

## Note

### Beating the Odds: The Public Policy of Drug Efficacy and Safety

Noah Lewellen\*

Your scientist friend just inherited a fortune and is looking to invest. Illiterate in finances and distrusting of bankers, he decides to test local psychics to see which is skilled enough to be entrusted with his investments. Being a scientist, he designed a test to determine the statistical significance of a psychic's predictions: he flips a coin 200 times and asks the psychic to call the flip. By random chance, the psychic should be right about fifty percent of the time, or on about 100 flips.<sup>1</sup> One psychic he found, however, was able to predict the coin toss on 114 flips, which, using a chi-square test, gives a p-value of 0.0477.<sup>2</sup> That means, your friend expounds excitedly, that he can be 95 percent confident that the psychic was using some ability—not random chance—to predict the flip.<sup>3</sup> He immediately delivers his fortune into the care of the psychic's financial wisdom.

---

\* J.D. Candidate 2015, University of Minnesota Law School; B.A. 2010, University of Chicago. Thank you to Professor Fred Morrison for his advice and guidance. Thank you also to my colleagues, the board and staff of *Minnesota Law Review*, for the great work they did to publish this Note. Finally, thank you to my parents, Phil and Linda, for their constant support and love, even when I ignored their wishes to become an engineer or a pastor and applied to law school. Copyright © 2015 by Noah Lewellen.

1. Kelly H. Zou et al., *Revisiting the P-value: A Comparison of Statistical Evidence in Clinical and Legal Medical Decision Making*, 8 L. PROBABILITY & RISK 159, 160 (2009).

2. This value is obtained by performing a chi-square test,  $\chi^2 = \sum (\text{observed value} - \text{expected value})^2 / (\text{expected value})$ , after determining the degrees of freedom in the experiment—in this case, 1—and applying the resulting  $\chi^2$  value to a chi-square distribution table. For a brief walkthrough on how to perform a chi-square test, including a chi-square distribution table, see Phillip McClean, *The Chi-Square Test*, N.D. ST. U. (2000), <http://www.ndsu.edu/pubweb/~mcclean/plsc431/mendel/mendel4.htm>.

3. A p-value of 0.0477 indicates that one would expect a random guess for each flip to result in 114 accurate “calls” out of 200 total to occur in roughly 4.77% of such trials (i.e.,  $0.0477 \times 100$ ). See Zou et al., *supra* note 1, at 160.

Even once you understand the significance of the p-value result, it is easy to criticize your friend's decision. Aside from the questionable transfer of skills at predicting coin flips to predicting the stock market, his test suggests that the psychic will be wrong roughly forty-three percent of the time.<sup>4</sup> Even if this misallocation of money never results in a loss—an unlikely proposition—the fact that the investment resulted in no positive benefits means that your friend has missed opportunities to gain money forty-three percent of the time.

If this seems nonsensical, it may be surprising to learn that the p-value test is a critical benchmark for drug approval by the Food and Drug Administration (FDA) in their efficacy testing for drugs that treat anything from the common cold to breast cancer.<sup>5</sup> In the last couple of years, however, the use of p-values in proving causation between drugs and their effects has come under close scrutiny: the Supreme Court held in *Matrixx Initiatives, Inc. v. Siracusano*<sup>6</sup> that statistical significance, through p-value analysis, is not determinative in finding causation between a drug and its side effects.<sup>7</sup> In a recent Ninth Circuit decision, *United States v. Harkonen*,<sup>8</sup> the court implied that p-values were determinative in linking a drug to its positive benefits.<sup>9</sup> These decisions are ostensibly in conflict and thus cast doubt on the ubiquitous use of p-values as critical benchmarks for drug approval and statistical significance in the FDA's drug approval process.

Part I of this Note sets forth the FDA's current drug approval process, including its regulations and guidelines regarding p-values and statistical significance, and provides examples of both approvals and rejections of drug applications. Part II examines the tension between *Matrixx Initiatives, Inc. v. Siracusano* and *United States v. Harkonen* and analyzes how the FDA's drug approval process has affected courts' decisions. Part III contends that *Matrixx* and *Harkonen* can be reconciled

---

4.  $(86/200)*100 = 43\%$  chance of being wrong.

