Washington and Lee Law Review

Volume 70 | Issue 4 Article 8

Fall 9-1-2013

Pharmaceutical Efficacy: The Illusory Legal Standard

Jonathan J. Darrow

Pharmaceutical Efficacy: The Illusory Legal Standard

Jonathan J. Darrow*

Abstract

The very long and expensive process of new drug research and development might suggest to observers that the efficacy standard for drugs is elevated and substantial, but this is not the case. Under the U.S. Federal Food, Drug, and Cosmetic Act, new drug approval merely requires that there be "substantial evidence that the drug will have the effect it purports or is represented to have." While the evidence of effectiveness must therefore be substantial, the efficacy attested to by that evidence need not surpass any particular threshold (other than zero), thus allowing drugs with de minimis *efficacy to be approved and sold at market rates. No other concept, principal, or standard applied during the approval process or after changes this result. The "gold standard," which includes the elements of blinding, randomization, and placebo control, is described in but not required by the drug statute, and in any event addresses various problems related to bias rather than magnitude of efficacy. Similarly, the concept of "statistical significance," which constitutes an essential element of modern research protocols, addresses the problem of certainty, not degree, of efficacy. The statutory requirement of "clinical significance," far from ensuring "substantial efficacy," demands no more than that there be statistical significance in a human study (as opposed to, for example, an animal study). Rather than specifying a fixed level of*

* The author is a Research Fellow at Harvard Medical School, a Postdoctoral Research Fellow in the Division of Pharmacoepidemiology and Pharmacoeconomics (Program On Regulation, Therapeutics And Law (PORTAL)) at Brigham and Women's Hospital, and a member of the law faculty at Bentley University. S.J.D., Harvard Law School; J.D., Duke University; LL.M., Harvard Law School (waived); M.B.A., Boston College. The author wishes to thank William Fisher, Benjamin Roin, Aaron Kesselheim, Donald Light, Tyler Black, and the editors of the Washington and Lee Law Review for their helpful contributions. Any errors remain the author's own.

efficacy or even a flexible standard of efficacy that a drug must possess, the U.S. drug approval framework thus fully delegates to drug companies and to the free market the determination of what level of efficacy is acceptable. The critical implication is that the public (and physicians and insurers) should not rely on the fact of FDA approval as an indication that medicines, including new and very highly priced ones, possess efficacy that is meaningfully greater than no efficacy at all.

Table of Contents

I. Introduction

David Kessler, a former Commissioner of the U.S. Food and Drug Administration (FDA), once boasted to Congress that the agency's "rigorous demand for safety and efficacy . . . makes FDA approval the international gold standard."[1] If the U.S. drug approval system does indeed set the global efficacy benchmark, it is a cause for concern. Although the FDA deserves praise for its significant efforts in promoting safety and promptly approving new medicines, the agency operates under a legal efficacy standard that is as likely to engender disbelief as it is to incite indignation:

1. *Testimony on FDA's Role in Protecting and Promoting Public Health: Hearing Before the S. Comm. on Labor and Human Res.*, 104th Cong. (Feb. 21, 1996) (statement of David A. Kessler, Comm'r, Food and Drug Administration), *available at* http://www.hhs.gov/asl/testify/t960221a.html.

Under U.S. law, there is no particular level of efficacy required for a new drug to be approved.[2] Instead, drugs with near-zero efficacy can be approved, prescribed, and sold to patients who have real and sometimes very serious diseases or conditions.[3]

Patients and physicians would be right to consider this assertion skeptically. How could an ineffective drug be approved under a legal system where it is axiomatic that all new drugs must be proven "safe and effective"[4] prior to government approval? This familiar and comforting rhetoric, however, conceals the illusory nature of what the underlying legal standard actually requires. The Federal Food, Drug, and Cosmetic Act[5] requires only that there be "substantial evidence that the drug will have the effect it purports or is represented to have."[6] It is the thesis of this Article that this standard, along with the related concepts of gold standard testing, statistical significance, and clinical significance, do not prevent FDA approval of substantially ineffective remedies.

The implications are alarming. Not only does an illusory efficacy standard create the possibility that many expensive drugs marketed today are essentially worthless, but the near-universal (but false) perception that FDA approval guarantees substantial effectiveness can lead physicians and patients to cede responsibility for critically evaluating drug value, thus adversely affecting treatment choices. While this harm must be balanced against the tremendous benefits flowing from FDA regulation, it nevertheless raises the specter that, in a significant sense, the stamp of "FDA approval" may actually work to harm public health.

---

2. *See* 21 U.S.C § 355(d) (2012) (stating that FDA Drug Approval guidelines require "substantial evidence" that "the drug will have the effect it purports or is represented to have").

3. *See* Joan E. Shreffler, *Bad Medicine: Good-Faith FDA Approval as a Recommended Bar to Punitive Damages in Pharmaceutical Products Liability Cases*, 84 N.C. L. REV. 737, 758 (2006) (discussing how a grant of FDA approval only requires the benefits of a drug to outweigh foreseeable risks).

4. *See, e.g.*, Pliva v. Mensing, 131 S. Ct. 2567, 2574 (2011) (noting that among the issues not in dispute was the fact that "[u]nder . . . the Federal Food, Drug and Cosmetic Act . . . a manufacturer seeking federal approval to market a new drug must prove that it is safe and effective").

5. 21 U.S.C. § 301.

6. *Id.* § 355(d).

II. The Illusory Legal Standard for Drug Efficacy

The legal standard that is most closely related to the level of efficacy a new drug must possess in order to receive FDA approval is set forth in a lengthy section of Title 21 of the United States Code, the relevant portion of which states:

> If the Secretary finds . . . that . . . there is a lack of *substantial evidence* that the drug will have *the effect it purports or is represented to have* . . . or [the drug's] labeling is false or misleading in any particular [then] he shall issue an order refusing to approve the [new drug] application.[7]

This language may sound as if it does not impose a requirement for any particular level of efficacy, and in fact it does not do so.[8] The phrases "substantial evidence" and "the effect it purports or is represented to have" do impose efficacy-related requirements on drug sponsors.[9] Neither, however, requires drugs to have anything more than next-to-zero levels of efficacy.[10] What these two phrases require—and do not require—is discussed next.

A. "The Effect It Purports or Is Represented to Have"

Section 355(d), quoted above, does specify a level of efficacy that new drugs must have if they are to be approved. That level of efficacy is defined by statute to be whatever level the drug "purports or is represented to have."[11] In other words, rather than

---

7. *Id.* § 355(d) (emphasis added).

8. *See* DANIEL CARPENTER, REPUTATION AND POWER: ORGANIZATION IMAGE AND PHARMACEUTICAL REGULATION AT THE FDA 146, 156, 192–93 (2010) (describing efficacy as "a slippery concept," "indefinable," and "robustly ambiguous," and discussing the absence of agreement or precision with respect to what was meant by "efficacy").

9. *See* 21 U.S.C. § 355(d) (2012) (describing grounds for refusing application, approval of application, and substantial evidence).

10. *Id.*

11. *Id.* Strictly speaking, the term "efficacy" refers to the benefits a drug produces under ideal conditions, that is, under the controlled conditions of a clinical trial. In contrast, the term "effectiveness" refers to the benefits a drug produces under the usual circumstances of health care practice. *See* STEDMAN'S MEDICAL DICTIONARY (27th ed. 2000) (defining "efficacy" and "effectiveness"); STAN N. FINKELSTEIN & PETER TEMIN, REASONABLE RX: SOLVING THE DRUG PRICE CRISIS 9, 21 (2008) (noting similar definitions used by the European Union and

specifying a fixed level of efficacy or a even a flexible standard of efficacy that a drug must possess, the statute fully delegates to drug companies the ability to specify any level of efficacy they desire.[12] If clinical test results demonstrate that a drug's efficacy that is only slightly above zero, that drug can nevertheless receive FDA approval so long as companies do not purport in the labeling that the drug is more effective than the evidence supports.[13] The accompanying regulations merely repeat the statutory words verbatim, adding no additional requirement or clarification regarding efficacy level.[14] As a result, the standard is almost entirely illusory because it leaves to the drug sponsor the ability to specify any non-zero level of efficacy.[15]

The real world result is predictable: Statements of efficacy contained in drug labels are often nearly meaningless, but are presented in such a way as to deemphasize the level of efficacy patients can expect while emphasizing the symptoms from which patients are seeking relief. For example, over-the-counter labeling for the analgesic Motrin (ibuprofen) states, innocently enough, that it "temporarily relieves minor aches and pains due to: headache . . . backache . . . muscular aches [and] toothaches . . . ."[16] A patient who wants a headache to go away might mistakenly assume that

---

by the International Network of Agencies for Health Technology Assessment, an international organization comprising agencies from twenty-nine countries); Hans-Georg Eichler et al., *Relative Efficacy of Drugs: An Emerging Issue Between Regulatory Agencies and Third-Party Payers*, 9 NATURE REVIEWS DRUG DISCOVERY 277, 279 box 1 (Apr. 2010) (same). However, because the terms are often used interchangeably or inconsistently with these definitions, the term "efficacy" will generally be used throughout this work for simplicity.

12. *See* 21 U.S.C. § 355(d) (2012) (discussing grounds for refusing applications and defining "substantial evidence").

13. *Id.*

14. *See* 21 C.F.R. § 314.125(b)(5) (2009) (echoing the identical "substantial evidence" requirement found in federal statute).

15. *Cf.* Kenyon v. Jennings, 560 F. Supp. 878, 881 (D. Kan. 1983) ("An illusory promise is 'a purported promise that actually promises nothing because it leaves to the speaker the choice of performance or nonperformance.'" (quoting BLACK'S LAW DICTIONARY 674 (5th ed. 1979))).

16. Drug Label for Motrin (ibuprofen), http://www.accessdata.fda.gov/drugsatfda_docs/label/2007/017463s105lbl.pdf (last visited Sept. 29, 2012) (on file with the Washington and Lee Law Review).

ibuprofen can accomplish this result, at least temporarily, in light of the "temporar[y] relie[f]" promised in the label.[17]

Those who swallow two caplets of ibuprofen, however, are unlikely to experience either immediate or complete disappearance of pain. One double-blind, randomized and controlled trial, for example, found that of seventy-five patients who took all three of ibuprofen, acetaminophen or placebo at different times, eighteen preferred ibuprofen, eighteen acetaminophen, twelve placebo, and twenty-seven expressed no preference, which differences in preference were not statistically significant.[18] Another randomized, blinded study assessed the intensity of headache on a four point scale (4 = intense; 3 = moderate; 2 = slight; 1 = none).[19] Three hours after treatment, those taking ibuprofen reported scores of 2.05, while those taking placebo reported an average score of 2.48, a statistically significant but hardly impressive difference.[20] The difference becomes even smaller (though still statistically significant) when one realizes that the baseline average scores for the two groups were different: 2.74 for those taking ibuprofen versus 2.98 for those taking placebo, yielding reductions in pain of 0.69 (2.74 minus 2.05) for ibuprofen versus 0.50 (2.98 minus 2.48) for placebo.[21] A number of other studies reveal similarly unimpressive efficacy in relieving headache pain.[22] The Motrin

---

17. *Id.*

18. Mirja L. Hamalainen et al., *Ibuprofen or Acetaminophen for the Acute Treatment of Migraine in Children: A Double-Blind, Randomized, Placebo-Controlled Crossover Study*, 48(1) NEUROLOGY 103, 106 (1997) (finding other measurements of pain relief were nevertheless statistically significant).

19. Seymour Diamond, *Ibuprofen Versus Aspirin and Placebo in the Treatment of Muscle Contraction Headache*, 23(5) HEADACHE: J. HEAD & FACE PAIN 206, 207 (1983).

20. *Id.* at 208 tbl.1.

21. *Id.*

22. *See, e.g.*, J.M.A. Van Gerven et al., *Self-Medication of a Single Headache Episode with Ketoprofen, Ibuprofen or Placebo, Home-Monitored with an Electronic Patient Diary*, 42 BRIT. J. CLINICAL PHARMACOLOGY 475, 478 fig.1 (1996) (showing no advantage of ibuprofen over placebo during the first two hours of observations); *cf.* Alan Branthwaite & Peter Cooper, *Analgesic Effects of Branding in Treatment of Headaches*, 282 (6276) BRIT. MED. J. 1576, 1577 tbl.II (1981) (reporting substantial improvements in headache pain for 86% of those given unbranded aspirin versus 74% of those given unbranded placebo); Roger Chou et al., *Comparative Effectiveness and Safety of Analgesics for Osteoarthritis*, OREGON EVIDENCE-BASED PRACTICE CTR.: DRUG EFFECTIVENESS REVIEW PROJECT, REVIEW NO. 4, Sept. 2006, at 10, http://effectivehealth

(ibuprofen) drug label, however, makes neither the claim that relief is immediate nor that it is complete.[23] In fact, it promises no particular level of relief at all.[24]

Although Motrin is an over-the-counter drug, many prescription drugs also promise no particular level of relief. Among the top-selling prescription drugs in a recent year were Cymbalta (duloxetine), Plavix (clopidogrel), and Enbrel (etanercept).[25] Cymbalta, a serotonin and norepinephrine reuptake inhibitor, is indicated for depression, anxiety, fibromyalgia, and musculoskeletal pain.[26] A television commercial for Cymbalta emphasizes the symptoms of depression but, with respect to efficacy, notes only that "Cymbalta can help" and that "Cymbalta . . . treats many symptoms of depression."[27] The television commercial for Enbrel (etanercept) similarly states only that "Enbrel can help relieve pain, stiffness, and stop joint damage."[28] Small print briefly viewable during the advertisement states that "Enbrel was shown to be effective in 50% of psoriatic

---

care.ahrq.gov/repFiles/AnalgesicsFinal.pdf (noting that several recent "good-quality systematic reviews" by the Cochrane Collaboration found no clear differences among nonselective NSAIDs in efficacy for treating knee, back, or hip pain). *But see* Bernard P. Schachtel & William R. Thoden, *Onset of Action of Ibuprofen in the Treatment of Muscle-Contraction Headache*, 28 HEADACHE: J. HEAD & FACE PAIN 471 (1988) (finding that ibuprofen provided substantially greater pain relief than placebo, but admitting that participating subjects were selected only if they reported previously satisfactory experience with nonprescription analgesics, i.e., participants may not have been representative subjects because they may have been unusually responsive to ibuprofen treatment).

23. Drug Label for Motrin (ibuprofen), *supra* note 16.

24. *Id.*

25. *See* Matthew Herper, *The Best-Selling Drugs in America*, FORBES (Apr. 19, 2011, 8:48 AM), www.forbes.com/sites/matthewherper/2011/04/19/the-best-selling-drugs-in-america (last visited Oct. 21, 2013) (listing the top-selling medicines in the United States in 2010) (on file with the Washington and Lee Law Review).

26. Drug Label for Cymbalta (duloxetine hydrochloride) Delayed-Release Capsules for Oral Use, Oct. 2010, Reference ID 2860327, at 1, http://www.accessdata.fda.gov/drugsatfda_docs/label/2010/022516lbl.pdf.

27. Tracy Shier, *Cymbalta Commercial (real one)*, YOUTUBE (Nov. 2, 2009), http://www.youtube.com/ watch?v=OTZvnAF7UsA (last visited Sept. 23, 2013) (on file with the Washington and Lee Law Review).

28. *Enbrel TV Spot Featuring Phil Mickelson*, http://www.ispot.tv/ad/7VfC/enbrel-featuring-phil-mickelson (last visited Apr. 10, 2013) (on file with the Washington and Lee Law Review).

arthritis patients at 6 months," but this statement addresses only the fraction of people who experienced any relief, not the amount of relief experienced by those patients.[29] Moreover, by implication the other half experienced no relief. A Plavix (clopidogrel) commercial boasts, as it changes ominous background music associated with symptoms to more uplifting music associated with the drug treatment, that "Plavix, in combination with aspirin and other heart medicines helps provide greater protection against heart attack or stroke than aspirin and other medicines alone."[30] The FDA has rebuked the maker of Plavix for overstating similar claims of efficacy in Plavix print advertisements.[31]

In each of these commercials, there is no claim to any particular level of efficacy.[32] How much Cymbalta can "help" mitigate symptoms of depression, how much "Enbrel helps . . . stop joint damage" and how much Plavix "helps" protect against heart attack and stroke, are questions all conveniently left for the television viewer to imagine.[33] The imagination, however, is not

---

29. *Id.*

30. LowPricePlavix, *Plavix TV Commercial*, YOUTUBE (Aug. 30, 2010), https://www.youtube.com/watch?v=oHRX0b4kPNY (last visited Oct. 21, 2013) (on file with the Washington and Lee Law Review).

31. *See* Letter from Andrew S.T. Haffer, U.S. Food & Drug Admin., to Kenneth Palmer, Sanofi Pharms. (May 9, 2001), http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/EnforcementActivitiesbyFDA/WarningLettersandNoticeofViolationLetterstoPharmaceuticalCompanies/UCM166467.pdf (advising that a Plavix visual aid overstated and mislead viewers about the drug's efficacy); *see also* Letter from Janet Norden, U.S. Food & Drug Admin., to Gregory M. Torre, Sanofi Pharms. (Dec. 18, 1998), http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/EnforcementActivitiesbyFDA/WarningLettersandNoticeofViolationLetterstoPharmaceuticalCompanies/UCM166391.pdf ("[C]laims that suggest Plavix has been 'proven' to be more effective than aspirin are misleading because they are not based on substantial evidence."); Deepak L. Bhatt et al., *Clopidogrel and Aspirin Versus Aspirin Alone for the Prevention of Atherothrombotic Events*, 354(16) NEW ENG. J. MED. 1706, 1714 (2006) ("[T]he combination of clopidogrel plus aspirin was not significantly more effective than aspirin alone in reducing the rate of myocardial infarction, stroke, or death from cardiovascular causes . . . .").

32. *See supra* notes 26–30 and accompanying text (asserting claims to help particular ailments but no degrees of aid).

