achieving ‘the variety and mix of devices that the federal regulation sought.’”⁶³⁴ The Court held that FDA’s decisions at issue in *Wyeth v. Levine* were not of this character, but applied an analytical framework that might well find preemptive effect in some of FDA’s future REMS decisions. These cases will require a fact-based inquiry geared to the specific drug and the specific circumstances in which particular use restrictions were or were not imposed. *Wyeth v. Levine* did not foreclose these inquiries; indeed, it maps out promising ways to frame the issues in future cases.

B. Federal Intrusion on Medical Practice

FDA long has described its role as determining which medical products are available, but letting physicians decide how the products should be used. REMS use requirements conflate availability and use. Under FDAAA, the agency will be deciding which products are available for which use.⁶³⁵ This blurs product and practice regulation in a

631 *Wyeth*, 129 S. Ct. at 1200–01.

632 *Id.* at 1201 (alteration in original) (quoting *Geier v. Am. Honda Motor Co.* 529 U.S. 861, 883 (2000)).

633 529 U.S. at 861.

634 *Wyeth*, 129 S. Ct. at 1203 (quoting *Geier*, 529 U.S. at 881).

635 Marc Wilenzick et al., *Interview with Gerald Masoudi, Outgoing Chief Counsel of the U.S. Food and Drug Administration*, MEMBER BRIEFING (Am. Health Law. Ass’n, Washington, D.C.), Feb. 2009, at 1, 2, available at http://www.healthlawyers.org/Members/PracticeGroups/LS/memberbriefings/Documents/Masoudi%20interview_Final.pdf (“[T]he REMS could have the effect of limiting the ability of physicians who might

way that will raise legal issues at both extremes: when patients are injured through physicians' undercompliance with REMS, and when physicians "overcomply" in ways that deny their patients needed treatments.

1. Denial of Access to FDA-Approved Treatments

REMS already are interfering with patient care. Perhaps the best example involves extra-REMS use of the drug mifepristone (RU-486). FDA approved the drug for use as an abortion agent in September, 2000 after a long and storied controversy.⁶³⁶ This drug blocks action of the female hormone progesterin⁶³⁷ and, therefore, has a wide array of potential uses to treat conditions influenced by the action of this hormone, including certain forms of cancer⁶³⁸ as well as benign tumors and other conditions of the reproductive tract,⁶³⁹ and certain brain tumors.⁶⁴⁰ Mifepristone's labeling states that its only indicated use is for medical termination of early pregnancy (through the forty-ninth day). FDA approved the drug subject to subpart H use restrictions and, in 2008, deemed these to be converted to a REMS. These use restrictions are aimed at addressing specific safety risks that can arise during a drug-induced abortion.⁶⁴¹ Distribution is tightly

use the drug for off-label uses to get their hands on the drug for their patients." (quoting Gerald Masoudi) (internal quotation marks omitted)).

636 JUDITH A. JOHNSON, ABORTION: TERMINATION OF EARLY PREGNANCY WITH RU-486 (MIFEPRISTONE) (2001), available at <http://www.law.umaryland.edu/marshall/crsreports/crsdocuments/RL30866.pdf>; see also Byron C. Calhoun & Donna J. Harrison, *Challenges to the FDA Approval of Mifepristone*, 38 ANNALS PHARMACOTHERAPY 163 (2004) (describing shortcomings of the approval process for mifepristone).

637 DIV. OF HEALTH PROMOTION & DISEASE PREVENTION, INST. OF MED., CLINICAL APPLICATIONS OF MIFEPRISTONE (RU 486) AND OTHER ANTIPROGESTINS 14 (Molla S. Donaldson et al. eds., 1993) [hereinafter IOM, CLINICAL APPLICATIONS], available at http://www.nap.edu/openbook.php?record_id=2203.

638 See Aimin Li et al., *Effect of Mifepristone on Proliferation and Apoptosis of Ishikawa Endometrial Adenocarcinoma Cells*, 84 FERTILITY & STERILITY 202, 207-09 (2005); Nat'l Insts. of Health, ClinicalTrials.gov, Mifepristone for Patients with Endometrial Cancer and LGESS, <http://clinicaltrials.gov/ct2/show/NCT00505739> (reporting the registration by M.D. Anderson Cancer Center of an eight-year clinical trial, starting in September 2001, of mifepristone in patients with endometrial cancer); see also Laura Sussman, *Clinical Trial Examines New Treatment for Endometrial Cancer*, TEX. MED. CTR. NEWS, Feb. 1, 2002, http://www.tmc.edu/tmcnews/02_01_02/page_20.html (calling the NIH study the first clinical trial to study mifepristone as a treatment for specific types of endometrial cancer).