5. See FDA, GUIDANCE FOR INDUSTRY: PROVIDING CLINICAL EVIDENCE OF EFFECTIVENESS FOR HUMAN DRUG AND BIOLOGICAL PRODUCTS 12 (1998), available at <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm078749.pdf> (noting approval of a drug based on a single study with a low p-value).

6. 131 S. Ct. 1309 (2011).

7. See *id.* at 1319 (noting that other factors, not just a p-value, are considered when determining causation).

8. 510 F. App'x 633 (9th Cir. 2013).

9. See *id.* at 637–38; *infra* Part I.C.

by framing statistical significance in the overall structure of the FDA's approval process regarding drug safety and efficacy. This framework can be extended to cover new areas of food and drug law, including the burgeoning field of personalized medicine, by proposing a balancing between statistical significance and other important factors facing consumers. This Note concludes that statistical significance, while highly relevant in the traditional efficacy context, is not as useful in examining the safe use of drugs in uniquely-populated fields like personalized medicine; it then details how courts might flexibly analyze such situations using the FDA's guidance and policy considerations tacitly outlined in *Harkonen* and *Matrixx*.

## I. HISTORY AND MODERN LANDSCAPE OF STATISTICAL SIGNIFICANCE

The importance of statistical significance in the FDA's decision-making processes must be placed in the context of its historical and current purpose of protecting consumers. Section A will briefly discuss the historical purpose of the FDA and the Federal Food, Drug, and Cosmetic Act (FDCA). Section B will provide a background on the FDCA's requirements regarding drug safety and efficacy. Sections C and D introduce *Harkonen* and *Matrixx*, respectively, and provide background on how these cases seem to challenge the FDA's current approach in using statistical significance in safety and efficacy.

### A. PURPOSE AND AUTHORITY OF THE FDA

The FDA was initially concerned solely with the safety of drugs, not their effectiveness.<sup>10</sup> The priorities of the FDA were mostly based on ensuring uniformity in the production of drugs.<sup>11</sup> However, weak statutory language allowed numerous ineffective or harmful drugs to come to market.<sup>12</sup> This changed with the 1962 amendments to the FDCA, which forced manufacturers to prove both the safety and effectiveness of their products to the FDA for its approval.<sup>13</sup> The FDA currently claims responsibility "for protecting the public health by assuring the safety, efficacy and security of human . . . drugs [and]

---

10. Wallace F. Janssen, *The Story of the Laws Behind the Labels*, FDA, <http://www.fda.gov/AboutFDA/WhatWeDo/History/Overviews/ucm056044.htm> (last updated Mar. 11, 2014).

11. *Id.*

12. *Id.*

13. *Id.*

for advancing the public health by helping to speed innovations that make medicines more effective, safer, and more affordable.”<sup>14</sup>

#### B. ACT REQUIREMENTS AND APPROVAL PROCESS

The FDCA requires drug manufacturers to show, through “substantial evidence,” both the safety and efficacy of their proposed drugs.<sup>15</sup> The standard of evidence used is only vaguely defined as

evidence consisting of adequate and well-controlled investigations . . . by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, on the basis of which it could fairly and responsibly be concluded by such experts that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling.<sup>16</sup>

In meeting this standard, manufacturers generally perform at least two large-scale, controlled clinical trials to demonstrate the safety and efficacy of the drug.<sup>17</sup> These studies must follow a variety of guidelines promulgated in FDA regulations generally establishing what an “adequate and well-controlled study” entails.<sup>18</sup>

---

14. *What We Do*, FDA, <http://www.fda.gov/aboutfda/whatwedo> (last updated Aug. 5, 2014).

15. FDA, *supra* note 5, at 2–3; *see also* 21 U.S.C. § 355(d) (2012) (stating that the Secretary shall not approve a drug if there is a lack of substantial evidence proving efficacy or safety).

16. 21 U.S.C. § 355(d). One notable exception to the vagueness of the approval or denial process is the strict Delaney Clause, which declares that no additive may be deemed safe if it is found to cause cancer in humans or experimental animals. Janssen, *supra* note 10. While it is outside the scope of this Note to discuss the Delaney Clause, it is worth noting that the Clause has been rather liberally interpreted by the FDA to allow additives which cause cancer in some circumstances, such as if those additives cause cancer only in doses which no human would be likely to ever consume. *See* Richard A. Merrill, *FDA’s Implementation of the Delaney Clause: Repudiation of Congressional Choice or Reasoned Adaptation to Scientific Progress?*, 5 YALE J. ON REG., Winter 1988, at 6–7.