33. The claim that "Enbrel can help relieve pain, stiffness, and stop joint damage" is particularly interesting because it is grammatically awkward. If an "and" were inserted between "pain" and "stiffness," then the word "help" would modify only those two words. The use of the comma instead of "and," by

left entirely to chance. The music, imagery, prescription status, and fact of FDA approval, among other things, all help the viewer to imagine that the drug must be substantially effective, without ever stating this directly.

Because the efficacy standard is a fully-adjustable hurdle, drugs that are able to jump that hurdle may be substantially ineffective, much as a five-year old can easily jump over a rope if that rope is lying on the ground. This is true not only for new drugs that are approved under the "purports or is represented to have" standard, but also for those approved prior to 1962, the year the standard was first introduced by the Kevauver-Harris Drug Amendments Act.[34] After the 1962 Act was passed, the FDA Commissioner decided to retrospectively evaluate about 4,000 pre-1962 drug formulations in what became known as the Drug Efficacy Study Implementation (DESI).[35] DESI engaged thirty panels of experts who were asked to classify the drug claims as either (1) effective, (2) probably effective, (3) possibly effective, or (4) ineffective.[36] It is often recognized that the large majority (81%) of drug claims subject to DESI review were classified in the latter three categories.[37] That is, drugs approved between 1938 and 1962 could not be said to be unqualifiedly effective for (on average) four

---

contrast, may be intended to make it sound as if "Enbrel can . . . stop joint damage" (a claim of 100% efficacy) when in reality the claim is only that "Enbrel can help . . . stop joint damage" (a claim of no particular level of efficacy).

34. Pub. L. No. 87-781, 76 Stat. 780 (1962) (codified at 21 U.S.C. § 301).

35. *See* NAT'L RESEARCH COUNCIL, DRUG EFFICACY STUDY: FINAL REPORT TO THE COMMISSIONER OF FOOD AND DRUGS, FOOD AND DRUG ADMINISTRATION 1 (1969) (describing the Commissioner's decision to examine pre-1962 approved drugs still on the market).

36. *Id.* at 42–43.

37. *See, e.g.*, Ralph F. Hall, *Right Question, Wrong Answer: A Response to Professor Epstein and the "Permititis" Challenge*, 94 MINN. L. REV. HEADNOTES 50, 70 (2010) (noting that only 19.1% of reviewed claims gained an "effective" rating); Matthew J. Seamon, *Plan B for the FDA: A Need for a Third Class of Drug Regulation in the United States Involving a "Pharmacist-Only" Class of Drugs*, 12 WM. & MARY J. WOMEN & L. 521, 541 n.172 (2006) (same); *see also* NAT'L RESEARCH COUNCIL, *supra* note 35, at 7. The 81% figure also included two additional rating categories, namely, "[e]ffective, but . . ." (i.e., where "much better safer or more conveniently administered drugs were . . . available") and "[i]neffective as a fixed combination" (i.e., based on the principle that multiple drugs should not be administered when a single drug alone would be effective). Hall, *supra*, at 70.

out of five claims, under the legal standard put in place in 1962.[38] Less recognized in discussions of DESI is that even for those drugs that were placed into the most favorable category ("effective"), this categorization was based only on whether the drug could be said, based upon substantial evidence, to have "the effect [it] purports or is represented to have."[39] Under this standard, drugs could have received (and probably did receive) the top rating even if evidence showed them to perform only marginally better than nothing (i.e., placebo), so long as manufacturers did not claim that the drugs could do more than that.

B. Substantial Evidence of Efficacy Versus Evidence of Substantial Efficacy

The statutory language that most directly constitutes an efficacy standard is the requirement that a new drug have that level of efficacy that it "purports or is represented to have."[40] This is not the only language relating to efficacy, however. Earlier in the same sentence is the requirement that a new drug cannot be approved by the FDA unless there is "*substantial evidence* that the drug will have the effect it purports or is represented to have."[41] This quoted language might be shortened to a requirement that there be "substantial evidence [of efficacy]." Critically, the statute does not require "evidence of *substantial efficacy*." Although a subtle distinction in word order, the statutory language makes clear something of great moment: it is not the efficacy of a drug that must be substantial in order to receive approval, but the evidence supporting that efficacy.

Rather than require any minimum level of efficacy, the requirement of "substantial evidence" is merely an evidentiary standard, that is, it refers to the amount or sufficiency of evidence needed to support a conclusion. Like evidentiary standards in general, "substantial evidence" has no precise definition. In the spectrum of evidentiary standards, however, it is among the

38. *See* Hall, *supra* note 37, at 70–71 (recognizing the 81% statistic); Seamon, *supra* note 37, at 541 n.172 (same).

39. NAT'L RESEARCH COUNCIL, *supra* note 35, at 43.

40. 21 U.S.C. § 355(d) (2012).

41. *Id.* (emphasis added).

easiest to satisfy. As construed by the Fourth Circuit, "substantial evidence requires more than a scintilla, but less than a preponderance, of the evidence."[42] This definition places the required sufficiency of evidence below not only a preponderance of the evidence (greater than 50% likelihood), but also far below "clear and convincing evidence" (greater than, perhaps, 80% likelihood) and toward the opposite end of the spectrum from "beyond a reasonable doubt" (perhaps, near 100%).[43] It is therefore no wonder that New York's highest court has described "substantial evidence" as "a minimal standard,"[44] while a Delaware court described it as "the lowest standard of proof."[45] The Sixth Circuit has added that even "the possibility of drawing two inconsistent conclusions from the evidence does not prevent an administrative agency's finding from being supported by substantial evidence."[46]

---

42. Jackson v. Astrue, No. 10–2226, 2012 WL 580239, at *1 (4th Cir. Feb. 23, 2012) (citing Mastro v. Apfel, 270 F.3d 171, 176 (4th. Cir. 2001)).

43. There is a notable lack of consensus on exact percentages equivalents, other than the preponderance of the evidence standard, for which there is substantial but not universal agreement. *See, e.g.*, United States v. Shonubi, 895 F. Supp. 460, 471 (E.D.N.Y. 1995) ("A survey of judges . . . found general agreement that 'a preponderance of the evidence' translates into 50+ percent probability. Eight judges estimated "clear and convincing" as between 60 and 70 percent probable . . . . Estimates for 'beyond a reasonable doubt' ranged from 76 to 90 percent, with 85 percent the modal response."), *vacated on other grounds*, 103 F.3d 1085 (1997); Barbara D. Underwood, *The Thumb on the Scales of Justice: Burdens of Persuasion in Criminal Cases*, 86 YALE L.J. 1299, 1311 (1977) ("[A]lmost a third of the responding judges put 'beyond a reasonable doubt' at 100%, another third . . . at 90% or 95%, and most of the rest put it at 80% or 85%. For the preponderance standard, by contrast, over half put it at 55%, and most . . . put it between 60%, and 75%."); Byron K. Warnken, *Litigating "Forfeiture by Wrongdoing" After* Crawford v. Washington, 41-AUG MD. B.J. 22, 24 (2008) (noting the "risk of factual error" to be "about 30 percent" under the clear and convincing evidence standard and "49 percent" under the preponderance of the evidence standard).

44. FMC Corp. v. Unmack, 699 N.E.2d 893, 896 (N.Y. 1998).

45. *In re* Susan S., No. 7764, 1996 WL 75343, at *12 (Del. Ch. Feb. 8, 1996) (citing Shipman v. Div. of Soc. Servs., 454 A.2d 767 (Del. Fam. Ct. 1982), *aff'd*, 460 A.2d 528 (Del. 1983)); *see also* Conn. Light & Power Co. v. Dep't of Pub. Util. Control, 830 A.2d 1121, 1131 (Conn. 2003) ("Th[e] substantial evidence standard is highly deferential and permits less judicial scrutiny than a clearly erroneous or weight of the evidence standard of review.") (quoting MacDermid, Inc. v. Dep't of Envtl. Prot., 778 A.2d 7 (Conn. 2001)).

46. Painting Co. v. NLRB, 298 F.3d 492, 499 (6th Cir. 2002) (quoting NLRB v. Ky. May Coal Co., 89 F.3d 1235, 1241 (6th Cir. 1996)). The court also noted

The Senate Report accompanying the 1962 Kefauver-Harris drug amendments clarifies what this standard means in the pharmaceutical context: "When a drug has been adequately tested by qualified experts and has been found to have the effect claimed for it, this claim should be permitted *even though there may be preponderant evidence to the contrary* based upon equally reliable studies."[47] The Senate Report, which is not itself law, is thus consistent with the general meaning of substantial evidence in that it will allow a claim to prevail even if the preponderance of the evidence indicates that it should fail. "Substantial evidence" is a "highly deferential standard" that is often used by courts when evaluating administrative agency decisions,[48] allowing those decisions to stand even if the court would have reached a different conclusion based upon the same evidence. In the context of pharmaceuticals, the Senate Report explains that the deferential standard was intended to allow approval in "a situation in which a new drug has been studied in a limited number of hospitals and clinics and its effectiveness established only to the satisfaction of a few investigators qualified to use it," or in other words, to ensure that minority viewpoints may be acted upon.[49]

The same statute that sets forth the substantial evidence requirement also elaborates with respect to its meaning,[50] explaining that:

> The term "substantial evidence" means evidence consisting of *adequate and well-controlled investigations, including clinical investigations*, by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, on the basis of which it could fairly and responsibly be concluded

---

that "[t]he substantial evidence standard is a lower standard than [the] weight of the evidence [standard]." *Id.* at 499.

47. S. REP. NO. 87-1744 (1962), *reprinted in* 1962 U.S.C.C.A.N. 2884, 2892 (1962) (emphasis added).

48. *Conn. Light & Power Co.*, 830 A.2d at 1131.

49. S. REP. NO. 87-1744 (1962), *reprinted in* 1962 U.S.C.C.A.N. 2884, 2892 (1962). The Senate Report also suggests that the substantial evidence standard may be important in allowing approval of drugs that help "a substantial percentage of the patients in a given disease condition but [that] will not be effective in other cases." *Id.*

50. *See* 21 U.S.C. § 355(d) (2012) (limiting the definition by its terms to the efficacy portion of the new drug approval statute).

> by such experts that the drug will have the effect it purports or is represented to have . . . .[51]

The usual substantial evidence standard is thus delineated more precisely in the context of drug approval to require "adequate and well controlled investigations, including clinical investigations . . . ."[52] This phrase constitutes the basis for the requirement of clinical (i.e., human) trials. Moreover, because "adequate and well-controlled investigations"[53] is in the plural form, it has been interpreted by the FDA to generally require at least two separate clinical trials,[54] although the FDA will sometimes approve a new drug on the basis of a single study.[55] The practice of approving a new drug on the basis of a single study, already a part of FDA practice, was codified by the FDA Modernization Act of 1997 (FDAMA),[56] which added a provision explicitly empowering the Secretary of Health and Human Services[57] to approve drugs on the basis of a single study under certain circumstances:

> If the Secretary determines, based on relevant science, that data from one adequate and well-controlled clinical investigation and confirmatory evidence (obtained prior to or after such investigation) are sufficient to establish effectiveness,

---

51. *Id*. (emphasis added).

52. *Id*.

53. *Id*.

54. *See* Warner-Lambert Co. v. Heckler, 787 F.2d 147, 151 (3d Cir. 1986) ("Because the Act uses the plural 'investigations,' the FDA requires drug manufacturers to submit at least two 'adequate and well-controlled' studies showing the effectiveness of the drug.").

55. *See* U.S. FOOD & DRUG ADMIN., GUIDANCE FOR INDUSTRY: PROVIDING CLINICAL EVIDENCE OF EFFECTIVENESS FOR HUMAN DRUG AND BIOLOGICAL PRODUCTS 3 (1998) [hereinafter FDA GUIDANCE], http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM078749.pdf (describing scenarios when a single clinical study was approved); New Drug, Antibiotic, and Biological Drug Product Regulations; Accelerated Approval, 57 Fed. Reg. 58,942 (Dec. 11, 1992) (codified at 21 C.F.R. §§ 314 and 601) (noting the FDA's existing practice of occasionally approving drugs on the basis of a single study "where the study was of excellent design [and] showed a high degree of statistical significance").

56. Pub. L. No. 105-115, § 115(a), 111 Stat. 2296 (codified at 21 U.S.C. § 355(d)).

57. *See* 21 U.S.C. § 321(d) (2012) (defining "Secretary" to mean "Secretary of Health and Human Services").

> the Secretary may consider such data and evidence to constitute substantial evidence for purposes of the preceding sentence.[58]

While approving a drug on the basis of a single study might be criticized as constituting a loosening of the standard, the more important concern is that even two trials will not ensure that new drugs possess substantial efficacy because the number of trials relates most directly only to the quantity of evidence rather than to any measure of efficacy level.[59] Drug companies may spend $1 billion over a decade or longer to test a drug for effectiveness in order to satisfy the substantial evidence standard.[60] This expense in both time and money, however, is primarily devoted not to increasing the efficacy of a drug nor even to ensuring that the drug's efficacy meets some minimum threshold (unless "zero efficacy" is considered to be a threshold), but to proving to a reasonable certainty that there is any efficacy at all.[61]

C. The Concepts of Evidence and Efficacy Are Often Conflated

The distinction between substantial evidence and substantial efficacy has not always been appreciated, even by those prominent in the field. David Kessler, for example, during his term as FDA Commissioner, expressed concern over a Congressional bill that would "explicitly lower the efficacy standard" in that it would allow a new drug to be approved based only on a single clinical trial rather than two clinical trials.[62] As just discussed, the number of

---

58. Food and Drug Administration Modernization Act of 1997, Pub. L. No. 105-115, § 115(a), 111 Stat. 2296 (codified at 21 U.S.C. § 355(d)).

59. *See* Adequate and Well-Controlled Studies, 21 C.F.R. § 126(b)(2) (2013) ("An adequate and well-controlled study . . . uses a design that permits a valid comparison with a control to provide a *quantitative* assessment of drug effect." (emphasis added)).

60. *See* Matthew Wynia & David Boren, *Better Regulation of Industry-Sponsored Clinical Trials Is Long Overdue*, 37 J.L. MED. & ETHICS 410, 411 (2009) ("For instance, it might take a billion dollars and ten years or more to bring a drug through testing to market . . . .").

61. Safety, of course, must also be a concern. *See* 21 C.F.R. § 314.124(d) (2013) ("The petitioner shall include in the petition information to show that the drug product was approved for safety and effectiveness . . . .").

62. *FDA Reform Legislation: Hearings Before the Subcomm. on Health and Env't of the H. Comm. on Commerce*, 104th Cong. 40 (1996) (statement of David A. Kessler, M.D., Comm'r of Food & Drugs).

clinical trials relates to the evidence standard and not to any efficacy standard.[63] Similarly, a New Jersey court noted that "[d]rugs approved between 1938 and 1962 were reevaluated [via DESI] to ensure compliance with the efficacy standard."[64] DESI, as discussed above,[65] examined only the substantiality of the evidence of efficacy and did not seek to ensure that the drugs had any particular level of efficacy.[66]

The distinction between substantial evidence (what the standard is) and substantial efficacy (what the standard perhaps ought to be), is reminiscent of the classic distinction made by scientists—and often conflated by laypersons—between the term "precise" and the term "accurate."[67] Scientists traditionally use a dartboard analogy to illustrate the difference: if all darts land close to each other, but far from the bull's eye of the dartboard, the darts have been thrown with precision, but they have not been thrown accurately.[68] See Figure 1 below:

63. *See supra* Part II.B (discussing the evidence–efficacy dichotomy).

64. Bailey v. Wyeth, Inc., 37 A.3d 549, 555 (N.J. Super. Ct. Law. Div. 2008).

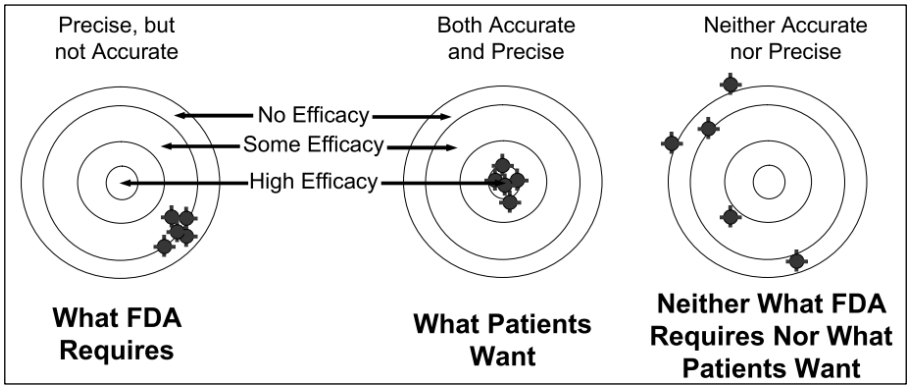
65. *See supra* notes 47–60 and accompanying text (discussing DESI in more depth).

66. Some commentators have also conflated the efficacy standard with the evidence standard. *See, e.g.*, Jennifer Kulynych, *Will FDA Relinquish the "Gold Standard" for New Drug Approval? Redefining "Substantial Evidence" in the FDA Modernization Act of 1997*, 54 FOOD & DRUG L.J. 127, 137–38 (1999) ("CBER reviewers are said to apply the effectiveness standard less rigidly than their CDER counterparts . . . on the basis of a single clinical trial.").

67. *See* AULAY MACKENZIE, MATHEMATICS AND STATISTICS FOR LIFE SCIENTISTS 4 (2005) ("Unfortunately, the terms 'accuracy' and 'precision' are often colloquially used interchangeably, but the distinction in scientific use is important.").

68. *See, e.g.*, B. ANTONISAMY ET AL., BIOSTATISTICS: PRINCIPLES AND PRACTICE 228 (2010) ("Often, a dartboard analogy . . . is . . . used to understand the difference between accuracy and precision.").