639 IOM, CLINICAL APPLICATIONS, *supra* note 637, at 189.

640 JOHNSON, *supra* note 636, at 4.

641 See Letter from Ctr. for Drug Evaluation & Research to Sarah P. Arnold, Vice President, Corporate Affairs, Population Council, (September 28, 2000), available at

restricted.⁶⁴² Labeling is silent about use of the drug in patients who are not pregnant and thus not in pursuit of an abortion. The impact of the REMS is to make such uses unlawful in the United States.

There is mounting clinical trial evidence that mifepristone, in small doses of five mg/day for ninety days (as opposed to the 600 mg one-time dose ordinarily used in conjunction with another drug, misoprostol, to induce an abortion) is highly effective in treating a form of benign tumor, leiomyomata (uterine fibroids).⁶⁴³ This condition is variously estimated to affect 30–80% of women of childbearing age,⁶⁴⁴ with an estimated 20–40% of women aged thirty-five or older having tumors of significant size.⁶⁴⁵ In severe cases, the condition creates major health problems including pressure on other internal organs and severe anemia that makes patients dependent on repeated blood transfusions.⁶⁴⁶ An estimated 15% of off-label uses of drugs in the United States have no clinical trial evidence whatsoever to back them.⁶⁴⁷ The quantum of evidence available to support this off-label use of mifepristone is substantial by comparison. However, off-label use of mifepristone to treat leiomyomata is unlawful in the United States. Women of sufficient means may travel to European nations for treatment, but unless they can remain in Europe for the full ninety-

http://www.accessdata.fda.gov/drugsatfda_docs/appltr/2000/20687appltr.pdf.

For example, the drug can be prescribed only by or under the supervision of physicians able to diagnose ectopic pregnancies and able to perform a surgical abortion if the medical abortion fails. *Id.*

642 *See id.* (requiring distribution in accordance with a plan submitted March 30, 2000, including specific requirements for storage, dosing, dosage tracking, and return of unused product); *see also* Danco Labs., Mifeprex (mifepristone) Tablets 200mg Labeling, *available at* http://www.earlyoptionpill.com/userfiles/file/Mifeprex%20Labeling%204-22-09_Final_doc.pdf (listing updated warnings for Mifeprex); Danco Labs. Mifeprex (mifepristone) Tablets 200mg Prescriber's Agreement, *available at* <http://www.fda.gov/downloads/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm111364.pdf> (stating necessary qualifications for prescribers of Mifeprex).

643 *See* Steven H. Eisinger et al., *Low-Dose Mifepristone for Uterine Leiomyomata*, 101 *OBSTETRICS AND GYNECOLOGY* 243, 245–50 (2003); Jody Steinauer et al., *Systematic Review of Mifepristone for the Treatment of Uterine Leiomyomata*, 103 *OBSTETRICS & GYNECOLOGY* 1331, 1332–35 (2004).

644 Rebecca L. Van Court, *Uterine Fibroids and Women's Right to Choose*, 26 *J. LEGAL MED.* 507, 509–10 (2005).

645 *Id.*

646 *Id.* at 508–09.

647 Radley et al., *supra* note 222, at 1023.

day course of treatment, they run the risk of having their medicine confiscated upon re-entry to the United States.⁶⁴⁸

Congress was aware that REMS use restrictions could create problems of this sort. Accordingly, FDAAA specifies that use restrictions must be commensurate with the specific serious risk that is being addressed.⁶⁴⁹ REMS must not unduly burden patient access to the drug, particularly in the case of patients with serious or life-threatening conditions.⁶⁵⁰ Congress fashioned a REMS bypass option⁶⁵¹ that would let drugs be supplied for extra-REMS use under Section 561 of FDCA.⁶⁵² Section 561 lets investigational (unapproved) drugs be used, under certain circumstances, by seriously ill patients who have run out of other treatment options. Congress instructed FDA to develop regulations allowing extra-REMS use under Section 561.⁶⁵³ To date, FDA has failed to issue these regulations, leaving no clear way for physicians to bypass REMS use restrictions when they are interfering with care of seriously ill patients.