17. Anup Malani et al., *Accounting for Heterogeneous Treatment Effects in the FDA Approval Process*, 67 FOOD & DRUG L.J. 23, 27 (2012). The statute also recognizes that sometimes one study, rather than two, may be acceptable in showing substantial evidence. 21 U.S.C. § 355(d). This is, however, disfavored by the FDA for a variety of reasons, including the higher risk of false positives. FDA, *supra* note 5, at 4–5. Of the single studies that qualify, the FDA notes that they should be, among other things, “particularly persuasive (low p-value).” *Id.* at 12.

18. *See* 21 C.F.R. § 314.126(b) (2014) (stating that studies, among other items, should use a design that allows for comparison and that the method of assigning patients to control groups should minimize bias).

After submission of a manufacturer's studies, reviewers undertake analyses of those studies in order to make a recommendation to approve or deny the new drug application.<sup>19</sup> In a reviewer's analysis of efficacy claims made in a manufacturer's clinical study, nearly every aspect of the claim is reviewed with statistical significance, in the form of a p-value, as a key determining factor.<sup>20</sup> The p-value is used extensively to uncover study flaws based on, among other factors, study group composition.<sup>21</sup> A finding that a study lacks the statistical significance to show effectiveness, especially in conjunction with a finding that serious adverse effects have been implicated, generally results in an application denial.<sup>22</sup>

### C. EFFECTIVENESS AND *HARKONEN*

Beyond study design, however, there are few hard-and-fast rules as to what may constitute substantial evidence of effectiveness. The Act does not define "effectiveness," and the Third Circuit upheld the FDA's interpretation that "statistical significance" is insufficient without substantial evidence of "therapeutic effect."<sup>23</sup> Some authors have claimed that the FDA employs

---

19. See generally FDA, *supra* note 5 (providing guidance to the industry on how to design a study to show clinical effectiveness for an effective NDA).

20. See, e.g., YUAN WHO CHEN, FDA, SBLA 103780/5010, STATISTICAL REVIEW 8–12 (2003), available at <http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/ApprovalApplications/TherapeuticBiologicApplications/ucm106184.pdf>.

21. *Id.*; see also 21 C.F.R. § 314.126(b)(4). This process is important because the FDA recognizes that studies are generally designed by manufacturers to achieve positive results. See FDA, *supra* note 5, at 4 (noting biases can lead to false conclusions). Analyses such as these keep manufacturers from hiding study manipulation behind acceptable efficacy ratings. See Malani et al., *supra* note 17, at 25 (discussing how the FDA wishes to avoid manipulative data mining by drug companies).

22. See, e.g., ERIC BASTINGS, FDA, CLINICAL/CLINICAL PHARMACOLOGY AND BIOPHARMACEUTICS REVIEW FOR NDA 20-626, SE5-004 (2004), available at <http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/DevelopmentResources/UCM163202.pdf>.

23. Warner-Lambert Co. v. Heckler, 787 F.2d 147, 155 (3d Cir. 1986). In making this decision, the court cited several studies regarding Chymoral, the drug in question. *Id.* One study relied on what the FDA found to be inappropriate post hoc stratification to show statistical significance. *Id.* It is important to note, as the court did, that the Commissioner stated that "the most reliable test of effectiveness is the comparison of the total drug group to the total placebo group." *Id.* This test relies on statistical analysis and indicates that the application may have been successful had there been a substantial effect shown by unbiased statistical analysis. The second study the court discussed, however, declared that, despite well-founded statistical significance in effect,

other standards, such as an amorphous “clinical significance” for prescription drugs.<sup>24</sup> However, it is largely undisputed that the FDA requires statistical significance to establish the basis of scientific claims.<sup>25</sup> Statistical significance at the very least provides some indication of effect,<sup>26</sup> which is perhaps why the FDA requires it as an indicator for drug approval. Lack of statistical significance in a study comparing a drug to a control, however, does not affirmatively indicate that there is no difference between the treatment and control.<sup>27</sup> For example, in a study that requires a p-value of 0.05 to find statistical significance, an empirical finding of a p-value of 0.055 will not indicate said significance, even though it is highly likely that the study reached the result due to an effect outside of random chance.<sup>28</sup>

Despite the fact that the lack of statistical significance does not automatically indicate a lack of efficacy in drug studies, the Ninth Circuit’s decision in *United States v. Harkonen*<sup>29</sup> indicates that a lack of statistical significance in a drug study is enough to prohibit a manufacturer from advertising its drugs as effective.<sup>30</sup> In that case, Harkonen, a physician, researcher, and former CEO of InterMune, Inc., was convicted of wire fraud for issuing a press release touting the effectiveness of a new

---

the achieved effect was “therapeutically trivial,” and therefore ineffective for its intended use. *Id.* at 156.