Figure 1: FDA Efficacy Standards Using the Dartboard Analogy[69]



Patients want drugs that reach reasonable efficacy targets, or, to use the language of the preceding analogy, a drug whose efficacy is accurate.[70] What the legal standard requires, however, is a drug whose efficacy is precise.[71] This is hardly a distinction that the average doctor likely considers on a daily basis, and is unlikely to be one that the average patient pauses to think about at all.[72] Nevertheless, it is an important distinction to make in order to ensure that patients, physicians, and others are not misled into believing mistakenly that regulatory evidence standards ensure that only drugs with high levels of efficacy are approved.[73]

---

69. This is the classic bull's-eye dartboard analogy depicting accuracy versus precision.

70. *See supra* notes 67–69 and accompanying text (illustrating the widely accepted "dartboard analogy").

71. *See* SHAYNE COX GAD, CLINICAL TRIALS HANDBOOK 442 (2009) (noting that it is the confidence interval, and not the *p*-value per se, that is a measure of precision).

72. Physicians may be less informed about drug efficacy than is commonly believed. *See* FOOD AND DRUG ADMIN. & DIV. OF MED. SCIES. NAT'L RESEARCH COUNCIL, DRUG EFFICACY STUDY: FINAL REPORT TO THE COMMISSIONER OF FOOD AND DRUGS 65 (1969) [hereinafter DESI FINAL REPORT] (alluding to "prescribing doctors who were not in a position of knowledge" with respect to actual drug efficacy and were therefore heavily influenced by advertising); *see also* Ben Goldacre, *What Doctors Don't Know About the Drugs They Prescribe*, TEDTALK (Apr. 5, 2013), http://www.huffingtonpost.com/ben-goldacre/prescription-drugs_b_3018272.html?icid=maing-grid7|main5|dl30|sec1_lnk3%26pLid%3D295315 (last visited Oct. 21, 2013) (noting that physicians would be "misled" due to publication bias, even if they were to review the literature) (on file with the Washington and Lee Law Review).

73. If academic institutions were to operate under legal standards

III. The Gold Standard and New Drug Approval

The very term "gold standard" evokes feelings of confidence and certainty. As one dictionary defines it, the gold standard is "the best, most reliable, or most prestigious thing of its type."[74] Certainly, if one were measuring something as important as drug efficacy, it would be desirable to use such a pristine and highly regarded standard, and the relevant FDA regulations do not disappoint. These regulations incorporate, as a general matter, all of the essential elements normally considered to be part of the gold standard, as will be explained next.[75]

A. FDA Regulations Generally Define "Adequate and Well-Controlled Investigations" According to the Gold Standard

Although the term "gold standard" is not used anywhere in the federal statute or in the accompanying regulations, the section of the regulations entitled "adequate and well-controlled studies" sets forth the three elements generally understood to constitute the core of the gold standard for clinical trials: randomization, double-blind administration, and placebo-control.[76]

---

analogous to those that bind the FDA, a student could pass a course with a failing grade so long as test results provided "substantial evidence"—to acceptable levels of certainty—that the student's grasp of the course material was slightly above absolute incompetence. To extend the analogy, professors would then provide these marginal students with glowing letters of recommendation and insist that any future employer compensate them at exorbitant annual salaries. These annual salaries would be so high that in some cases they could only be paid with the help of government-regulated third parties who would provide at least partial reimbursement to the employers. If those third parties refused to pay, in light of the student's (now graduate's) test results, those third parties would be condemned as being cold and calculating, or even immoral. The analogy would be absurd and worthy of summary dismissal were it not so suggestive of what actually occurs with pharmaceuticals.

74. *Gold Standard Definition*, OXFORD DICTIONARIES, http://oxforddictionaries.com/definition/gold%2Bstandard?region=us (last visited Sept. 8, 2013) (on file with the Washington and Lee Law Review).

75. *See* 21 C.F.R. § 314.126 (2013) (containing the elements of the scientific gold standard: randomization, double-blind administration, and placebo control). For an explanation of the gold standard elements, see *infra* Part III.A.

76. *See, e.g.*, FTC v. QT, Inc., 448 F. Supp. 2d 908, 938 (N.D. Ill. 2006) ("Dr. Feldstein agrees that a double-blind, placebo-controlled, randomized trial is the

According to the regulations, "adequate and well-controlled studies," more commonly known as clinical trials, must "us[e] a design that permits a valid comparison with a control to provide a quantitative assessment of drug effect."[77] That is, study participants taking the new drug must be compared to a scientific "control" group, which in its most basic form consists of patients taking an inactive placebo, the appearance of which is indistinguishable from the active treatment.[78] The purpose of placebo control (combined with blinding, discussed below)[79] is to address response bias, which is the tendency of study participants to respond because they believe they are being treated.[80] The use of placebo thus helps to determine the extent of the effect caused by the chemical composition of the drug itself.[81] Note that, although the regulations indicate that the drug effect should be assessable quantitatively, there is no minimum quantum of effect required.[82] In other words, efficacy must be measured, but it need not be measured against any standard (other than zero).

Researchers in clinical trials must ensure more generally that "[a]dequate measures are taken to minimize bias on the part of the subjects, observers, and analysts of the data."[83] The most common means for avoiding bias, and the one specifically

---

'gold standard' in the scientific community . . . ."); DAVID MISCHOULON & JERROLD F. ROSENBAUM, NATURAL MEDICINES FOR PSYCHIATRIC DISORDERS 12 (2008) ("Double-blind, placebo-controlled, randomized clinical trials (RCTs) have been the gold standard of clinical research for several decades."); David W. Barnes, *General Acceptance Versus Scientific Soundness: Mad Scientists in the Courtroom*, 31 FLA. ST. U. L. REV. 303, 320 (2004) ("The *gold standard* for experimental testing that will yield an error rate measuring reliability is the randomized, controlled, double-masked study." (emphasis in original)).

77. 21 C.F.R. § 314.126(b)(2) (2013).

78. Laura Lee Johnson et al., *An Introduction to Biostatistics: Randomization, Hypothesis Testing, and Sample Size Estimation*, *in* PRINCIPLES AND PRACTICE OF CLINICAL RESEARCH 165, 167 (John I. Gallin & Frederick P. Ognibene eds., 2d ed. 2007).

79. *See infra* notes 83–96 and accompanying text (discussing placebo control and blind testing).

80. Johnson et al., *supra* note 78, at 167.

81. *Id.* at 165–67.

82. *See* 21 C.F.R. § 314.126(b)(2) (2013) (neglecting to set a minimum quantitative efficacy level).

83. *Id.* § 314.126(b)(5).

mentioned in the regulations, is "blinding,"[84] sometimes referred to as "masking."[85] A "single-blind" study involves taking measures to ensure that patients do not know whether they are receiving the active treatment or the placebo treatment, which helps avoid response bias.[86] The more highly regarded "double-blind" study involves utilizing a protocol that ensures that neither the patient nor the treating researcher knows whether the active or placebo treatment is being administered.[87] Blinding the researcher helps to avoid observer bias, a type of bias introduced when the researcher manages, treats, or assesses an outcome differently depending on whether the researcher believes an active treatment is being administered.[88] While most trials follow the double-blind approach,[89] some trials follow a "triple-blind" approach in which members of the committee that monitors the patient responses are also unaware of which patients are receiving the treatment.[90]

A third characteristic of adequate and well-controlled investigations, as defined by the FDA regulations, is that patients must be assigned to treatment or placebo groups so as to minimize bias.[91] The regulations note that minimizing bias is "ordinarily" achieved by "randomization," that is, assigning patients randomly

---

84. *Id*.

85. MIQUEL S. PORTA, A DICTIONARY OF EPIDEMIOLOGY (Oxford Univ. Press 2008) ("To avoid confusion about the meaning of the word *blind*, some authors prefer to describe such studies as *masked*.").

86. LAWRENCE M. FRIEDMAN & CURT D. FURBERG, FUNDAMENTALS OF CLINICAL TRIALS 120–21 (2010).

87. *Id.* at 122 ("The main advantage of a truly double-blind study is that the risk of bias is reduced.").

88. U.S. FOOD & DRUG ADMIN., INT'L CONF. ON HARMONIZATION, GUIDANCE FOR INDUSTRY: E 10 CHOICE OF CONTROL GROUP AND RELATED ISSUES IN CLINICAL TRIALS 4 (2001). *See* Bradley Evanoff, *Reducing Bias*, *in* TRANSLATIONAL AND EXPERIMENTAL CLINICAL RESEARCH 67, 68–69 (Daniel P. Schuster & William J. Powers eds., 2005) (discussing various types of bias and techniques to counter them).

89. *See* U.S. FOOD & DRUG ADMIN., INT'L CONF. ON HARMONIZATION, GUIDANCE FOR INDUSTRY: E 9 STATISTICAL PRINCIPLES FOR CLINICAL TRIALS 10 (1998) ("Most . . . trials follow a double-blind approach . . . .").

90. *See* FRIEDMAN & FURBERG, *supra* note 86, at 123 ("A triple-blind study is an extension of the double-blind design; the committee monitoring response variables is not told the identity of the groups.").

91. 21 C.F.R. § 314.126(b)(4) (2013).

to either the treatment group or the control group.[92] Randomization helps to counter selection bias, which occurs when nonrandom selection results in active and control groups that are not statistically representative of the population under study.[93] In the case of clinical trials, for example, selection bias could occur if patients placed on active treatment (i.e., the drug being tested) are on average more severely afflicted with disease, or more susceptible to treatment, than are those in the control group. Randomization helps to ensure against such bias.[94]

The regulations thus make reference to all three core elements of the gold standard, and these three elements help to counter a number of biases that can undermine the validity of trial outcomes.[95] Figure 2 summarizes these elements and the biases they are designed to address. Note that even the total elimination of these biases does not necessarily imply any minimum threshold of drug efficacy. Instead, mitigating bias helps to ensure that results are correct, regardless of whether the results reveal great efficacy, little efficacy, no efficacy, or even negative efficacy (i.e., harm).

Figure 2: Research Biases and Means of Addressing Them[96]

Problem	**Regulatory Solution**
Biases, e.g. • Selection bias • Observer bias • Response bias	Gold standard • Randomization • Blinding • Placebo control

B. FDA Regulations Do Not Necessarily Require the Gold Standard

Although FDA regulations describe the three main elements of the gold standard in defining "adequate and well-controlled

92. *Id*.

93. Ian S. Lustick, *History, Historiography, and Political Science: Multiple Historical Records and the Problem of Selection Bias*, 90 AM. POL. SCI. REV. 605, 606 (1996).

94. *See* Johnson et al., *supra* note 78, at 167.

95. *See* 21 C.F.R. § 314.126(b) (2013) (requiring randomization, blinding, and placebo control).

96. *See generally supra* Part II.A.

investigations,"[97] the regulations contain considerable flexibility and new drugs can in fact be approved without being subjected to gold-standard testing.[98] For example, although placebo-control is usually considered to be a key element of the gold standard, the regulations specifically contemplate the use of other types of controls, including: (1) "dose-comparison" controls, where patients are divided into groups and given different doses of the drug being tested;[99] (2) "active treatment" controls, where the control group is given a different active ingredient (such as a previously-approved alternate drug) from that of the test group;[100] (3) "no treatment" controls, reserved for cases where objective measurements of efficacy are available;[101] and (4) "historical" controls, where data from the test group is compared to historical data of patients not associated with the current trial.[102] "Uncontrolled" studies are specifically noted to be unacceptable.[103]

The FDA regulations also allow flexibility with respect to both blinding and randomization.[104] These techniques are offered in the regulations as examples of means to avoid bias, but they are not indicated as requirements.[105] Moreover, even if they were nominally required, the regulations provide that the Director for the Center for Drug Evaluation and Research (CDER, a division of the FDA) "may . . . waive in whole or in part any of the criteria

---

97. 21 C.F.R. § 314.126 (2013).

98. *See infra* notes 99–107 and accompanying text (discussing the flexibility of the testing process under the current regulations).

99. 21 C.F.R. § 314.126(b)(2)(ii).

100. *Id.* § 314.126(b)(2)(iv).

101. *Id.* § 314.126(b)(2)(iii).

102. *Id.* § 314.126(b)(2)(v).

103. *Id.* § 314.126(e).

104. *See id.* § 314.126(b)(2)(i) ("A placebo-controlled study . . . *usually* includes randomization and blinding . . . ." (emphasis added)); *id.* § 314.126(b)(2)(ii) ("Dose comparison trials *usually* include randomization and blinding . . . ." (emphasis added)); *id.* § 314.126(b)(2)(iii) ("No treatment concurrent control trials *usually* include randomization." (emphasis added)); *id.* § 314.126(b)(2)(iv) ("Active treatment trials *usually* include randomization and blinding . . . ." (emphasis added)); *id.* § 314.126(b)(5) (offering "blinding" as an *example* of a measure taken to avoid bias); *id.* § 314.126(b)(4) ("*Ordinarily*, in a concurrently controlled study, assignment is by randomization . . . ." (emphasis added)).

105. *Id.* § 314.126(b)(2)(i).

[i.e., placebo-control, randomization, and blinding] . . . either prior to the investigation or in the evaluation of a completed study."[106]

The flexibility provided by the regulations is not necessarily a cause for concern. New drugs are submitted to the FDA under a wide variety of circumstances, and flexibility is therefore needed to appropriately tailor clinical trials to each drug and disease against which it is being tested.[107] At the same time, the brief review just presented makes clear that not all new drugs are necessarily subjected to gold standard testing. The important point, however, is that even if all new drugs were subject to the gold standard, that standard would not be sufficient to ensure that only meaningfully effective drugs reached the market.

IV. The Gold Standard Does Not Ensure Substantial Efficacy

The most significant weakness of the gold standard is not that the standard may not always be required,[108] but the insufficiency of the gold standard itself. There are two reasons for this insufficiency. First, the statistical framework supporting the gold standard does not account for the possibility that drug companies may undertake multiple trials until one or more of them demonstrates efficacy.[109] Second, and more importantly, the standard is inadequate because its statistical framework requires no particular level of efficacy.[110] All that is required is reasonable

---

106. *Id.* § 314.126(c).

107. *See generally* Barry S. Roberts & Sara M. Biggers, *Regulatory Update: The FDA Speeds Up Hope for the Desperately Ill and Dying*, 27 AM. BUS. L.J. 403 (1989) (arguing that an experimental drug to treat AIDS during the 1980s benefited from the flexibility of the clinical trial process mandated by the FDA).

108. In some cases, insisting on the gold standard would be unethical. For example, when testing a new drug against a deadly condition for which there is already an effective treatment, it would be unethical to condemn half of the patients in the clinical trial to a placebo group. Instead, the existing treatment could be used as a comparator.

109. Each controlled clinical study pertinent to a proposed use of a new drug must be submitted to the FDA as part of a New Drug Application. 21 C.F.R § 314.50(d)(5) (2013). Thus, although companies cannot "cherry-pick" the two best trials to submit to the FDA while withholding others from its consideration, the FDA may approve a drug based upon two positive clinical trials notwithstanding that additional trials known to the FDA failed to show promising results. *See infra* Part V.C.

110. *See* 21 C.F.R. § 314.126(b)(2) (2013) (neglecting to set a minimum

certainty as to the existence of *any* level of efficacy.[111] Each of these shortcomings is explored in turn.

A. The Gold Standard Does Not Prevent the Undertaking of Multiple Trials

Each year, hundreds of clinical trials are undertaken with the hope that favorable results can be used to win FDA approval.[112] At a statistical certainty level of p=.05 (and assuming a one-tailed distribution),[113] the FDA could expect that one in every twenty trials would be positive, even if all of the drugs tested had no efficacy whatsoever.[114] Such periodic "false positives," or Type I errors, are due to the natural random variation of clinical trial results combined with the tolerances that are inherent in the statistical measurements.[115] In order to reduce the possibility that a drug is approved based on a Type I error, the statute generally requires drug sponsors to conduct two clinical trials rather than one.[116] Assuming those clinical trials are unbiased and independent, the two-trial requirement reduces the chance of a Type I error from one in twenty to one in four hundred,[117] substantially reducing but still not eliminating the chance that a

---

quantitative efficacy level); *supra* Part III (evaluating the elements of the scientific gold standard as required by the FDA in clinical drug trials).

111. *See* 21 C.F.R. § 314.126(b)(2) (2013) (neglecting to set a minimum quantitative efficacy level); *supra* Part II (evaluating the level of efficacy required by standard clinical trials).

112. FDA GUIDANCE, *supra* note 55, at 5.

113. *See* U.S. FOOD & DRUG ADMIN., INT'L CONF. ON HARMONIZATION, GUIDANCE FOR INDUSTRY: E9 STATISTICAL PRINCIPLES FOR CLINICAL TRIALS 32 (1998), http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073137.pdf (noting that "[t]he issue of one-sided or two-sided approaches to inference is controversial, and a diversity of views can be found in the statistical literature").

114. FDA GUIDANCE, *supra* note 55, at 5, n.5.

115. *Id.* at 31–32.

116. *Id.* at 5 ("Confirmatory trials are intended to provide firm evidence in support of claims; hence adherence to protocols and standard operating procedures is particularly important.").

117. The figure of one in four hundred is calculated as the product of the p-values of each individual trial, or 1 in 400 assuming a p-value of .05 (that is: 0.05 * 0.05 = .0025, or 1 in 400).

drug under investigation will appear to be effective when in fact it is not.

A larger challenge, however, derives not from statistical uncertainty, which can never be absolutely eliminated, but from the deliberate actions of study sponsors. In theory, a drug sponsor could simply run clinical trials in sequence, stopping only after two positive trials have been accumulated.[118] Although drug sponsors must submit pertinent information from all studies to the FDA as part of the New Drug Application process,[119] data from the two trials showing statistically significant results could be relied upon by the FDA as substantial evidence of efficacy in order to approve the drug under the applicable legal standard, notwithstanding its knowledge of less flattering data from other trials.[120] The requirements of the gold standard itself do not stand in the way of such multiple sequential trials.[121] The fact that the FDA can approve new drugs after two positive trials are accumulated, even though other trials failed to show efficacy, is entirely consistent with the intent of Congress as expressed in the 1962 Senate Report quoted above: drugs may be approved "*even though there may be preponderant evidence to the contrary* based upon equally reliable studies."[122] More generally, it is consistent with the substantial

---

118. *See* MELODY PETERSON, OUR DAILY MEDS: HOW THE PHARMACEUTICAL COMPANIES TRANSFORMED THEMSELVES INTO SLICK MARKETING MACHINES AND HOOKED THE NATION ON PRESCRIPTION DRUGS 49 (2008) ("[T]he sponsor could just do studies until the cows come home until he gets two of them that are statistically significant by chance alone . . . .").