As applied to nonpregnant patients who are not seeking an abortion, the mifepristone REMS violates FDAAA's statutory requirement of narrow tailoring. These patients have zero risk of the safety problems (such as incomplete abortion or a ruptured ectopic pregnancy) that the REMS use restrictions were designed to address. For these patients, the REMS use restrictions serve no valid safety purpose and can only be viewed as an instrument of abortion policy. To be lawful, laws restricting access to abortion must have an exception for life or health of the mother. The mifepristone REMS presents a novel but closely related question: must there also be an exception if laws restricting abortion endanger the life or health of a person other than the mother? The mifepristone REMS is doing that. Unable to get mifepristone, women whose fibroids are seriously symptomatic face an array of bad options including unnecessary hysterectomies which, in women of childbearing age, interfere with their fundamental liberty interest in reproduction.

648 *Benten v. Kessler*, 505 U.S. 1084 (1992), *denying cert. to Benten v. Kessler*, 799 F. Supp. 281 (E.D.N.Y. 1992) (refusing to hear the case of a woman from whom mifepristone was confiscated as she attempted to bring the product into the U.S. for her own use to end a pregnancy); *see also* JOHNSON, *supra* note 636, at 2–3 (discussing this case).

649 FDAAA § 901(b), 21 U.S.C.A. § 355-1(f)(2)(A) (West Supp. 2009).

650 *Id.*, 21 U.S.C.A. § 355-1(f)(2)(C)(i).

651 *Id.*, 21 U.S.C.A. § 355-1(f)(6).

652 *See* 21 U.S.C. § 360bbb (2006).

653 FDAAA § 901(b), 21 U.S.C.A. § 355-1(f)(6).

Because of the tortured approval history of this drug, U.S. pharmaceutical companies were concerned about boycotts if they produced it and its French manufacturer ultimately licensed it to a nonprofit group which sponsored its FDA approval application.⁶⁵⁴ As a result, there is no commercial sponsor that now has an incentive to press FDA to allow wider use of the drug in nonpregnant patients. Patients are left to advocate for themselves. In contrast to the experimental therapy that was at issue in *Abigail Alliance*, mifepristone is an FDA-approved drug. The proper question is not whether clinical trial evidence supports its current approval for use in treating fibroid tumors, but whether this use has a level of evidentiary support comparable to that of other off-label uses that FDA allows. If FDA fails to provide a meaningful REMS bypass option, the mifepristone REMS is ripe for litigation. Exhausting administrative remedies requires only that a patient file a Citizen's Petition⁶⁵⁵ with FDA and wait 180 days.

2. Enforcing Physician Compliance with Use Restrictions

Even before Congress had passed FDAAA, the question was raised whether REMS use restrictions amount to FDA regulation of medical practice.⁶⁵⁶ The answer has many implications, not all of which will be distasteful to physicians. For example, there is a possibility that doctors who comply with a REMS may be able to assert a regulatory compliance defense to medical malpractice suits brought by patients who suffer drug-related injuries.

FDAAA does not, on its face, grant FDA any new powers to regulate doctors or to enforce their compliance with use restrictions in a REMS.⁶⁵⁷ Congress, instead, authorized FDA to require drug manufacturers to provide a REMS, which ostensibly regulates only the manufacturer while setting conditions for clinical use of the drug. If those restrictions are violated, the manufacturer, rather than the physician, would be in violation of federal law⁶⁵⁸ and would be subject to signifi-

654 JOHNSON, *supra* note 636, at 4–5.

655 21 C.F.R. § 10.30 (2009).

656 Wilenzick et al., *supra* note 635, at 2; *see also* Michael B. Enzi & Edward M. Kennedy, *Risk Management and Intrusions on Medical Practice: Striking a Balance*, 26 HEALTH AFFAIRS 678 (2007) (arguing, prior to passage of FDAAA, that its proposed REMS provisions strike a realistic balance between managing risks and intruding on medical practice).

657 *See* Wilenzick et al., *supra* note 635, at 2 (“REMS provisions generally bind the sponsor of the drug, not physicians.” (quoting Gerald Masoudi)).

658 FDAAA § 901(a), 21 U.S.C.A. § 355(p) (West Supp. 2009) (making it unlawful for a person to “introduce or deliver for introduction into interstate commerce a new

cant civil penalties.⁶⁵⁹ The drug can be deemed misbranded, which allows FDA to pursue various sanctions against the manufacturer including seizure of the drug.⁶⁶⁰ FDAAA lays the burden of enforcing physician compliance on drug manufacturers. In 1992, FDA took a similar view of its subpart H use restrictions: "The burden is on the [new drug or biologics license] applicant to ensure that the conditions of use under which the applicant's product was approved are being followed."⁶⁶¹ FDAAA purports not to regulate physicians, yet already there are complaints that physicians bear the burden of implementing and documenting compliance with REMS use restrictions.⁶⁶²

This REMS enforcement mechanism raises a number of questions. By what means, precisely, can drug manufacturers control the behavior of errant physicians? The drug manufacturer may be too conflicted to be a credible enforcer of patient safety: off-label (or, more correctly, extra-REMS) use of drugs bolsters sales, and the manufacturer would likely be shielded from tort liability for extra-REMS prescribing under the learned intermediary doctrine. Does it comport with norms of justice for FDA to penalize manufacturers for REMS violations by physicians? If this seems unjust, will FDA decline to enforce REMS, so that FDA's policies post-REMS are ultimately little different from today's lax policy on off-label use?