24. See Sarah M.R. Cravens, Note, *The Usage and Meaning of “Clinical Significance” in Drug-Related Litigation*, 59 WASH. & LEE L. REV. 553, 568 (2002) (noting that FDA approval cases are not clear about how to define significance).

25. *Id.* at 591.

26. See Brief for the United States as Amicus Curiae Supporting Respondents at 14, *Matrixx Initiatives, Inc. v. Siracusano*, 131 S. Ct. 1309 (2011) (No. 09-1156) [hereinafter U.S. Brief] (“[S]tatistical significance provides some indication about the validity of a correlation between a product and [an effect] . . .”).

27. See Brief of Amici Curiae Statistics Experts Professors Deirdre N. McCloskey and Stephen T. Ziliak in Support of Respondents at 7–8, *Matrixx*, 131 S. Ct. 1309 (No. 09-1156) [hereinafter Brief for Statistics Experts] (“[F]ailing to reject the null hypothesis does not mean that one should then accept it.”).

28. *Cf. id.* at 8 (discussing how failing to reject the null hypothesis does not automatically indicate a certain lack of causality). When one fails to reject the null hypothesis, despite the existence of causation, this is known as a Type II error, or a false negative. *Id.* at 9. This Note primarily discusses Type II errors.

29. 510 F. App’x 633 (9th Cir. 2013).

30. See *id.* at 637–38 (holding that evidence was sufficient to find a statement of efficacy misleading).

prescription medication, Actimmune.<sup>31</sup> The press release reported the preliminary results of a standard clinical trial which showed that patients with idiopathic pulmonary fibrosis were 40 percent more likely to survive if taking Actimmune instead of a placebo.<sup>32</sup> Despite this appearance of effectiveness, the results were not statistically significant with regard to the study's pre-set endpoint: at 0.084, the p-value exceeded 0.05.<sup>33</sup> In its prosecution of Harkonen, the government contended that the study merely suggested a survival benefit, but failed to affirmatively demonstrate one.<sup>34</sup>

Harkonen's press release contained such headlines as "InterMune Announces Phase III Data Demonstrating Survival Benefit of Actimmune in IPF,"<sup>35</sup> and "Reduces Mortality by 70% in Patients with Mild to Moderate Disease."<sup>36</sup> These announcements were made after two members of the FDA's medical review staff—not acting for the Agency—informed Harkonen that the data were inconclusive and insufficient for FDA approval to treat IPF; they indicated further studies would be required.<sup>37</sup> Moreover, the p-value for the primary endpoint of the study was 0.5.<sup>38</sup> The government argued that the lack of statistical significance for this preset endpoint meant that no one could draw any conclusions from the trial and that any claim that Actimmune had any survival benefit was "just false."<sup>39</sup>

In affirming Harkonen's conviction of wire fraud, the Ninth Circuit noted that such a conviction does not require a showing that the defendant's representations are universally considered

---

31. Petition for Writ of Certiorari at 1–2, *Harkonen*, 510 F. App'x 633 (9th Cir. 2013) (No. 13–180).

32. *Id.* at 2.

33. *Id.* at 2, 5.

34. *Id.* at 2. It is interesting to note that the set endpoint of the clinical trials discussed in *Harkonen* was not based purely on survival, but on "an approximately ten percent improvement in survival without progression in disease severity." *Id.* at 5. The results for this endpoint were not statistically significant. *Id.*

35. IPF stands for idiopathic pulmonary fibrosis, a progressive lung disease. See *What Is Idiopathic Pulmonary Fibrosis?*, NAT'L HEART, LUNG & BLOOD INST. (Sept. 20, 2011), <http://www.nhlbi.nih.gov/health/health-topics/topics/ipf>.