119. *See* 21 C.F.R § 314.50(d)(5) (2013) (describing the requirements of the New Drug Application process); 21 C.F.R. § 314.125(b)(14) (2013) ("The [FDA] will refuse to approve the application . . . if . . . [t]he application does not contain an explanation of the omission of a report of any investigation of the drug product sponsored by the applicant, or . . . of other [pertinent] information.").

120. *See* 21 U.S.C. § 355(d) (2012) ("[D]ata from one adequate and well-controlled clinical investigation and confirmatory evidence . . . are sufficient to establish effectiveness . . . .").

121. *Supra* Parts II–IV.

122. S. REP. NO. 87-1744, at 16 (1962), *reprinted in* 1962 U.S.C.C.A.N. 2884, 2892 (emphasis added). The Senate Report continued:

> There may also be a situation in which a new drug has been studied in a limited number of hospitals and clinics and its effectiveness established only to the satisfaction of a few investigators qualified to use it. There may be many physicians who would deny the effectiveness simply on the basis of a disbelief growing out of their past experience with other drugs or with the diseases involved.

evidence standard, which requires less than a preponderance of the evidence, as discussed above.[123]

1. Using Multiple Trials to Obtain Drug Approval: The Antidepressants

At least some borderline-ineffective medicines appear to have been approved by undertaking multiple trials until sufficient positive evidence is accumulated. Irving Kirsch and a group of scholars at the University of Connecticut and the George Washington School of Public Health used the federal Freedom of Information Act (FOIA) to obtain data from clinical trials of the six most widely prescribed depression medicines.[124] Given the legal requirement that two favorable trials be submitted to the FDA to satisfy the substantial evidence standard, one might expect the FOIA request to have yielded data from the twelve trials necessary to obtain FDA approval.[125] In fact, the FOIA request submitted by Kirsch and his colleagues uncovered an astonishing forty-seven randomized and placebo controlled efficacy trials for the six medicines in question: five trials each for fluoxetine (Prozac) and citalopram (Celex); six for venlafaxine (Effexor); seven for

---

> Again, the studies may show that the drug will help a substantial percentage of the patients in a given disease condition but will not be effective in other cases. What the committee intends is to permit the claim for this new drug to be made to the medical profession with a proper explanation of the basis on which it rests.

Id.; *see also* Karen Baswell, Note, *Time for a Change: Why the FDA Should Require Greater Disclosure of Differences of Opinion on the Safety and Efficacy of Approved Drugs*, 35 HOFSTRA L. REV. 1799, 1826 (2007) (interpreting the Senate Report to mean that, under the standard expressed in that Report, "differences in opinion would always be resolved in favor of drug approval").

123. *See supra* Part II (discussing the evidentiary standard).

124. Irving Kirsch et al., *The Emperor's New Drugs: An Analysis of Antidepressant Medication Data Submitted to the U.S. Food and Drug Administration*, 5 PREVENTION & TREATMENT, no. 1, 2008, at 3 [hereinafter Kirsch et al., *Emperor's*], http://alphachoices.com/repository/assets/pdf/EmperorsNewDrugs.pdf.

125. *See* Food and Drug Administration Modernization Act of 1997, Pub. L. No. 105-115, § 115(a), 111 Stat. 2296 (codified at 21 U.S.C. § 355(d)) ("[D]ata from one adequate and well-controlled clinical investigation and confirmatory evidence . . . are sufficient to establish effectiveness . . . .").

sertraline (Zoloft); eight for nefazodone (Serzone); and sixteen for paroxetine (Paxil).[126]

While there may have been legitimate reasons for conducting the plethora of clinical trials observed by Kirsch et al.,[127] the data they report suggest one obvious possibility, namely, that many of the trials were not as favorable as the clinical researchers had hoped.[128] Earlier work by Kirsch and Guy Sapirstein that relied on studies that had been published suggested that at least 75% of the benefit supposedly due to the antidepressants was also seen in the placebo group.[129] According to Kirsch, the data from the unpublished studies uncovered via the FOIA request cast an even darker shadow over the efficacy claims of the depression medications: "More than half of the clinical trials [of antidepressants] sponsored by the pharmaceutical companies showed no significant difference at all between drug and placebo."[130] Although Kirsch and his colleagues advocate caution in interpreting the results, they note that the data "suggest that the effect[s] of antidepressant drugs are very small and of questionable

---

126. *Id.*; *see also* Eric H. Turner et al., *Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy*, 358 NEW ENG. J. MED. 252, 256 (2008) (examining seventy-four FDA-registered studies of twelve depression medications and concluding that thirty-seven of the thirty-eight positive studies were published while, of the negative studies, twenty-two were not published and eleven were published in a way that misleadingly "conveyed a positive outcome").

127. *See, e.g.*, FDA GUIDANCE, *supra* note 55, at 6 (1998) (noting that multiple trials may be conducted in order "to find an appropriate dose, to study patients of greater and lesser complexity or severity of disease, to compare the drug to other therapy, to study an adequate number of patients for safety purposes," and to otherwise study a drug).

128. *Infra* notes 129–30 and accompanying text.

129. *See* Irving Kirsch & Guy Sapirstein, *Listening to Prozac but Hearing Placebo: A Meta-Analysis of Antidepressant Medication*, 1 PREVENTION & TREATMENT, no. 2, 1998, Article 2a (comparing the effects of active drugs and placebos).

130. Irving Kirsch, *Antidepressants: The Emperor's New Drugs?*, HUFFINGTON POST (Jan. 20, 2010, 1:35 PM) [hereinafter Kirsch, *Antidepressants*], http://www.huffingtonpost.com/irving-kirsch-phd/antidepressants-the-emper_b_442205.html (last visited Sept. 21, 2013) (on file with the Washington and Lee Law Review); s*ee also* Arif Khan et al., *Severity of Depression and Response to Antidepressants and Placebo: An Analysis of the Food and Drug Administration Database*, 22 J. CLINICAL PSYCHOPHARMACOLOGY 40 (2002) (analyzing the relationship between the severity of depression symptoms and clinical trial results).

clinical significance."[131] Kirsch, now a faculty member at Harvard Medical School,[132] would later share his findings with the broader public in his book *The Emperor's New Drugs: Exploding the Antidepressant Myth*.[133]

2. The Larger Phenomenon: Publication Bias

The selective use of multiple trials to obtain drug approval is part of a larger phenomenon of publication bias, i.e., where studies with positive results are more likely to be published than studies with negative or inconclusive results.[134] Critics have warned of publication bias for decades.[135] An influential 1988 paper by researchers at the Dana-Farber Cancer Institute surveyed the literature on publication bias and concluded that "the problem [of publication bias] is indeed a serious one."[136] The following year, some of the same researchers attempted to quantify the magnitude of publication bias in published cancer clinical trials, concluding that small studies are more prone to bias than larger ones and finding the bias to be very large.[137] Small trials might be more subject to publication bias in part because their small size makes their non-publication less likely to attract notice. In addition, the lower cost of small trials means that it is easier to shine a media spotlight on favorable ones and simply write off negative trials as a

131. Kirsch et al., *Emperor's*, *supra* note 124, at 8*; see also id.* at 10 ("[C]linical significance is dubious.").

132. Program in Placebo Studies & Therapeutic Encounter (PiPS), *Our Team*, http://programinplacebostudies.org/about/people/ (last visited Sept. 21, 2013) (on file with the Washington and Lee Law Review).

133. IRVING KIRSCH, THE EMPEROR'S NEW DRUGS: EXPLODING THE ANTIDEPRESSANT MYTH (2010).

134. *See* I. Peterson, *Publication Bias: Looking for Missing Data*, SCI. NEWS, Jan. 7, 1989, at 5 (1989) (discussing disproportionate publication of studies with positive results and suggestions for reform).

135. *Infra* note 136 and accompanying text.

136. Colin B. Begg & Jesse A. Berlin, *Publication Bias: A Problem in Interpreting Medical Data*, 151 J. ROYAL STAT. SOC'Y [SERIES A] 419, 421 (1988); *see also* Goldacre, *infra* note 148 (discussing the effects of publication bias on doctors, patients, and the industry in general).

137. *See* Jesse A. Berlin et al., *An Assessment of Publication Bias Using a Sample of Published Clinical Trials*, 84 J. AM. STAT. ASS'N 381, 391 (1989) ("The results appear to confirm that small studies are more prone to bias and that the bias is substantial . . . .").

cost of doing business. Statistically, small trials tend to vary more in outcome and therefore are more likely than large trials to sometimes produce misleadingly favorable results.[138]

3. Statutory Attempts to Address Publication Bias

Concerns about the use of multiple trials to game the system were ostensibly addressed by legislation that mandated increased transparency of clinical trials. The Food and Drug Administration Modernization Act of 1997 (FDAMA)[139] tasked the Secretary of Health and Human Services with the creation of a publicly-accessible data bank of clinical trials of drugs for serious or life-threatening diseases or conditions.[140] The data bank is now available at ClinicalTrials.gov.[141] In 2007, the Food and Drug Administration Amendments Act (FDAAA) significantly expanded the scope of the data bank such that it now includes clinical trials for drugs treating other than serious or life-threatening diseases or conditions.[142] Both federally and privately funded trials are required to be reported,[143] including many trials that are conducted overseas.[144] The rationale behind this legislation is that

---

138. *See id.* at 391 (evaluating the effect of small sample size on generating positive clinical trial results).

139. Food and Drug Administration Modernization Act of 1997, Pub. L. No. 105-115, § 113, 111 Stat. 2296 (1997) (codified as amended at 42 U.S.C. § 282 (2012)).

140. *Id*. § 113, 111 Stat. at 2311.

141. The European Union maintains a similar database called EudraCT (European Union Drug Regulating Authorities Clinical Trials). EU CLINICAL TRIALS REGISTER, https://www.clinicaltrialsregister.eu (last visited Oct. 21, 2013) (on file with the Washington and Lee Law Review).

142. *See* Tamsen Valoir & Shubha Ghosh, *FDA Preemption of Drug & Device Labeling: Who Should Decide What Goes on a Drug Label?*, 21 HEALTH MATRIX 555, 589–90 (2011) (citing Food and Drug Administration Amendments Act of 2007, Pub. L. No. 110-85, § 801(a), 121 Stat. 823 (2007)).

143. U.S. FOOD & DRUG ADMIN., GUIDANCE FOR THE INDUSTRY: INFORMATION PROGRAM ON CLINICAL TRIALS FOR SERIOUS OR LIFE-THREATENING DISEASES AND CONDITIONS 2 (2002), http://www.fda.gov/downloads/regulatoryinformation/guidances/ucm126838.pdf (citing the FDA Modernization Act of 1997, Pub. L. No. 105-115, § 113, 111 Stat. 2296 (codified as amended at 42 U.S.C. § 282(j)(3)(A)) (2012)).

144. *See id*. at 6 (discussing reporting requirements for overseas clinical trials of investigational new drugs).

if clinical trials are registered before results are obtained, it will be more difficult to hide unfavorable trial results.[145]

Unfortunately, ClinicalTrials.gov has in practice proven far less effective than the rhetoric suggests, or even than the law requires. According to a 2012 study in the British Medical Journal, almost four out of five clinical trials covered by FDAAA that should have been reported on ClinicalTrials.gov were not.[146] Despite what appear to be flagrant violations of the law and authorized penalties of up to $10,000 per day,[147] the New York Times recently reported that, "[a]mazingly, no fine has yet been levied."[148]

Efforts outside of FDAAA to address selective publication have met with similarly limited success. A 2004 policy announcement by the International Committee of Medical Journal Editors—an organization whose membership includes the New England Journal of Medicine, the Journal of the American Medical Association, and other leading publishers of clinical trial results[149]—to refrain from publishing papers that are based upon

---

145. *See* 153 Cong. Rec. S1411-02, S1448 (Jan. 31, 2007) (statement of Christopher Dodd) ("We owe it to patients to make sure that their participation in a trial will benefit other individuals suffering from the same illness or condition by making the results of the trial public, no matter the outcome of the trial.").

146. *See* Andrew P. Prayle et al., *Compliance with Mandatory Reporting of Clinical Trial Results on ClinicalTrials.gov: Cross Sectional Study*, 344 BRIT. MED. J. d7373, at 3 (Jan. 3, 2012) (stating that approximately 78% of trials from 2009 were not reported); *see also* Kay Dickersin & Drummond Rennie, *The Evolution of Trial Registries and Their Use to Assess the Clinical Trial Enterprise*, 307 J. AM. MED. ASS'N 1861, 1861 (2012) ("[T]rial investigators . . . largely ignored any legal requirement to register [under FDAMA]."); Fiona Godlee, *Research Misconduct Is Widespread and Harms Patients*, 344 BRIT. MED. J. e14 (Jan. 5, 2012) (emphasizing reporting bias and suppression); Joseph S. Ross et al., *Publication of NIH Funded Trials Registered in ClinicalTrials.gov: Cross Sectional Analysis*, 344 BRIT. MED. J. d7292, at 3 (Jan. 3, 2012), http://www.bmj.com/content/344/bmj.d7292.pdf%2Bhtml (finding that a third of NIH funded trials remain unpublished more than three years after study completion).

147. 21 U.S.C. § 333(f)(3)(B) (2012).

148. Ben Goldacre, *Health Care's Trick Coin*, N.Y. TIMES, Feb. 2, 2013, at A23.

149. Int'l Comm. Med. J. Editors, *About the International Committee of Medical Journal Editors*, ICMJE, http://www.icmje.org/about.html (last visited Sept. 21, 2013) (on file with the Washington and Lee Law Review).

unregistered studies,[150] has not been adequately followed by journal editors themselves.[151]

Earlier in 2004, New York State attorney general Eliot Spitzer had sued GlaxoSmithKline alleging that it fraudulently concealed negative efficacy (and safety) data about antidepressant Paxil (paroxetine) through selective publication.[152] Glaxo's stock was pushed down by 3.2%,[153] likely based more on anticipated court damages or fines rather than on any anticipated decrease in sales;[154] sales of Paxil remained over $200 million per year.[155] In 2012, Glaxo finally paid $3 billion in fines to settle criminal and civil charges over sales and marketing practices related to several of its drugs, including Paxil,[156] but less than a third of this amount

---

150. *See* Catherine D. DeAngelis et al., *Clinical Trial Registration: A Statement From the International Committee of Medical Journal Editors*, 292 J. AM. MED. ASS'N 1363 (2004) (discussing registration of trials as a requirement for publication in their journals); *see also* Int'l Comm. Med. J. Editors, *Obligation to Register Clinical Trials*, ICMJE, http://www.icmje.org/publishing_10regi ster.html (last visited Sept. 21, 2013) (setting forth the current policy) (on file with the Washington and Lee Law Review).

151. *See* Sylvain Mathieu et al., *Comparison of Registered and Published Primary Outcomes in Randomized Controlled Trials*, 302 J. AM. MED. ASS'N 977, 981 (2009) (noting that fewer than half of trials were adequately registered, and more than a quarter were unregistered).

152. Gardiner Harris, *Spitzer Sues a Drug Maker, Saying It Hid Negative Data*, N.Y. TIMES, June 3, 2004, at A1. *See generally* Abigail Kagle, Driven to Settle: Elliot Spitzer v. GlaxoSmithKline and Undisclosed Trial Data Regarding Paxil (unpublished student paper) (on file with the Columbia Law School Attorney General Program), http://web.law.columbia.edu/sites/default/files/microsites/attorneys-general/files/Eliot-Spitzer-v-GlaxoSmithKline.pdf.

153. Kagle, *supra* note 152, at 26.

154. *See* GLAXOSMITHKLINE PLC, ANNUAL REPORT 2004, 120 (Mar. 4, 2005), http://www.gsk.com/content/dam/gsk/globals/documents/pdf/annual-report-2004.pdf (noting that the New York State lawsuit was settled in August 2004 but that "similar cases, some of which purport to be class actions, have been filed in state and federal courts"); *id.* at 145 (noting that impairments of £633 million were recorded following the launch in the U.S. of a generic version of Paxil).

155. *See* Aaron Smith, *Suicide Label Unlikely to Squeeze Glaxo Sales*, CNN MONEY (May 12, 2006, 1:28 PM), http://money.cnn.com/2006/05/12/news/companies/paxil/index.htm (last visited Sept. 21, 2013) (noting sales of Paxil and Paxil CR totaled $242 million in 2005) (on file with the Washington and Lee Law Review). These sales levels were achieved notwithstanding the entry of generic competition to Paxil in September 2003. *See* ANNUAL REPORT OF 2004, *supra* note 154, at 79 (noting decreased pharmaceutical turnover in the United States after September 2003).

156. Katie Thomas & Michael S. Schmidt, *Glaxo Agrees to Pay $3 Billion in*

would stem from Paxil (paroxetine), a product that has earned Glaxo at least $11.6 billion.[157]

Unfortunately, the public's conscience seems to be less sensitive to questionable but important selective publication practices than to the lurid personal dalliances of those attempting to check such practices: Spitzer, a consumer protection champion, was exposed in 2008 as a client in a prostitution scandal and charged under a century-old federal law that is rarely invoked against prostitution clients,[158] thus ending his otherwise admirable career as a public advocate.[159] At the same time, even billions of dollars in fines do not seem to be deterring misconduct.[160] Because "[t]he benefits of aggressive marketing often outweigh the cost of settlements," even those in the billion-dollar range, "it is not clear that companies are changing their ways," notes *The Economist* magazine.[161]

Although ClinicalTrials.gov and other efforts to reduce publication bias are important steps in the right direction, they would not change the underlying standard for drug approval, even if successful.[162] According to Congressional testimony, one reason to make trial results publicly accessible is to counter the "bias toward publication of positive results [and the suppression of

---

Fraud Settlement, N.Y. TIMES, July 3, 2012, at A1, *available at* http://www.nytimes.com/2012/07/03/business/glaxosmithkline-agrees-to-pay-3-billion-in-fraud-settlement.html?_r=0.