It seems implausible that FDA would deem a drug misbranded and pull it off the market if a few physicians were violating elements of safe use included in its REMS. State tort actions offer a more selective approach to deterring extra-REMS prescribing that injures patients. The 1938 Act left it for states to develop their own approaches to promote physician compliance with safety information reflected in drug labeling.⁶⁶³ States generally eschewed a direct regulatory approach

drug" that requires a REMS, if the person fails to maintain compliance with the REMS).

659 *Id.* § 902, 21 U.S.C.A. § 333(f) (West Supp. 2009) (calling for penalties of \$250,000 per REMS violation, up to \$1 million in a single proceeding, with doubling of penalties for each 30 days of continued violation after the Secretary provides notice, not to exceed \$1 million in a 30-day period and \$10 million for all violations adjudicated in a single proceeding).

660 *Id.* § 902(a), 21 U.S.C.A. § 352(y) (West Supp. 2009) (permitting drugs to be regarded as misbranded if a REMS is required and the responsible person (such as the manufacturer) fails to comply with its terms).

661 New Drug, Antibiotic, and Biological Drug Product Regulations; Accelerated Approval, 57 Fed. Reg. 58,942, 58,952 (Dec. 11, 1992).

662 *See Gever, supra* note 619 (discussing paperwork responsibilities previous REMS have imposed on physicians and requirements to implement pregnancy testing under the REMS for teratogenic medical products such as Accutane).

663 *See Evans & Flockhart, supra* note 81, at 51.

and medical malpractice suits became the de facto compliance mechanism at the state level.⁶⁶⁴ The result has been an unclear, inconsistent framework of physician compliance with FDA's safety information.⁶⁶⁵ States vary in whether the tort standard of care requires safety warnings and contraindications in labeling to be heeded.⁶⁶⁶ It remains for the medical profession and individual states to assess how standards of care (and possible defenses to malpractice actions) will be affected by REMS and by the Internet-based communication system that will be feeding continuously updated warnings to both physicians and patients. There is much work still to be done at the state level, if REMS are to achieve their promise of improving patient safety.

CONCLUSION

Congress has set the pillars of a new evidentiary paradigm, but it is still unclear what the finished edifice will look like. It appears likely to have the following contours: Products will move into clinical use after premarket studies establish a basic level of safety and efficacy (or a plausible basis for projecting the efficacy of products intended for long-term predictive and preventive uses). These initial determinations may rely on surrogate endpoints. Evidence development will continue after approval, with new forms of regulatory action available in response to emerging evidence of risks and benefits.

FDAAA implies a culture shift both at FDA and among the public. Today, when FDA announces a new risk with a previously approved drug, both the agency and the public tend to regard this as a regulatory failure: why did FDA not detect the problem before it approved the drug? This question was rooted in a naïve, twentieth-century belief that clinical trials are, or should be, fully informative. Zero late-emerging risk was never a realistic goal. The goal should be to detect risks as quickly as possible, to minimize them using the best available risk mitigation technologies, and to communicate emerging evidence swiftly and effectively. By this view, the regulatory failure, in the case of Vioxx, was not that new cardiovascular risks emerged after the product was approved. The regulatory failures were: first, that it took sixty-five months to get the product off the market when, with appropriate information infrastructure, the risk might have been detected

664 *Id.*

665 *Id.* at 52.

666 Linda A. Sharp, Annotation, *Malpractice: Physicians' Liability for Injury or Death Resulting from Side Effects of Drugs Intentionally Administered to or Prescribed for Patient*, 47 A.L.R. 5th 433, § 2[a] (1997, updated through 2004).

in three months; second, that even now—almost five years after the scandal and ten years after the product was approved—the factors that predispose a small subset of patients to cardiovascular risks are still unknown; and, third, that FDA’s old evidentiary paradigm failed to foster cooperative development of technologies to predict and manage this risk so that patients who respond well to the drug can have it. A new paradigm was needed and FDAAA has supplied one. As with any major change in law, there are many unresolved legal questions with which scholars now need to engage.