36. Petition for Writ of Certiorari, *supra* note 31, at 7.

37. *Id.* at 8, 10.

38. *Id.* at 9.

39. *Id.* at 10. For comparison, a p-value of 0.5034 in a coin-flipping experiment measuring the frequency of heads and tails for bias is generally considered to indicate a non-significant result and an unbiased coin. Zou et al., *supra* note 1, at 160.

false.<sup>40</sup> The district court noted that a number of witnesses testified that the data demonstrated a survival benefit.<sup>41</sup> In dismissing the relevance of this testimony, the court stated that it could not acquit someone of fraud simply because other experts in the field would have made similar misrepresentations given the set of data presented.<sup>42</sup> At the same time, the district court found that a “p-value of 0.05 [was] somewhat of a magic number,” and that results that report a p-value above it “are generally considered unreliable and [thus] not statistically significant.”<sup>43</sup>

The government’s position in *Harkonen* was a stark change from its stance in an amicus brief in *Matrixx* urging the Supreme Court to reject the theory that statistical significance determines scientific truth.<sup>44</sup> In fact, the *Matrixx* court agreed with the government in that instance, contrasting the Ninth Circuit’s findings in *Harkonen*.<sup>45</sup>

#### D. SAFETY AND *MATRIXx*

Despite safety being held to the same substantial evidence standard as efficacy by statute,<sup>46</sup> in practice, the FDA has applied this standard much differently. Even the indication—even if not statistically significant—of adverse effects may have a bearing on a new drug application’s (NDA) denial.<sup>47</sup> Adverse effects may manifest long after the conclusion of a clinical trial, and statisticians have persuasively argued that statistical significance is an incomplete and anemic standard for determining

---

40. *United States v. Harkonen*, 510 F. App’x 633, 637 (9th Cir. 2013). In fact, the court cited case law to show that fraud encompassed a variety of behavior, including any “dishonest method[] or scheme[],” and any “trick, deceit, chicanery, or overreaching.” *Id.* at 637 (quoting *Carpenter v. United States*, 484 U.S. 19, 27 (1987)). The court noted that the fraudulent nature of a defendant’s actions is measured by a non-technical standard. *Id.* (citing *United States v. Woods*, 335 F.3d 993, 998 (9th Cir. 2003)).

41. Petition for Writ of Certiorari, *supra* note 31, at 11.

42. *Id.*

43. *Id.* (quoting *United States v. Harkonen*, No. C 08–00164 MHP, 2010 WL 2985257, at \*5 (N.D. Cal. Jul. 27, 2010)). The court apparently reached this conclusion at the behest of other experts. *See Harkonen*, 2010 WL 2985257, at \*5.

44. Petition for Writ of Certiorari, *supra* note 31, at 3.

45. *See Matrixx Initiatives, Inc. v. Siracusano*, 131 S. Ct. 1309, 1321–23 (2011) (holding that reports can be material despite lack of statistical significance).

46. *See* 21 U.S.C. § 355(d) (2012).

47. *See, e.g., BASTINGS*, *supra* note 22, at 3.

the relevancy of adverse effects.<sup>48</sup> Indeed, the FDA's own guidance documents reflect the understanding that the clinical tests required for efficacy and safety testing generally lack the power required to fully explore all potential adverse effects.<sup>49</sup> This, too, was the Solicitor General's position in an amicus brief filed in *Matrixx Initiatives, Inc. v. Siracusano*.<sup>50</sup>

Instead, the FDA endorses a more holistic approach, adopting factors such as association strength, temporal relationships between adverse events and drug use, evidence of dose response, biological plausibility, and even the seriousness of the effect in relation to the treated disease.<sup>51</sup> The FDA has indicated that it may withdraw approval of a drug based on even the suspicion of causation shown in post-market research.<sup>52</sup>

In *Matrixx*, the Supreme Court found that Matrixx could be liable to its investors for securities fraud stemming from its lack of disclosure of reported adverse effects of its over-the-counter cold remedy, Zicam, even though the incidence of those reported effects was not statistically significant.<sup>53</sup> In doing so, the Court affirmed a decision by the Ninth Circuit—the same circuit that decided *Harkonen*.<sup>54</sup>

Reports of adverse effects began to emerge around 1999, when a neurologist called Matrixx after discovering a possible link between Zicam and anosmia, a loss of smell.<sup>55</sup> Other reports began to arrive, and Matrixx's vice president for research

---

48. See Brief for Statistics Experts, *supra* note 27, at 21 (noting that factors such as sample size and lack of data may make calculating statistical significance “futile”).