157. *Id.*

158. *See* Danny Hakim & William K. Rashbaum, *Spitzer Is Linked to Prostitution Ring*, N.Y. TIMES, Mar. 10, 2008, http://www.nytimes.com/2008/03/10/nyregion/10cnd-spitzer.html?pagewanted=all&r_=0 (last visited Sept. 21, 2013) (discussing the allegations against Mr. Spitzer, his press release, and the potential career repercussions) (on file with the Washington and Lee Law Review).

159. *See* Michael M. Grynbaum, *Spitzer Resigns, Citing Personal Failings*, N.Y. TIMES (Mar. 12, 2008), www.nytimes.com/2009/03/12/nyregion/12end-resign.html?pagewanted=all (last visited Sept. 21, 2013) ("Governor Spitzer, whose rise to political power as a fierce enforcer of ethics in public life was undone by revelations of his own involvement with prostitutes, resigned on Wednesday . . . .") (on file with the Washington and Lee Law Review).

160. *Infra* note 161 and accompanying text.

161. *Johnson & Johnson Out of the Mire?: The Justice Department May Spoil the Drugmaker's Fresh Start*, ECONOMIST, Apr. 28, 2012, at 75, *available at* http://www.economist.com/node/21553512.

162. *See supra* Part II.A (discussing the current legal standard for drug approval).

negative results that] may distort the community's overall understanding of an intervention's effectiveness or risk profile."[163] Although the legislation increases the accessibility of clinical trial data and makes it more difficult to hide the results of unfavorable trials, it does not prevent companies from continuing to undertake trials until achieving two that are positive.[164]

In addition, the advent of ClinicalTrials.gov is a change in degree, but not in kind. Clinical trial information submitted to government agencies such as the FDA has long been available to the public, to the extent it is not secret or otherwise exempted, under the Freedom of Information Act.[165] This availability is reflected in the work of Kirsch and his colleagues noted above.[166] Moreover, even if ClinicalTrials.gov improves public access to trial information, it does not affect the standard for approval.[167] For example, the evidence uncovered by Kirsch suggests that FDA officials were already aware of the multiple trials being conducted and the unimpressive (if occasionally statistically-significant) evidence of efficacy, but approved the drugs anyway in accordance with the existing legal standard.[168] Figure 3 illustrates the role of ClinicalTrials.gov as a means to address the problem of selective publication, which can undermine the certainty that reported results are in fact accurate and representative of a given drugs characteristics.[169]

---

163. *Publication and Disclosure Issues in Antidepressant Pediatric Clinical Trials: Hearing Before the Subcomm. on Oversight and Investigations of the H. Comm. on Energy & Commerce*, 108th Cong. 18 (2004) (statement of Dr. Janet Woodcock, Deputy Commissioner for Operations, Food and Drug Administration).

164. *See supra* Part IV.A (discussing FDA drug approval after two positive trials, despite other reported failed trials).

165. *See* Freedom of Information Act, Pub L. No. 89-554, 80 Stat. 383 (1966) (codified as amended at 5 U.S.C. § 552 (2012)) (requiring disclosure of information submitted to the FDA, unless exempted under § 552(a)).

166. *See* Kirsch et al., *Emperor's*, *supra* note 124, at 3 (noting that the study was conducted based on information reported to the FDA and obtained under the Freedom of Information Act).

167. *See supra* Part II.A (discussing the legal standard for FDA drug approval).

168. *See* Kirsch, *Antidepressants*, *supra* note 130 (discussing FDA approval of drugs with statistically significant, but clinically unimportant, positive results after two successful trials).

169. *See* Peterson, *supra* note 134, at 5 ("Because positive results are more

Figure 3: Efficacy Evaluation Problems and Existing Regulatory Solutions

Nature of Problem	Problem	Regulatory Solution
Certainty	Selective publication	ClinicalTrials.gov (also, private efforts)

4. Publication Bias and Comparative Efficacy Claims

The multiple-trial problem also manifests itself in the context of comparative efficacy claims. For example, of six clinical studies comparing the efficacy of Pepcid Complete (famotidine, magnesium hydroxide, and calcium carbonate) with calcium carbonate (the active ingredient in Tums), only two showed clear statistical significance, while another two "showed no statistically significant differences" and the final two "showed borderline significance."[170] Although the FDA was aware of at least three equivocal studies that attempted but failed to show superiority of Pepcid Complete, which had formed the basis of an earlier NDA rejection,[171] the two studies showing statistical significance provided the basis for FDA approval of Pepcid Complete as an over-the-counter product.[172] Similarly, of four clinical trials comparing single-enantiomer Nexium (esomeprazole) with racemic Prilosec (omeprazole), only two showed statistical significance.[173] The two statistically significant trials showed only very modest differences in efficacy.[174]

---

likely to be reported, the overall picture may appear rosier than justified by all available evidence.").

170. SmithKline Beecham Consumer Healthcare, L.P. v. Johnson & Johnson-Merck Consumer Pharms. Co., No. 01 Civ. 2775(DAB), 2001 WL 588846, at *3 (S.D.N.Y. June 1, 2001).

171. *See* Michael Elasoff, *Statistical Review and Evaluation*, *in* STATISTICAL REVIEW(S): APPLICATION NO. 20-958 13,13 (U.S. Food & Drug Admin., Ctr. For Drug Evaluation & Research 2000), http://www.accessdata.fda.gov/drugsatfda_docs/nda/2000/20-958_Pepcid%20Complete_statr_P1.pdf (noting the previous drug application denial based on similar clinical trials).

172. *See SmithKline Beecham*, 2001 WL 588846, at *3 ("These two studies were the so-called 'pivotal' studies that led the FDA to approve Pepcid Complete for [over-the-counter] use.").

173. AstraZeneca, *Drug Label for Nexium (esomeprazole magnesium)*, 10–11 (2001), http://www.accessdata.fda.gov/drugsatfda_docs/label/2001/21153lbl.pdf.

174. *Id*. at 10.

One of them, for example, showed that Nexium had a 93.7% healing rate at the end of the eight-week study, less than ten percentage points higher than omeprazole's 84.2% rate.[175] This small difference becomes even less meaningful when one observes that the trial compared 40 mg of Nexium to 20 mg of omeprazole,[176] reflecting the well-established technique of "proving" a new drug's superiority by using a higher dose.

Where drugs are approximately equivalent in efficacy to effective, previously-approved drugs, the concern is less that the new drug is not effective at all, and more that it may not be meaningfully more effective or that the claimed comparative effectiveness may mislead consumers into overpaying.[177] In the case of Pepcid Complete and Tums, for example, the parties were arguing over whether the difference in effectiveness of the products was 7% or 11%[178]—hardly any difference at all—but the competing advertisements at issue in the case each portrayed the respective products as greatly superior to the other, as advertisements tend to do.[179] Both parties agreed that the price difference between the products was 300%.[180]

More important than price is the uncertainty over whether there is any difference in efficacy at all. Despite the rigorous elements of the gold standard that are intended to mitigate bias, ample evidence suggests that drug sponsors can and do influence trial outcomes.[181] The identity of the trial sponsor has been repeatedly shown to correlate with the outcome of the study.[182]

---

175. *Id.*

176. *Id.*

177. *See* SmithKline Beecham Consumer Healthcare, L.P. v. Johnson & Johnson-Merck Consumer Pharms. Co., No. 01 Civ. 2775(DAB), 2001 WL 588846, at *2, *4, *11 (S.D.N.Y. June 1, 2001) (discussing the commercial depiction of (the more expensive) Pepcid Complete as a faster, more effective alternative to Tums, despite studies showing limited differences in effectiveness).

178. *Id.* at *3.

179. *Id.* at *2.

180. *Id.* at *11.

181. *Infra* notes 182, 187 and accompanying text.

182. *See* Stephan Heres et al., *Why Olanzapine Beats Risperidone, Risperidone Beats Quetiapine, and Quetiapine Beats Olanzapine: An Exploratory Analysis of Head-to-Head Comparison Studies of Second-Generation Antipsychotics*, 163 AM. J. PSYCHIATRY 185, 189 (2006) (finding that 90% of pharmaceutical sponsored trials yielded positive results).

One influential paper, for example, found that 90% of head-to-head comparative studies of antipsychotic drugs reported outcomes that favored the trial sponsor's drug.[183] It might be tempting to dismiss this finding as unsurprising, reasoning that drug companies are likely to invest in expensive head-to-head trials only where they have good reason to believe, in advance, that their drug in fact possesses superior efficacy.[184] This cannot be the principal explanation, however, because the researchers found that "different comparisons of the same two antipsychotic drugs led to contradictory overall conclusions depending on the sponsor of the drug."[185]

Inconsistent results can arise from otherwise high-quality trials by making non-neutral comparisons, such as a trial that shows Lipitor (atorvastatin) "superior" to Pravachol (pravastatin)—by using 80mg of Lipitor but only 40mg of Pravachol.[186] Through such techniques, relationships that are logically impossible can be "proven," as reflected in the provocative, M.C. Escher-like title of one research paper: *Why Olanzapine Beats Risperidone, Risperidone Beats Quetiapine, and Quetiapine Beats Olanzapine*.[187] Other studies have consistently confirmed similar sponsor bias in other therapeutic areas.[188]

---

183. *Id.* at 189.

184. *See* Warren Ross, *Head to Head: Prizes, Pitfalls and Pains of Comparative Clinical Testing*, MED. MKTG. & MEDIA, Aug. 2006, at 96 (noting that pharmaceutical companies sponsor head-to-head studies "to prove their drug is superior").

185. Heres et. al., *supra* note 182, at 189.

186. Ross, *supra* note 184, at 92, 98.

187. *See* Heres et al., *supra* note 182, at 189 ("On the basis of these contrasting findings in head-to-head trials, it appears that whatever company sponsors the trial produces the better antipsychotic drug.").

188. *See, e.g.*, Paula A. Rochon et al., *A Study of Manufacturer Supported Trials of Nonsteroidal Anti-Inflammatory Drugs in the Treatment of Arthritis*, 154 ARCH. INTERNAL MED. 157, 161 (1994) ("Our results indicate that the manufacturer-associated drug is always reported as being either superior to or comparable with the comparison drug."); Justin E. Bekelman et al., *Scope and Impact of Financial Conflicts of Interest in Biomedical Research: A Systematic Review*, 289 J. AM. MED. ASS'N 454, 456, 458 tbl.2 (2003) (aggregating in table format eleven previous studies of sponsorship bias, all of which "concluded that industry sponsored research tends to yield pro-industry conclusions").

5. Government-Run Clinical Trials

If industry-sponsored clinical trials are problematic because they are biased, an obvious solution is to instead engage some unbiased third party such as the government to conduct the trials.[189] Alternately, the government might contract out the task to a trusted intermediary such as a medical school, which would report to the government rather than to the drug sponsor.[190] A number of prominent academics have made similar suggestions in the past,[191] and indeed government-run trials might well go a long way toward addressing issues of bias, which include not only publication bias but various types of fraud, misleading presentation, and poor quality.[192] Acting as a representative of public rather than private interests, the government might be better positioned to stop development of new drugs when anticipated efficacy levels fail to materialize rather than continuing to pour money into additional or larger trials in an effort to show the minimal levels of efficacy needed for approval.

Nevertheless, solving a market failure by legislatively abolishing the market represents an extreme step. Not only would such a change be politically challenging and potentially disruptive, but it might also lead to unforeseen consequences. For example, placing the obligation to run clinical trials on the government may cause drug companies to merely shift the focus of their attempted influence from the private entities that currently conduct clinical trials to government entities, making such influence even less transparent than it currently is and further promoting the

189. *See* Heres et al., *supra* note 182, at 190 (suggesting that the FDA evaluate the methodology of studies).

190. *See, e.g.*, Marc A. Rodwin, *Independent Clinical Trials to Test Drugs: The Neglected Reform*, 6 ST. LOUIS U. J. HEALTH L. & POL'Y 113, 117 (2012) ("Researchers would not depend on drug firms to select them and so would lack incentive to favor the drug firm.").

191. *See, e.g.*, *id.* at 113 (exploring the advantages and disadvantages of various reform measures). Rodwin also notes the earlier proposals of Drs. Marcia Angell, Jerry Avorn, and others. *Id.* at 126–27.

192. *See, e.g.*, Michael Hochman & Danny McCormick, *Endpoint Selection and Relative (Versus Absolute) Risk Reporting in Published Medication Trials*, 26 J. GEN. INTERNAL MED. 1246, 1247 tbl.1 (2011) (noting that means to make outcomes appear better than they are include the use of (1) relative rather than absolute measures; (2) surrogate endpoints; (3) composite endpoints; and (4) disease-specific endpoints).

institutional corruption of government. Said differently, the desirability of government-run clinical trials rests on the assumption that the government would be better insulated than private parties from the influence of drug sponsors, an assumption that deserves critical examination.

Fortunately, a softer and politically more feasible approach may be able to facilitate market functioning. One such approach, for example, is to increase the transparency of efficacy information. Markets tend to work better when market participants have access to good information,[193] and it may be a relatively simple matter to cast light on the magnitude of a drug's efficacy so that patients, doctors, and others could evaluate for themselves the drug's value. Helping the public become aware of the magnitude of drug efficacy might be accomplished, for example, by promulgating guidelines for measuring and describing efficacy, and by making simple and clear statements of efficacy easily available, such as by placing them on product labeling or by making them available on a user-friendly and freely accessible website.

The FDA already regulates trial quality and bias through its extensive regulations.[194] Therefore, even if a manufacturer can make its product appear slightly better than placebo (or slightly better than a comparator product), it is unlikely in a regulated environment that publication bias can make a completely ineffective drug appear extremely effective, or vice versa. It is more likely that bias will only nudge the trial outcome slightly in favor of the party in interest, as appears to be happening with many pharmaceutical trials. The ability to nudge trial outcomes would explain why comparative studies tend to favor the trial sponsor,[195] and also why so many new drugs show benefits over placebos that are of only the most modest relevance (if there is any benefit at

193. *See* ROUTLEDGE HANDBOOK OF MODERN ECONOMIC HISTORY 73 (Robert M. Whaples & Randall E. Parker eds., 2013) ("Neoclassical economic models often assume costless information and perfect markets."); James A. Ohlson, *The Social Value of Public Information in Production Economies*, *in* ECONOMIC ANALYSIS OF INFORMATION AND CONTRACTS 95, 107 (John E. Butterworth et al. eds. 1988) ("Any movement away from perfect information . . . leads to . . . a welfare loss . . . .").

194. 21 C.F.R. § 314.126 (2013).

195. *Supra* notes 127–31 and accompanying text.

all).[196]

Casting light on the magnitude of efficacy can appropriately shift the focus of the prescription/consumption decision to the amount of benefit a drug confers rather than the binary question of whether the drug is FDA-approved, mitigating the effects of bias. If simple, minimally intrusive efforts to facilitate more rational consumption decisions prove inadequate, government-run trials could then be considered as a last resort.

B. A "Significant Difference" Is Not Always a Significant Difference

To be clear on the conceptual framework, the "substantial evidence" standard of 21 U.S.C. § 355(d) has been construed in the accompanying FDA regulations[197] to incorporate generally, but flexibly, the elements of randomization, blinding, and placebo control.[198] These elements are widely accepted as defining the gold standard of clinical research.[199] The gold standard, in turn, incorporates a requirement that research results be statistically significant.[200]

---

196. *See, e.g.*, Alex Berenson, *Cancer Drugs Offer Hope, But at a Huge Expense*, N.Y. TIMES (July, 12, 2005), http://www.nytimes.com/2005/07/12/business/12cancer.html?pagewanted=all&_r=0 (last visited on Sept. 17, 2013) ("[T]he new cancer drugs help most patients only marginally, prolonging life by a few weeks or months.") (on file with the Washington and Lee Law Review); P. Bentham et. al., *Long-Term Donepezil Treatment in 565 Patients with Alzheimer's Disease (AD2000): Randomized Double-Blind Trial*, 363 LANCET 2105, 2105 (2004) ("Donepezil is not cost effective, with benefits below minimally relevant thresholds.").

197. 21 C.F.R. § 314.126 (2013).

198. *See In re* Avandia Mktg., Sales Practices & Prods. Liability, No. 2007–MD–1871, 2011 WL 13576, at *2 (E.D. Pa. Jan. 4, 2011) ("Double-blind randomized control trials, and particularly monotherapy trials comparing Avandia use to a placebo, are the 'gold standard' of epidemiology.").