49. See International Conference on Harmonisation; Guidance on Statistical Principles for Clinical Trials; Availability, 63 Fed. Reg. 49,583, 49,596 (Sept. 16, 1998) [hereinafter Harmonisation]. The FDA notes that p-values are “sometimes useful” in safety analyses, but an approach that combines both qualitative and quantitative approaches is preferred. *Id.*

50. See U.S. Brief, *supra* note 26, at 13 (“Statistical significance is a limited and non-exclusive tool for inferring causation[.]”).

51. See FDA, GUIDANCE FOR INDUSTRY: GOOD PHARMACOVIGILANCE PRACTICES AND PHARMACOEPIDEMIOLOGIC ASSESSMENT (2005), available at <http://www.fda.gov/downloads/RegulatingInformation/Guidances/UCM126834.pdf>.

52. See FDA, THE CLINICAL IMPACT OF ADVERSE EVENT REPORTING 6–7 (1996), available at <http://www.fda.gov/downloads/Safety/MedWatch/UCM168505.pdf>; see also *Matrixx Initiatives, Inc. v. Siracusano*, 131 S. Ct. 1309, 1320 (2011).

53. *Matrixx*, 131 S. Ct. at 1317–18.

54. *Id.* at 1314; *United States v. Harkonen*, 510 F. App'x 633, 633 (9th Cir. 2013).

55. See *Matrixx*, 131 S. Ct. at 1314.

and development, Timothy Clarot, reached out to at least one reporting doctor to discuss the effects.<sup>56</sup> During one of these phone calls, Clarot was informed about studies from the 1930s and 1980s discussing zinc's<sup>57</sup> toxicity and ability to cause anosmia.<sup>58</sup> Clarot also reached out to prevent a doctor from explicitly mentioning Zicam in a poster presentation to the American Rhinologic Society which discussed patients' resulting anosmia after using the drug.<sup>59</sup> At least two plaintiffs had sued Matrixx for allegedly losing their smell due to Zicam use by the time Matrixx made public statements claiming that they expected revenues to be up by eighty percent in January 2004.<sup>60</sup> This was also in spite of Matrixx's own report to the Securities and Exchange Commission, which described a potential material adverse effect which it expected could result in product liability claims.<sup>61</sup> On January 30, 2004, the FDA was reportedly looking into complaints about anosmia resulting from Zicam use.<sup>62</sup>

In a public statement responding to this report, Matrixx announced that "the safety and efficacy of zinc gluconate for the treatment of . . . the common cold have been well established in two double-blind, placebo-controlled, randomized clinical trials," that neither study had reported any adverse effects of anosmia, and that the overall incidence of adverse effects associated with zinc gluconate was statistically insignificant.<sup>63</sup> Matrixx argued to the Court that the lack of statistical significance eliminated its responsibility to disclose adverse effects reports to its investors.<sup>64</sup> Matrixx's claim "rest[ed] on the premise that statistical significance is the only reliable indication of causation."<sup>65</sup> The government explicitly opposed this position,

---

56. *See id.*

57. Zicam's active ingredient, zinc gluconate, used a zinc base. *Id.*

58. *See id.* at 1314–15.

59. *Id.* at 1315.

60. *Id.*

61. *Id.*

62. *Id.* Zicam was approved by the FDA for use under its guidelines regarding homeopathic treatments. *See id.* at 1316.

63. *Id.* at 1316

64. *Id.* at 1318–19. The Court noted that such a bright-line rule would artificially exclude information that a reasonable investor might use in making a financial decision. *See id.* at 1319 (citing *Basic, Inc. v. Levinson*, 485 U.S. 224, 236 (1988)).