199. *Id.*

200. *See id.* at *4 ("[U]nder a scientific standard one must have statistically significant findings to justify a causal inference."); Tucker v. SmithKline Beecham Corp., 701 F. Supp. 2d 1040, 1062 (S.D. Ind. 2010) (stating that the gold standard utilizes the concept of statistical significance); *see also* Capizzano v. Sec'y of the Dep't of Health and Hum. Servs., Nos. 00–759V, 01–221V, 99–609V, 99–591V, 99–628V, 2003 WL 21432586, at *3 (Fed. Cl. June 20, 2003) ("Respondent's expert Dr. Moulton testified that the gold standard for proof of causation would be a double-blind controlled study with a statistically significant sample."); Jonathan Denne & Gregory Enas, *"Substantial Evidence"*

1. Statistical Significance Is a Measure of Certainty, Not Efficacy

The FDA regulations note that an adequate and well-controlled study should "us[e] a design that permits a valid comparison with a control to provide a *quantitative assessment* of drug effect."[201] Although the term "quantitative assessment" might suggest a quantification of the level of efficacy itself, the context ("a valid comparison with a control") makes clear that what is quantified is not the drug's level of efficacy *per se*, but rather the relationship between the results from the control group and those from the active group. This relationship is embodied in the concept of "statistical significance."[202]

Statistical significance is a measure of the probability that an outcome occurred as a result of a particular chosen variable, rather than as a result of chance.[203] To illustrate, suppose researchers administered a double-blind study in which they randomly divided 1,000 patients with a history of heart attack into two groups. Suppose further that all members of one group received placebo (the "placebo group"), while all members of the other group received the drug under investigation (the "active group"). Finally, suppose that at the end of ten years, all of those in the placebo group suffered at least one additional heart attack, while none of those in the active group suffered another heart attack. Examining these results, the researchers could reasonably conclude that the extremely favorable outcome was related to the presence or absence of the drug, and was not the result of random chance.

In practice, results are rarely so clear-cut. Continuing the previous example, suppose that at the end of ten years 55% of

---

from a Replicated Secondary Analysis, Followed by a Single Prospective Confirmatory Study, 42 DRUG INFO. J. 131, 131 (2008) ("[T]he standard paradigm for providing substantial evidence is usually two well-controlled confirmatory studies in which a statistically significant difference has been shown . . . .").

201. 21 C.F.R. § 314.126(b)(2) (2013) (emphasis added).

202. For background information on statistical significance and hypothesis testing, see DAVID RAY ANDERSON ET AL., STATISTICS FOR BUSINESS & ECONOMICS 338–92 (2008).

203. *See* Ficken v. Clinton, 841 F. Supp. 2d 85, 91 (D.D.C. 2012) ("Statistical significance is a measure of the probability that the outcome of a statistical analysis would have occurred by chance. . . ." (citing Segar v. Smith 738 F.2d 1249, 1268 (D.D.C. 1984))).

those in the "control group" experienced heart attacks, while only 51% of those in the "active group" did so. Under these circumstances, it is more difficult to determine whether the difference of 4% is the result of the drug or merely of random chance. Fortunately, statistical methods can be used to quantify the likelihood that the difference is the result of mere chance. These methods result in the calculation of "*p*-values," which represent the probability that the outcome was the result of chance.[204] For example, a *p*-value of 0.10 would indicate a 10% probability that the observed difference was the result of mere chance, and a 90% probability that the difference was due to the variable under study, such as a new drug. Smaller *p*-values are therefore desirable when testing for drug efficacy, since the smaller the *p*-value, the greater the probability that the observed improvement was due to the drug under study.

Statistical significance is therefore a concept that measures the certainty of an outcome rather than its magnitude. It is not a measurement of efficacy level. In addition, the cutoff for statistical significance is essentially arbitrary:[205] By convention, a *p*-value calculated to be 5% or less (p=0.05) is considered to be "statistically significant,"[206] but other *p*-values could also be used. Nevertheless, p=0.05 is the standard generally used by the FDA (and generally accepted by the scientific community) when evaluating the efficacy attested to by clinical trial data in order to be reasonably certain that a Type I error is unlikely to have occurred.[207] Figure 4 highlights the function of statistical significance as a measure of

---

204. *See* Henry I. Miller & David R. Henderson, *The FDA's Risky Risk-Aversion*, 145 POL'Y REV. 3, 14 (2007), *available at* http://hoover.org/publications/policy-review/article/6147 (discussing what constitutes a *p*-value).

205. *See id.* at 14 (stating that, while normally established at p=.05, statistical significance is arbitrary).

206. Vanderwerf v. SmithKlineBeecham Corp., 529 F. Supp. 2d 1294, 1301 n.9 (D. Kan. 2008) (citing Coates v. Johnson & Johnson, 756 F.2d 524, 537 n.13 (7th Cir. 1985); Federal Judicial Center, *Reference Manual on Scientific Evidence* 124, 357–58 (2d ed. 2000)).

207. *See* Joseph W. Cromier, *Advancing FDA's Regulatory Science Through Weight of Evidence Evaluations*, 28 J. CONTEMP. HEALTH L. & POL'Y 1, 9 (2011) ("FDA generally considers a clinical trial to be a success if the p value is less than or equal to 0.05 when comparing the treatment group to the control group."); Russell Katz, *FDA: Evidentiary Standards for Drug Development and Approval*, 1 NEURORX 307, 311 (2004) ("[T]he typical 'cap' on the type I error rate is set at 5% . . . .").

certainty that study results are real rather than the result of random chance.

Figure 4: Addressing Random Variation: Statistical Significance

Problem	Regulatory Solution
Type I error (i.e., the appearance of efficacy is actually due to random variation)	Statistical Significance

2. When Is a "Significantly Effective" Drug Not Significantly Effective?

The greatest—or, one might say, most significant—problem with the standard of statistical significance is the double meaning that can be ascribed to it. In the context of statistics, saying that a drug is "significantly" more effective than placebo would be understood to mean that the difference in efficacy is statistically significant.[208] For example, a book entitled *A Guide to Treatments that Work* notes that "clomipramine was . . . significantly more effective than citalopram,"[209] that "risperidone was significantly more effective than placebo"[210] and that "[a]ll SSRI's have been reported to be significantly more effective than a placebo."[211] To the statistician, this means only that the difference in efficacy, regardless of how small, was not due to random chance. To the layperson, however, a "significantly" more effective drug may be understood to mean one that is *much more effective*. Statistical texts have warned of this double meaning since at least the 1950s:

> It is essential not to confuse the statistical usage of "significance" with the everyday usage. In everyday usage, "significant" means "of practical importance," or simply

208. *See* PETER E. NATHAN & JACK M GORMAN, A GUIDE TO TREATMENTS THAT WORK 279 (2007) (discussing statistical studies that show that in the context of statistics "significantly more effective" means statistical significance); JOHN O'GRADY ET AL., MEDICINES, MEDICAL DEVICES AND THE LAW 189 (2011) ("If the probability of the difference occurring by chance is 1 in 20, or less, then the result is said to be 'significant.'").

209. NATHAN & GORMAN, *supra* note 208, at 275.

210. *Id.* at 210.

211. *Id.* at 279.

> "important." In statistical usage, "significant" means "signifying a characteristic of the population from which the sample is drawn," regardless of whether the characteristic is important.[212]

This double entendre can be skillfully employed to a drug company's advantage. For example, Pfizer boasts on one consumer-oriented website that "[i]n clinical studies, for patients with RA [rheumatoid arthritis], CELEBREX demonstrated *significant* reduction in joint tenderness/pain and joint swelling."[213] Similarly, Amgen proudly states in the headline to an online press release that Enbrel (etanercept) "[s]ignificantly [r]educed [l]evels of C-[r]eactive [p]rotein."[214] Notwithstanding that it is probably not at all clear to the lay reader why it is beneficial for a patient to achieve lower levels of C-reactive protein, these promotional materials seem intended to convey the message that the drugs are not only beneficial, but beneficial enough to justify a trip to the doctor for a prescription and perhaps also an out-of-pocket co-payment. That the drugs offer "significant" benefits may be true in the statistical sense, but lay readers not trained in statistics may perceive the message quite differently.

Even if a claim of significance is true according to one of its meanings, courts have the power to deem that claim to be in violation of the Lanham Act[215] "by necessary implication if it is

---

212. STEPHEN THOMAS ZILIAK & DIERDRE N. MCCLOSKEY, THE CULT OF STATISTICAL SIGNIFICANCE 110 (2008) (quoting W. ALLEN WALLIS & HARRY ROBERTS, STATISTICS: A NEW APPROACH 385 (1956)); *see also* Michael D. Maltz, *Deviating from the Mean: The Declining Significance of Significance*, 31 J. RES. CRIME & DELINQUENCY 434, 440 (1994) ("Statistical significance does not imply substantive significance, and most researchers know this—but this does not stop them from implying that it does.").

213. *About Celebrex*, PFIZER (2013), http://www.celebrex.com/about-celecrex.aspx (last visited Sept. 17, 2013) (emphasis added) (on file with the Washington and Lee Law Review); *see also Symbicort "Fishing" Commercial (2012)*, http://www.youtube.com/watch?v=oG9MxLwnapE (last visited Sept. 17, 2013) ("[Symbicort] significantly improved my lung function starting within five minutes.") (on file with the Washington and Lee Law Review).

214. Sonia Fiorenza et al., *New Findings Show Enbrel(R) (etanercept) Significantly Reduced Levels of C-Reactive Protein, a Marker of Inflammation, in Patients with Moderate to Severe Plaque Psoriasis*, AMGEN (Feb. 1, 2008), http://www.amgen.com/media/media_pr_detail.jsp?releaseID=1103148 (last visited Oct. 22, 2013) (on file with the Washington and Lee Law Review).

215. *See* Pub. L. No.79-489, § 43, 60 Stat. 427 (1946) (codified as amended at 15 U.S.C. § 1125(a) (2013) (discussing civil actions for false designations of origin, false descriptions, and dilution).

susceptible to more than one interpretation" and the other interpretation is false.[216] In addition, the FDA has general authority to censure companies that disseminate advertisements that are misleading, including those that "[u]s[e] the concept of 'statistical significance' to support a claim that has not been demonstrated to have clinical significance or validity."[217]

However, neither of these avenues of redress has been particularly powerful in policing the misuse of the concept of statistical significance. The Lanham Act, the principal federal trademark statute, is generally invoked by competitors seeking to use the necessary implication doctrine to enjoin another competitor's advertisements, as in the Pepcid Complete case described above.[218] Comparative advertisements, however, seem to be more pronounced in the over-the-counter (OTC) market, perhaps because advertised prescription drugs rely primarily on patent status to ward off competitors, while OTC products focus more on branding.[219] Whatever the reason, there are few reported opinions that invoke the Lanham Act to combat the misuse of claims of statistically significant efficacy differences in the prescription drugs sector.

FDA efforts are also unlikely to be adequate. Due to funding constraints, the FDA like all enforcement agencies must prioritize its efforts.[220] As a result, low priority may be assigned to

---

216. SmithKline Beecham Consumer Healthcare, L.P. v. Johnson & Johnson-Merck Consumer Pharms. Co., 906 F. Supp. 178, 184 (S.D.N.Y. 1995) (citing Cuisinarts, Inc. v. Robot-Coupe Int'l Corp., No. 81 Civ. 731-CSH, 1982 WL 121559, at *1 (S.D.N.Y. June 9, 1982)).

217. 21 C.F.R. § 202.1(e)(7) (2013); *see also id.* § 202.1(e)(6)(vii) (stating that an advertisement is false or misleading if it "[c]ontains favorable data or conclusions from nonclinical studies of a drug, such as in laboratory animals or in vitro, in a way that suggests they have clinical significance when in fact no such clinical significance has been demonstrated"); 21 U.S.C. § 353b (2012) ("Prereview of Television Advertisements"); *cf.* 21 C.F.R. § 201.200 (2013) ("Disclosure of drug efficacy study evaluations [DESI] in labeling and advertising.").

218. *See SmithKline Beecham*, 906 F. Supp. at 184 (S.D.N.Y. 1995) (applying the necessary implication doctrine).

219. *See* Simon P. Andersen & Regis Renault, *Comparative Advertising: Disclosing Horizontal Match Information*, 40 RAND J. OF ECON. 558, 577 (2009) (discussing why over-the-counter drugs such as Tylenol engage in comparative advertising to fend off competition).

220. *See* Bryan A. Liang, *Fade to Black: Importation and Counterfeit Drugs*, 32 AM. J.L. & MED. 279, 300 (2006) ("[T]he FDA is chronically underfunded,

enforcement activities directed against claims of "significance" that, while literally true according to one meaning of the term, are less so under another. Bringing the FDA's modest enforcement resources to bear against the massive advertising campaigns of powerful industries can be, as the FDA itself has complained, "like bringing a butter knife to a gun fight."[221] Given the volume of advertisements disseminated during prime-time television alone, thirty-one warning letters in a single year may indicate the extent of the problem rather than stand as an assurance that all misleading claims are dealt with swiftly.[222]

If prescription drug advertisements are misleading and current avenues of redress are inadequate, one possible solution is to legislatively ban such advertisements altogether. While this may seem an extreme step, almost every country to consider the issue has concluded that prohibiting or severely restricting direct-to-consumer prescription drug advertisements is appropriate. Only the United States and New Zealand (which has a population less than that of greater Atlanta)[223] have concluded otherwise,[224] and in the United States the decision to liberalize advertising came only in 1997.[225] In response,

---

leading to situations in which scarce resources must be stretched and policies prioritized for enforcement. . . .").

221. R.J. Reynolds Tobacco Co. v. Food & Drug Admin., 696 F.3d 1205, 1221 (D.C. Cir. 2012).

222. *See* Jacqueline West, *National Marketing Gone Unintentionally Global: Direct-To-Consumer Advertising of Pharmaceutical Products and the Internet*, 10 J. INT'L BUS. & L. 405, 414 (2012), http://law.hofstra.edu/pdf/academics/journals/jibl/jibl_volxii_national_marketing_gone_unintentionally_global_west.pdf (stating that thirty-one warning letters were sent in 2011 by the Office of Prescription Drug Promotion).

223. *See* U.S. CENSUS BUREAU, STATISTICAL ABSTRACT OF THE UNITED STATES: 2012 (2012), http://www.census.gov/compendia/statab/2012/tables/12s0020.pdf (indicating an Atlanta area population of 5,269,000); *Population Clock*, STATISTICS NEW ZEALAND, http://www.stats.govt.nz/tools_and_services/population_clock.aspx (last visited Sept. 17, 2013) (indicating an estimated New Zealand resident population of 4,458,047) (on file with the Washington and Lee Law Review).

224. *See* Marjorie Delbaere, *Metaphors and Myths in Pharmaceutical Advertising*, 82 SOC. SCI. & MED. 21, 21 (2013) (stating that, of the countries that have addressed the issue, only the United States and New Zealand permit direct to consumer advertisements for pharmaceuticals).

225. *See* Draft Guidance for Industry: Consumer-Directed Broadcast Advertisements, 62 Fed. Reg. 43,171–72 (Aug. 12, 1997) (discussing the FDA's requirements for consumer-directed broadcasting); Lars Noah, *Advertising*

state[226] and federal[227] bills have emerged to limit such advertising, and academics have argued the merits of ad prohibition.[228]

Given the serious public health issues at stake combined with the fact that very few advertised drugs offer significant therapeutic advantages over other drugs that may be much cheaper but unadvertised,[229] a prohibition on drug advertising might well be preferable to the status quo. The suppression of misleadingly optimistic presentations of minimally advantageous new drugs might among other things prevent the crowding out of more balanced, healthy and realistic views of drug efficacy.[230] It would

---

Prescription Drugs to Consumers: Assessing the Regulatory and Liability Issues, 32 GA. L. REV. 141, 141 (1997) (noting the dramatic marketing shift toward direct-to-consumer prescription drug advertising during the fifteen years prior to 1997).

226. *See, e.g.*, H.B. 2061, 188th Gen. Ct. (Mass. 2013) ("An Act Prohibiting Advertising by Pharmaceutical Companies"); H.B. 2646, 188th Gen. Ct. (Mass. 2013) ("An Act to Eliminate the Tax Deduction for Direct to Consumer Pharmaceutical Marketing"); H.C.R. 66, 2003 Leg., Reg. Sess. (Ky. 2003), *available at* http://www.lrc.state.ky.us/record/ 03rs/HC66/bill.doc (proposing a resolution "to limit, ban, or otherwise impose strict standards on direct-to-consumer advertising of drugs by pharmaceutical companies").

227. *See, e.g.*, H.R. 722, 112th Cong. (2011) (proposing the "Say No to Drug Ads Act," which would "deny any [tax] deduction for direct-to-consumer advertisements of prescription drugs"); H.R. 2966, 111th Cong. (2009) ("Say No to Drug Ads Act"); H.R. 5105, 107th Cong. (2002) ("Say No to Drug Ads Act").

228. *See* Kurt C. Stange, *Time to Ban Direct to Consumer Prescription Drug Marketing*, 5 ANNALS FAM. MED. 101, 102 (2007) (arguing that direct-to-consumer advertisements should be banned for prescription drugs); Joel Lexchin & Barbara Mintzes, *Direct-to-Consumer Advertising of Prescription Drugs: The Evidence Says No*, 21(2) J. PUB. POL'Y & MARKETING 194, 196–97 (2002) (providing evidence that direct-to-consumer advertising should not be used for prescription drugs).

229. *See* Lexchin & Mintzes, *supra* note 228, at 194 ("There is no evidence that direct-to-consumer advertising results in any improvement in health outcomes.").

230. *See* George Loewenstein, *Out of Control: Visceral Influences on Behaviour*, 65 ORG. BEHAV. & HUM. DECISION PROCESSES 272, 272 (1996) ("[V]isceral factors have a disproportionate effect on behavior and tend to 'crowd out' virtually all goals . . . ."); *cf.* Catherine MacKinnon, *Pornography, Civil Rights, and Speech*, 20 HARV. C.R.-C.L. L. REV. 1, 18 (1985) (asserting that pornography "is not imagery in some relation to a reality . . . [but] is a sexual reality"). Analogously, television advertisements can create a false imagery of drug efficacy that becomes a perceived reality, leading patients to demand nearly worthless drugs no matter the cost or potential side effects.

also help to rein in wasteful expenditures that needlessly inflate healthcare costs.

Nevertheless, preventing businesses from communicating with their customers is a drastic step. From the business perspective, it can raise First Amendment concerns,[231] while from the consumer's perspective, it can block off a potential channel of useful information[232] (even if little useful information is currently flowing through that channel). Most importantly, a less restrictive means of reforming direct-to-consumer advertising is available, namely, increasing the utility of the information in advertisements by requiring a clear presentation of efficacy data.[233] Such a tempered approach would preserve channels of communication, increase transparency, leave the decision in the hands of consumers (and their doctors), and embody free market ideals that have been the traditional underpinning of the United States economic system.[234] Should the reform prove insufficient within a reasonable period of time, prohibition could always be instituted as a last resort.

---

231. *See* Mark I. Schwartz, *To Ban or Not to Ban—That Is the Question: The Constitutionality of a Moratorium on Consumer Drug Advertising*, 63 FOOD & DRUG L.J. 1, 3 n.5 (2008) (noting that legislation proposed in 2007 that would have allowed the FDA to impose a moratorium on advertising was abandoned following claims that it would violate the First Amendment); *see also* Thompson v. W. States Med. Ctr., 535 U.S. 357, 374 (2002) ("We have previously rejected the notion that the Government has an interest in preventing the dissemination of truthful commercial information in order to prevent members of the public from making bad decisions with the information."); Gerald Masoudi & Christopher Pruitt, *The Food and Drug Administration v. The First Amendment: A Survey of Recent FDA Enforcement*, 21 HEALTH MATRIX 111, 112 (2011) (noting that the FDA's "curtailment of constitutionally protected commercial speech" can "remov[e] truthful (and useful) product communications from the marketplace").

232. *See generally* Anthony D. Cox & Dena Cox, *A Defense of Direct-to-Consumer Prescription Drug Advertising*, 53 BUS. HORIZONS 221 (2010) (acknowledging problems, but concluding that advertising can increase consumer knowledge and awareness). Cox and Cox also argue that direct-to-consumer advertising, despite its problems, is at least better than physician targeted promotion, which should be a greater source of public concern. *Id.* at 227.

233. Other proposals that fall short of a full ban have also been suggested. *See, e.g.*, Margaret Gilhooley, *Commercial Speech, Drugs, Promotion and a Tailored Advertisement Moratorium*, 21 HEALTH MATRIX 97, 98–99 (2011) (discussing a prohibition on advertisements for only recently approved drugs, or alternately, for only the most high-risk recently approved drugs).

234. *See* City of Lafayette v. La. Power & Light Co., 435 U.S. 389, 416 (1978) (acknowledging "the Nation's free-market goals").

3. Statistical Significance Is a De Minimis Requirement that Can Be Met by Diet, Exercise, and Other Mundane, Inexpensive Treatments

The low bar imposed by the statistical significance requirement might well allow for a number of mundane, inexpensive treatments to receive FDA approval, so long as they could qualify under the statutory definitions of a "drug" (or "device").[235] For example, clinical trials have demonstrated that the drug Aricept (donepezil) is statistically significantly more effective than a placebo in treating Alzheimer's disease, and as a result the drug was approved by the FDA.[236] But fruit juice, exercise, music, and even coffee might also be able to meet the lax significance standard. One study, for example, followed almost 2,000 people for seven years and concluded that fruit and vegetable juices were "highly significant" in delaying the onset of Alzheimer's,[237] suggesting what may be in any event a sensible dietary change to improve health. A meta-analysis of multiple studies concluded that exercise has a "robust and beneficial influence on the cognition of sedentary older adults,"[238] while another meta-analysis concluded that music therapy was an effective treatment for Alzheimer's.[239] Research

---

235. *See* 21 U.S.C. §§ 321(g)(1), (h), (p) (2012) (defining the terms "drug," "new drug," and "device").

236. *See* S.L. Rogers et al., *A 24-Week, Double-Blind, Placebo-Controlled Trial of Donepezil in Patients with Alzheimer's Disease*, 50 NEUROLOGY 137, 137 (1998), http://www.stvincents.ie/dynamic/File/Donepezil%20study.pdf (describing how donepezil is more effective than a placebo in treating Alzheimer's disease).

237. *See* Qi Dai et al., *Fruit and Vegetable Juice and Alzheimer's Disease: The Kame Project*, 199 AM. J. MED. 751, 751 (2006) (discussing how fruit and vegetable juices play an important role in delaying the onset of Alzheimer's disease, particularly among those who are at high risk for the disease).

238. Stanley Colcombe & Arthur F. Kramer, *Fitness Effects on the Cognitive Function of Older Adults: a Meta-Analytic Study*, 14 PSYCH. SCI. 125, 128 (2003); *see also* Patricia Heyn et al., *The Effects of Exercise Training on Elderly Persons with Cognitive Impairment and Dementia: A Meta-Analysis*, 85 ARCHIVES PHYSICAL MED. & REHABIL. 1694, 1694 (2006) ("Exercise training increases fitness, physical function, cognitive function, and positive behavior in people with dementia and related cognitive impairments.").

239. *See* Susan M. Kroger et al., *Is Music Therapy an Effective Intervention for Dementia? A Meta-Analytic Review of the Literature*, 36 J. MUSIC THERAPY 2, 2 (1999) (finding the effect of music therapy to be "highly significant" for individuals with dementias).

has also shown that caffeine is effective in treating Alzheimer's Disease.[240]

Although diet, exercise, music, or caffeine may not necessarily be substantially effective in treating Alzheimer's, the likelihood that these ordinary, inexpensive and easily available treatments could meet the statistical significance standard casts light on just how de minimis that standard is. It is perhaps no surprise that studies confirm the efficacy of diet and exercise in improving cognition.[241] Nor is it surprising that caffeine, a stimulant, stimulates brain activity in some manner, or that music, which is self-evidently associated with emotion, can favorably affect the brain.[242] These studies merely corroborate the conventional wisdom and lend a measure of scientific credibility to what the public thought it already knew. Few people, however, place their hopes for relief from Alzheimer's in walking, listening to music, drinking coffee, or eating vegetables, and fewer still would be willing to pay $100 for a glass of vegetable juice or a 10-minute walk. Yet desperate patients will pay this much and more for an FDA-approved pill that may do as little, or less, than any of these ordinary treatments.

The vulnerability of the statistical significance standard, therefore, is that it utterly fails to differentiate between factors that have a statistically significant effect in treating diseases or conditions, and those that are substantially effective in treating them. To offer yet another example, one randomized controlled trial concluded that "light therapy and fluoxetine [Prozac] are comparably effective treatments for patients with [seasonal affective disorder]," a type of depression.[243] If this approximate

240. Gary W. Arendash & Chuanhai Cao, *Caffeine and Coffee as Therapeutic Agents Against Alzheimer's Disease*, 20 J. ALZHEIMER'S DISEASE S117 (2010).

241. *See* Qi Dai et al., *supra* note 237, at 751 (finding fruits and vegetables to delay Alzheimer's); Colcombe & Kramer, *supra* note 238, at 128 (finding exercise to improve cognitive function of elderly adults).

242. *See* Kroger et al., *supra* note 239, at 2 (finding caffeine to be an effective treatment against Alzheimer's).

243. Raymond W. Lam et al., *The Can-SAD Study: A Randomized Controlled Trial of the Effectiveness of Light Therapy and Fluoxetine in Patients with Winter Seasonal Affective Disorder*, 163 AM. J. PSYCHIATRY 805, 811 (2006). The Drug Effectiveness Review Project cited this study with approval. GERALD GARTLEHNER ET AL., DRUG CLASS REVIEW: SECOND-GENERATION ANTIDEPRESSANTS: FINAL UPDATE 5 REPORT 47 (2011), http://www.healthandwelfare.idaho.gov/Portals/0/Medical/MedicaidCHIP/AntidepressantsFin

equivalence were understood either by the medical community or the marketplace, it would be difficult to explain the $21 billion spent on Prozac during its first thirteen years on the market.[244] One could instead simply buy very bright lights, move one's workspace closer to a window, step outside during daylight hours, etc.

V. "Clinical Significance" Does Not Imply Greater Efficacy

Proponents of the current efficacy standard sometimes argue that drugs cannot be approved unless the level of efficacy is "clinically significant," apparently suggesting that "clinical significance" requires an elevated degree of efficacy.[245] There is no statutory or regulatory basis for such a distinction, nor can such a distinction be found in FDA guidance documents or court decisions. If clinical significance has any meaning distinct from statistical significance, it is that clinical significance means statistical significance in humans (as opposed to in animals or in vitro).

A. The Law Requires Clinical Significance

FDA regulations do include a clinical significance requirement in a number of provisions.[246] For example, "effectiveness" in the context of over-the-counter products is defined as "a reasonable expectation that, in a significant proportion of the target population, the pharmacological effect of the drug, when used under adequate directions for use and warnings against unsafe

---

alReport.pdf.

244. Bethany McLean, *A Bitter Pill*, FORTUNE, Aug. 13, 2001, at 118.

245. *See, e.g.*, DAVID MACHIN, YIN BUN CHEUNG & MAHESH K.B. PARMAR, SURVIVAL ANALYSIS: A PRACTICAL APPROACH 1, 17 (2d. 1995), *available at* http://onlinelibrary.wiley.com.ezp-prod1.hul.harvard.edu/doi/10.1002/0470034572.ch1/pdf (arguing that clinical significance, as opposed to statistical significance, implies a difference in efficacy between two treatments that is "substantial"); MICHAL J. CAMPBELL, DAVID MACHIN & STEPHEN J. WALTERS, MEDICAL STATISTICS: A TEXTBOOK FOR THE HEALTH SCIENCES 1, 288 (4d. 2010), http://tinyurl.com/lal6sb3 (arguing that results can be statistically but not clinically significant).

246. *Infra* notes 247–49 and accompanying text.

use, will provide *clinically significant* relief of the type claimed."[247] Parallel provisions require that biologics "serve a *clinically significant* function in the diagnosis, cure, mitigation, treatment, or prevention of disease in man,"[248] and that medical devices "provide *clinically significant* results."[249] With respect to the efficacy needed for new prescription drug approval, the term "clinically significant" appears nowhere in the statute[250] or regulations,[251] but the statute does require the undertaking of "clinical trials" that "form the primary basis of an effectiveness claim."[252] Thus, the clinical significance requirement is incorporated into the new drug statute as well as applying in other areas of FDA regulation.

B. Clinical Significance Means Statistical Significance in Humans

Although the law requires clinical significance in a variety of contexts,[253] nowhere in the regulations or statute is it stated that a showing of clinical significance for new drugs necessitates an elevated degree of efficacy vis-à-vis statistical significance. Instead the most plausible reading of the law is that clinical significance

247. 21 C.F.R. § 330.10(a)(4)(ii) (2013) (emphasis added).

248. *Id.* § 601.25(d)(2) (emphasis added).

249. *Id.* § 860.7(e)(1) (emphasis added).

250. *See* 21 U.S.C. § 355 (d)(7) (2012) (defining substantial evidence to include "evidence consisting of adequate and well-controlled investigations, including clinical investigations").

251. *See, e.g.*, 21 C.F.R. § 201.57(a)(7) (2013) (noting that "clinically significant clinical pharmacologic information" must appear in the "[h]ighlights of prescribing information" section of drug labeling); *id.* § 201.57(c)(3)(i)(J) (requiring drug labeling to indicate "[e]fficacious . . . concentration ranges . . . if established and clinically significant"); *see also* Karen M. Becker et al., *Scientific Dispute Resolution: First Use of Provision 404 of the Food and Drug Administration Modernization Act of 1997*, 58 FOOD & DRUG L.J. 211, 220 (2003) (noting that in one case where clinical significance was specifically at issue, the FDA "shifted the scientific dispute from a complex specific question focused on clinical significance . . . to any scientific issue that . . . provided a basis for its not-approvable decision"); *cf.* 21 C.F.R. § 860.7(e)(1) (requiring reasonable assurance that medical devices provide clinically significant results).

252. 21 U.S.C. § 355(b)(5)(B) (2012).

253. *Supra* Part V.A.

merely requires statistical significance *in human trials*, as opposed to animal trials or in vitro studies.[254]

This distinction is made explicitly in section 202.1 of the FDA drug advertising regulations, which provide that "as used in this section, 'clinical investigations,' 'clinical experience' and 'clinical significance' mean in the case of drugs intended for administration to man, investigations, experience, or significance in humans."[255] Elsewhere in section 202.1, the regulation clarifies what does not constitute clinical significance, stating that an advertisement is false or misleading if it "contains favorable data or conclusions from nonclinical studies of a drug, *such as in laboratory animals or in vitro*, in a way that suggests they have clinical significance."[256] Another provision in the same section states that an advertisement may be false or misleading if it "[u]ses the concept of 'statistical significance' to support a claim that has not been demonstrated to have clinical significance or validity," i.e., if it has not been demonstrated to have statistical significance in humans.[257] Figure 5 reflects the function of the clinical significance standard.

Figure 5: Ensuring Relevance in Humans: Clinical Significance

Nature of Problem	Problem	Regulatory Solution
Relevance	Efficacy data may not be relevant to humans	Clinical significance

By the terms of the FDA drug advertising regulation, this definition of clinical significance as meaning statistical significance in humans applies only to section 202.1.[258] Nevertheless other documents such as FDA guidance and government reports are not inconsistent with this definition.[259] A Government Accountability Office report, for example, notes that although there is no definition of "clinical significance" in the FDA medical device regulations, in the context of medical devices, the term is

254. *Infra* text accompanying notes 256–64.
255. 21 C.F.R. § 202.1(e)(4)(ii)(b) (2013).
256. *Id.* § 202.1(e)(6)(vii) (emphasis added).
257. *Id.* § 202.1(e)(2)(7)(ii).
258. *Id.* § 202.1(e)(4)(ii)(b).
259. *Infra* notes 261–64 and accompanying text.

understood to mean results that "have a positive effect on the disease being treated according to the standard of care for the related field."[260] An FDA guidance document explains how to "provid[e] clinical evidence of effectiveness for human drug and biological products,"[261] which at no point indicates that clinical significance implies an elevated level of efficacy. Section 2 of the guidance document, for example, discusses quantity of evidence, while Section 3 discusses quality of evidence.[262] Both sections explain at length the flexible nature of the evidence standard, noting that in some situations "effectiveness of a new use may be extrapolated entirely from existing efficacy studies" without the need to conduct an additional study,[263] and that under other circumstances "it is possible for sponsors to rely on [certain] studies to support effectiveness claims, despite less than usual documentation or monitoring."[264]

C. Cases Addressing Absolute Efficacy Fail to Distinguish Clinical and Statistical Significance

The few cases to address both clinical and statistical significance fail to distinguish the terms on the basis of efficacy level. In general, they do not clearly distinguish the two terms at all, sometimes implying that there is a distinction but then declining to reach a decision on the basis of any distinction and often failing to clearly articulate or define any distinction.[265] In

260. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-07-996, MEDICAL DEVICES: FDA'S APPROVAL OF FOUR TEMPOROMANDIBULAR JOINT IMPLANTS 8 n.9 (2007), http://www.gao.gov/new.items/d07996.pdf.

261. FDA GUIDANCE, *supra* note 55, at 1 n.1.

262. *See id.* at 6–16 (providing guidance on the quantity of evidence needed in particular circumstances to establish substantial evidence of effectiveness); *id.* at 16–20 (discussing the factors that influence the quality of documentation evidence needed to support approval of a new human drug and biological product).

263. *Id.* at 6.

264. *Id.* at 17.

265. *See infra* notes 267–82 and accompanying text (discussing a case that addresses but fails to define a distinction between clinical and statistical significance).

other cases, clinical significance has been treated as if it is essentially synonymous with statistical significance.[266]

One of the cases to discuss both clinical and statistical significance is the 1986 Third Circuit case of *Warner-Lambert Co. v. Heckler*,[267] which involved the FDA's withdrawal of approval of several oral proteolytic enzymes that had long been promoted as effective in relieving inflammation and pain, especially that arising from surgery, trauma, infection and allergic reactions.[268] The drugs had initially received FDA approval prior to the Drug Amendments of 1962,[269] at a time when approval formally required only a showing of safety, but not efficacy.[270] The 1962 amendments required proof of efficacy not only for any new drugs submitted for approval after the effective date of the amendments, but also required the FDA to go back and review the efficacy of drugs that had already been approved prior to 1962, including the oral proteolytics at issue in *Warner-Lambert*.[271] Following extensive review of the data submitted by Warner-Lambert and the other manufacturers in the case, an FDA administrative law judge (ALJ) concluded that the manufacturers had failed to establish that the drugs were effective.[272] The FDA Commissioner upheld this finding.[273] More than twenty years after the 1962 amendments,

---

266. *See infra* notes 282–87 and accompanying text (discussing two cases that did not distinguish between or base their holdings on a distinction between clinical and statistical significance).

267. 787 F.2d 147 (3d Cir. 1986).

268. *See id.* at 149 ("The Commissioner withdrew approval of the new drug applications for these drugs after concluding that there was a lack of substantial evidence that the OPEs will have effects they are purported or represented to have for their intended conditions of use.").

269. Pub. L. No. 87-781, 76 Stat. 780 (1962) (codified in scattered sections of 21 U.S.C.).

270. *See Warner-Lambert*, 787 F.2d at 149 ("At the time approval was granted, the Food, Drug, and Cosmetic Act required the FDA to determine only that a drug was safe for human use.").

271. *See id.* ("The 1962 amendments also required the FDA to reevaluate drugs that it had previously approved.").

272. *See id.* at 150 ("The ALJ thus found that the drug manufacturers had not met their statutory burden of producing evidence demonstrating that the OPEs were effective.").

273. *See id.* ("The Commissioner also found that there was a lack of substantial evidence that the . . . OPEs have the effects represented, and, accordingly, withdrew approval.").

the case finally reached the Third Circuit, which upheld the Commissioner's findings.[274]

On the surface, certain statements in *Warner-Lambert* appear to support the proposition that statistical significance is distinctly different from clinical significance. According to the Third Circuit, "[t]he Commissioner's interpretation of the statute as requiring a showing of clinical significance, rather than merely statistical significance, is persuasive."[275] However, the rejection by the FDA (and court) for lack of efficacy seems in fact to have been based primarily on the faulty methodology of the studies under consideration rather than on any distinction between clinical and statistical significance.[276] For example, one study had made 240 comparisons between the placebo and study groups, finding six of those comparisons to be statistically significant.[277] However, as discussed above, at the level of statistical certainty usually required for drug approval (p=.05), one in twenty studies can be expected to reflect a Type I error, erroneously indicating efficacy where there is none.[278] With 240 comparisons, this would suggest (assuming independence) that perhaps twelve comparisons might erroneously show statistical significance where none exists, which is in the neighborhood of what was in fact observed.[279] More importantly, the FDA Commissioner had found that the post-hoc "stratification of the subjects into subgroups . . . had no scientific basis."[280] What this means in lay terms is that it appeared to the

---

274. *See id.* at 159–60 (finding that the Commissioner had a reasonable basis for disqualifying each submission of Warner-Lambert's studies attempting to establish that the drugs were effective and noting that the rejection of the studies is consistent with the possibility that the OPEs do not work).

275. Warner-Lambert Co. v. Heckler, 787 F.2d 147, 155 (3d Cir. 1986).

276. *See id.* at 160 (discussing the Commissioner's rejection of faulty studies).

277. *See id.* at 155 (discussing the results of the study on the therapeutic effect of OPE Chymoral).

278. *See* 44 Fed. Reg. 51512, 51520 col. 3 (Aug. 31, 1979) ("As a matter of scientific custom, a statistically significant difference has sometimes been considered one that is likely to occur by chance 1 in 20 times or less. . . .").

279. *See Warner-Lambert*, 787 F.2d at 155 ("Of the 240 tests, only six provided statistical results indicating that Chymoral had some effectiveness . . . .").

280. *Id.* at 155. The FDA has repeatedly urged caution when statistical significance is found after multiple comparisons are made, owing to the elevated risk of Type I errors. *See, e.g.*, William B. Hood, *More on Sulfinpyrazone After*

FDA that the manufacturers had inappropriately mined the data, after the fact, in such a way as to create the appearance of statistically significant results where none existed.[281]

In another Third Circuit case from the 1980s, *United States v. 225 Cartons . . . of an Article or Drug*,[282] the court noted with approval the FDA's requirement that new drugs demonstrate "a clinically, *i.e.*, therapeutically, significant benefit . . . ."[283] Once again, however, the drug product in question was not rejected on the basis of any distinction between clinical and statistical significance, but on the basis that the studies submitted by the manufacturer were inadequate to show that each of the putative active ingredients contributed to the drug's efficacy.[284] Nowhere did the court hold that statistically significant, but not clinically significant, results had been established.

Similarly, in *American Home Products Corp. v. Johnson & Johnson, Inc.*,[285] the Second Circuit found "no reliable evidence showing that [the aspirin in Anacin®] reduces inflammation to a clinically significant extent in the conditions listed in the

---

Myocardial Infarction, 306(16) NEW ENG. J. MED. 988, 988–89 (1982) (discussing one study with inconsistent findings in which the FDA had concerns about data misclassification and exclusion).

281. *See* Warner-Lambert Co. v. Heckler, 787 F.2d 147, 155 (3d Cir. 1986) (showing that the six statistically significant results demonstrating effectiveness for reduction of pain were spread out over several different subgroups suffering various ailments and no comparison of the total drug group to the total placebo group was ever made). Repeated testing for statistical significance at frequent intervals during the trial process can also bias results and lead to Type I errors. *See* S.J. Pocock, *Size of Cancer Clinical Trials and Stopping Rules*, 38(6) BRIT. J. CANCER 757, 761 (1978) ("The more often one performs a significance test on the accumulating results in a trial, the greater is the chance that some significant difference will eventually be detected, even if the treatments are really equally effective."); *see also* J.L. Haybittle, *Repeated Assessment of Results in Clinical Trials of Cancer Treatment*, 44 BRIT. J. RADIOLOGY 793, 796 (1971) (urging similar caution).

282. 871 F.2d 409 (3d Cir. 1989).

283. *Id.* at 416.

284. *See id.* at 415 (following the district court's ruling that "Sandoz had failed to produce such studies for its FWC combination products, and thus the court rejected its claim of general recognition" (citing United States v. 225 Cartons . . . of an Article or Drug, 687 F.Supp. 946, 962 (D.N.J. 1988)). Under the applicable FDA regulation, it must be shown that "each component [of a combination drug product] makes a contribution to the claimed effects." *Id.* (quoting 21 C.F.R. § 300.50(a) (1988)).

285. 436 F. Supp. 785 (S.D.N.Y. 1977), *aff'd*, 577 F.2d 160 (2d Cir. 1978).

advertisements at OTC [over-the-counter] dosages." As in *225 Cartons*, the Second Circuit based its holding not on any distinction between clinical and statistical significance, but on the basis of inadequate studies. The court noted that some of the documents submitted in support of efficacy consisted of "informal pieces" that were "not studies or tests at all."[286] Other documentation provided some evidence of statistically significant efficacy with respect to rheumatoid arthritis, but rheumatoid arthritis was not among the conditions for which the manufacturer was now claiming efficacy.[287]

A 1979 administrative decision by then-FDA Commissioner Donald Kennedy also appears at first to endorse a meaningful distinction between clinical and statistical significance.[288] In a Final Decision following a formal evidentiary public hearing in an adjudicative proceeding for the cough suppressant Benylin (diphenhydramine), Kennedy emphatically rejected "the fallacy of equating statistical significance with clinical significance."[289] He dismissed a 9% reduction in coughing, stating that it "may be statistically significant . . . but is not clinically significant."[290] In the end, however, the decision to decline approval for the antitussant indication of Benylin (diphenhydramine) seems to have been based on a finding that neither of the two studies qualified as an "adequate and well controlled investigation[]," and that therefore there was a "lack of 'substantial evidence' . . . that Benelyn will have the effect it purports . . . to have."[291] The Commissioner found troubling (1) the fact that statistically significant results were obtained only on the first day of the study;[292] (2) that those results were not strongly statistically

---

286. *Id.* at 800–01.

287. *See id.* at 799 (discussing the general acceptance that Anacin has an anti-inflammatory effect in the treatment of rheumatic diseases at dosages exceeding that recommended for over-the-counter use).

288. *Infra* notes 289–90 and accompanying text.

289. Benylin Final Decision, 44 Fed. Reg. 51,512, 51,521 col. 1 (Aug. 31, 1979).

290. *Id.* at 51,521 col. 2.

291. *Id.* at 51,537 col. 1.

292. *See id.* at 51,521 col. 3 ("It should also be noted that the results of the . . . study were statistically significant only on the first day of the study. . . .").

significant;[293] (3) that overall more patients were satisfied with the placebo than with Benelyn (diphenhydramine);[294] (4) that overall the physician-investigators did not rate Benelyn (diphenhydramine) differently from placebo with any statistical significance;[295] (5) that the placebo controls were inadequate because they contained ingredients that may have had pharmacological activity;[296] and (6) that the statistically significant results that were found were based on subjective responses (given by children aged six to twelve years) which the Commissioner found unreliable.[297] In short, a number of factors other than any difference between statistical and clinical significance led to the Commissioner's finding. In any event, the skeptical views of statistical significance expressed by a single FDA Commissioner who served in that role for only twenty-six months during the 1970s[298] do not seem to have had a lasting impact on subsequent judicial decisions interpreting the substantial evidence standard, as seen in the cases just discussed.[299]

The lack of any meaningful and enduring distinction between clinical and statistical significance with respect to drug efficacy has not gone entirely unnoticed.[300] One scholar, after exhaustively

---

293. *See id.* ("[T]he results even on [the first day] were just on the borderline of statistical significance. A change in the reports of just a few patients would have eliminated this significance.").

294. *See id.* ("More patients were satisfied with the [placebo] (90.3 percent) than with Benylin (84.3 percent).").

295. *See id.* ("The other such measure was a question directed to investigators, which called for an overall rating as to beneficial drug-attributable results from medication. The results did not reveal any statistically significant differences between Benylin and the [placebo]. . . .").

296. *See id.* at 51,527 col. 3 ("I find that ammonium chloride and sodium citrate in the amounts used in the [placebo] may have expectorant, demulcent, or other pharmacological activity.").

297. *See id.* at 51,529–30 (describing the inability of the patients to provide "valid subjective evaluations").

298. U.S. Food & Drug Admin., *About FDA: Donald Kennedy, Ph.D.* (Feb. 20, 2009), http://www.fda.gov/AboutFDA/WhatWeDo/History/Leaders/Commissioners/ucm093736.htm (last visited Sept. 18, 2013) (on file with the Washington and Lee Law Review).

299. *See supra* notes 267–87 and accompanying text (discussing judicial decisions interpreting the "substantial evidence" standard).

300. *See infra* note 301 and accompanying text (discussing an article that compares clinical and statistical significance).

analyzing the meaning of clinical significance in the context of FDA approval, patent infringement, false advertising, and product liability cases, concluded:

> Having presented a survey of the various indications from case law, legislative materials, and academic sources, it is apparent how little clarity, and certainly how little consensus, there is concerning the meaning and appropriate usage of the phrase "clinical significance." . . . In fact, it appears that these requirements [for clinical significance] are so much like those for finding statistical significance as to be simply redundant. . . . As the phrase stands now, it makes little if any positive contribution to drug-related litigation, while at the same time causing several important problems that will only grow worse . . . .[301]

Oddly, the distinction between clinical significance and statistical significance contained in the advertising regulations discussed above has been largely overlooked by the courts, commissioners and commentators that have discussed the subject.[302]

D. Cases Addressing Comparative Efficacy Fail to Distinguish Clinical and Statistical Significance

Even in the context of the FDA drug advertising regulation,[303] the meaning of clinical significance has sometimes been seemingly misunderstood. Section 202.1 defines as possibly false or misleading those advertisements that "[u]se[] the concept of 'statistical significance' to support a claim that has not been demonstrated to have clinical significance or validity."[304] Only two reported cases quote or cite this provision, and only one of these engages with it substantively.[305] In that case, AstraZeneca had

---

301. Sarah M.R. Cravens, *The Usage and Meaning of "Clinical Significance" in Drug-Related Litigation*, 59 WASH. & LEE L. REV. 553, 594–96 (2002).

302. Similar language defining clinical significance has appeared in the drug advertising regulations since at least 1969. 21 C.F.R. § 1.105(e)(4)(iii)(b) (1969).

303. 21 C.F.R. § 202.1 (2013).

304. 21 C.F.R. § 202.1(e)(7)(ii).

305. *See* AstraZeneca LP v. TAP Pharm. Prods., Inc., 444 F. Supp. 2d 278, 282–89 (D. Del. 2006) (engaging in substantive discussion on 21 C.F.R. §202.1(e)(7)); Pa. Emps. Benefit Trust Fund v. Zeneca Inc., 499 F.3d 239, 248 (3d Cir. 2007) (mentioning 21 C.F.R. § 202.1(e)(7)), *vacated*, 129 U.S. 1578 (2009).

promoted Nexium (esomeprazole) in a "Better is Better" campaign that claimed, among other things, that "recent medical studies . . . prove Nexium heals moderate to severe acid related damage in the esophagus better than the other leading prescription medicine."[306] One of the human studies put forth by AstraZeneca reported an overall healing rate of 92.6% for Nexium (esomeprazole) versus 88.8% for Prevacid (lansoprazole), an unimpressive difference that was nevertheless statistically significant.[307] TAP Pharmaceuticals, the maker of Prevacid (lansoprazole), asserted that AstraZeneca had engaged in false advertising in violation of the Lanham Act,[308] and cited in support the FDA drug advertising regulations that prohibit the misleading use of statistical significance to mean clinical significance.[309] TAP argued that the advertisement was "literally false because Nexium is only marginally better at healing EE [erosive esophagitis], and that difference is clinically meaningless."[310] Elsewhere, TAP asserted that "the Castell and Fennerty studies [on humans], as well as other studies, while statistically significant, are not clinically significant."[311]

Although the court declined to find a violation of the FDA drug advertising regulations, it did so without any reference to the distinction between animal and in vitro studies on the one hand, and clinical studies on the other.[312] Indeed, the court seemed to tacitly accept the suggestion that clinical significance meant meaningful significance, finding that "the Castell and Fennerty studies are relevant to a consumer's use of Nexium" because "Nexium is at least statistically significantly better at healing

---

306. *AstraZeneca LP*, 444 F. Supp. 2d at 282.

307. *Id.* at 282–83.

308. Trademark Act of 1946, 15 U.S.C. §§ 1051–1072.

309. *See AstraZeneca LP*, 444 F. Supp. 2d at 295 (discussing TAP's assertion that AstraZeneca had engaged in false advertising).

310. *Id.* at 282.

311. *Id.* at 289.

312. *See id.* at 295.

> Thus, citation to the FDA guidelines, in the absence of proof of literal falsity or misleading of the public, is insufficient to show that the claims in the Better is Better campaign are false. Because there are no genuine issues of material fact, and TAP cannot meet its burden on literal falsity, summary judgment will be granted to AstraZeneca on this issue.

[esophageal] damage."[313] The court also pointed out that because of the "'separate jurisprudence that has evolved' under the Lanham Act and under the FDA," the mere violation of the FDA drug advertising regulation in question, even if established, would be insufficient to show that the claims were literally false or misleading under the Lanham Act.[314] TAP could not assert the FDA drug advertising regulations directly because, in general, only the United States may bring actions under the Food, Drug and Cosmetic Act[315] or FDA drug advertising regulations.[316]

VI. Conclusion

The very long and expensive process of new drug research and development might suggest to casual observers that the efficacy standard for drugs is elevated and substantial, but the review of the law just presented reveals that while the *evidence* standard may be substantial, the efficacy standard itself is almost entirely illusory. Under 21 U.S.C. § 355(d), drug sponsors may obtain FDA approval of a new drug so long as they do not "purport[] or . . . represent[]" the drug to have greater efficacy than can be shown by substantial evidence, essentially delegating to drug companies the ability to set as low an efficacy bar as they wish (safety aside). Moreover, the substantial evidence standard itself allows for approval of a new drug based upon two (in some cases, one) statistically significant clinical trials "even though there may be preponderant evidence to the contrary based upon equally reliable studies."[317] Not only do the trials that might constitute

313. *Id.* at 282.

314. *Id.* at 295 (quoting Sandoz Pharm. Corp. v. Richardson-Vicks, Inc., 902 F.2d 222, 229 (3d Cir. 1990)).

315. *See* 21 U.S.C. § 337(a) (2012) ("[A]ll such proceedings for the enforcement, or to restrain violations, of this chapter shall be by and in the name of the United States."); State ex rel. McGraw v. Johnson & Johnson, 704 S.E.2d 677, 687 n.6 (W.Va. 2010) ("[T]ypically, only the United States is entitled to enforce an action under the FDCA. As such, claims brought to enforce violations of the FDCA by any party other than the United States are generally preempted." (citation omitted)).

316. *See* Kemp v. Medtronic, Inc., 231 F.3d 216, 236 (6th Cir. 2000) ("[N]o private cause of action exists for a violation of the FDCA.").

317. S. REP. NO. 87-1744 (1962), *reprinted in* 1962 U.S.C.C.A.N. 2884, 2892.

"preponderant evidence to the contrary" not necessarily bar approval, but selective publication means that they often may not even be known to prescribing doctors, legislators, insurance companies and others, frustrating private efforts at rational drug use. Despite both legislative and private efforts to curb selective publication of only positive trials, publication bias remains a problem.[318]

The sensible elements that underlie the substantial evidence standard also fail to ensure that new drugs possess any particular level of efficacy. "Substantial evidence" is defined by statute to mean "evidence consisting of adequate and well-controlled investigations, including clinical investigations."[319] Accompanying FDA regulations suggest, but do not necessarily require, that these investigations conform to the "gold standard." The gold standard elements of randomization, blinding, and placebo control may help to counter the various types of bias in the experimental process, but none specifies the magnitude of efficacy needed for a trial to succeed. Similarly, the concept of statistical significance, which is generally used by the FDA in its evaluations of efficacy based on gold standard trials, is merely a measure of certainty and not a measure of efficacy. "Clinical significance" and "statistical significance" are distinct concepts, and the law does require that new drugs demonstrate evidence of clinical significance, but the term "clinical significance" merely means significance in humans as opposed to significance in vitro or in non-human animals. The difference in these two terms thus lies not in the magnitude of efficacy, but in the nature of the trial.

Figure 6 summarizes the various challenges in evaluating drug efficacy and the regulatory or scientific standards designed to address each:

---

318. *See supra* Part IV.A.3 and Figure 3 (discussing publication bias).

319. 21 U.S.C. § 355 (d)(7) (2012).

Figure 6: Efficacy Evaluation Problems and Existing Regulatory Solutions

Nature of Problem	**Problem**	**Regulatory Solution**
Relevance	Efficacy data may not be relevant to humans	Clinical significance
Quality	Biases, e.g., • Selection bias • Observer bias • Response bias Anecdotal or unreliable evidence or testimonials	Gold standard • Randomization • Blinding • Placebo control Substantial evidence standard
Certainty	Type I error, (i.e., the appearance of efficacy is actually due to random variation) Selective publication	Statistical significance standard ClinicalTrials.gov
Magnitude	Efficacy level is too low	None (i.e., allow the "free-market" to evaluate the effect a drug "purports or is represented to have")

In short, there is no legal efficacy standard. Although the need for statistical significance might imply that new drugs must at least be better than nothing (zero efficacy), there is no minimum quantum of difference from zero that is required, making the standard illusory in a way reminiscent of how mathematics describes $.\overline{9}$ (point nine repeating) as exactly equal to 1. The influencing of trial results by drug sponsors, suggested by studies that demonstrate inconsistent comparative efficacy results that correlate with the study sponsor, can make whatever tiny quantum of efficacy difference that may be required entirely disappear. Making highly effective drugs may be complex, expensive, and difficult, and the law must be sensitive to the significant technical challenges drug companies face. At the same time, greater awareness of the illusory efficacy standard is badly needed in order to enable physicians, patients, governments, and society at large to make rational choices about the risks they are

willing to undertake and the medicines for which they are willing to pay.