65. *Id.* at 1319

noting that experts, such as medical specialists, routinely “consider multiple factors in assessing causation. . . .”<sup>66</sup>

In affirming the Ninth Circuit’s decision that a claim of securities fraud against Matrixx could go forward, the *Matrixx* decision noted that there are a number of ways both courts and medical experts show causation outside of statistical significance.<sup>67</sup> The Court held that something more than statistical significance in clinical trials must be considered when deciding what to disclose to investors, suggesting that the source, content, and context of adverse reports may be relevant.<sup>68</sup>

Harkonen, in his petition for certiorari, claimed that his conviction by the Ninth Circuit directly conflicts with the Supreme Court and government in the *Matrixx* decision.<sup>69</sup> In a way, Harkonen is correct—the government and courts are weighting p-values differently by holding statistical significance to be conclusive when deciding effectiveness and insufficiently conclusive when considering safety. Through these decisions, the government is beginning to shape the way statistical significance should be used in drug manufacturing, marketing, and post-market analysis.

---

66. U.S. Brief, *supra* note 26, at 8. Medical researchers, filing an independent brief, pointed out that some adverse effects, while undoubtedly linked to a drug’s use, might occur so infrequently or with such subtlety that no researcher or manufacturer would likely be able to compile a dataset of appropriate quality or quantity. See Brief of Amici Curiae Medical Researchers in Support of Respondents Urging Affirmance at 11, *Matrixx Initiatives, Inc. v. Siracusano*, 131 S. Ct. 1309 (2011) (No. 09-1156).

67. See *Matrixx*, 131 S. Ct. at 1319–20.

68. *Id.* at 1321. The language involving examination of the context of reports bears a strong resemblance to the FDA’s approach to treating postmarket reports of adverse effects. Compare *Matrixx*, 131 S. Ct. at 1321, with Harmonisation, *supra* note 49, at 49,596. Prior to the *Matrixx* decision, a number of securities fraud cases with similar claims regarding statistically insignificant adverse effects reports had been dismissed. See Joseph B. Kadane, *Matrixx v. Siracusano: What Do Courts Mean by ‘Statistical Significance’?*, 11 L. PROBABILITY & RISK 41, 42–44 (2011). While it is not within the scope of this Note to discuss the case history leading up to the *Matrixx* decision, at least one author has come to the conclusion that *Matrixx* follows that line of cases, but comes to a different conclusion only due to the strength of the facts presented in the case. *Id.* at 46. This would indicate that, despite previous decisions dismissing cases regarding statistically insignificant adverse facts, the courts have always espoused the idea that statistical significance is not the end-all, be-all of showing causation. See *id.* at 46–47.

69. See Petition for Writ of Certiorari, *supra* note 31, at 26 (“The First Amendment does not permit the government to prosecute a scientific viewpoint in one courtroom while championing that same viewpoint in another.”).

## II. THE FDA TREATS P-VALUES DIFFERENTLY FOR DIFFERENT PURPOSES, AND THE FDA'S POLICY CHOICE HAS EXTENDED INTO THE COURTS

While the decisions in *Harkonen* and *Matrixx* seem to cast light on a conflict which requires the FDA's attention, this Note will demonstrate how the FDA's policies regarding safety and efficacy already support the reasoning inherent in those opinions. Section A describes how the FDA actually values statistical significance minimally when considering the safety of a drug and poses some suggestions as to why this may be. Section B contrasts this approach with the high value the FDA places on statistical significance in the context of efficacy, and suggests the reasons behind that focus. Section C recognizes that the FDA occasionally errs from the standard framework of analysis set out in Part II and describes when that deviation might occur.

### A. STATISTICAL SIGNIFICANCE IS OF LITTLE IMPORTANCE WHEN CONSIDERING THE SAFETY OF A DRUG

Statistical significance is less important to the FDA's analysis of the safety of a drug than the mere fact that adverse effects occur after the administration of that drug. While not codified in regulation, this stance is reflected in the FDA's rampant use of other factors<sup>70</sup> and their use of tests that are generally frowned upon in proving the effectiveness of a drug.<sup>71</sup> The FDA and courts disregard stringent methods of linking adverse effects to product use because public policy dictates that safety be paramount, both in restricting potentially non-effective drugs and in withdrawing potentially dangerous ones.

#### 1. Current State of Statistical Significance in Safety Analyses

The FDA uses a number of factors outside of statistical significance to assess the safety of a product, such as temporality between use and event, consistency across studies, evidence of a dose-dependent response, biological plausibility, and seriousness of the side effect relative to the disease being treated, among others.<sup>72</sup> In analyzing these factors, the FDA approves a drug sponsor's data mining of its studies in order to further

---

70. See FDA, *supra* note 51, at 18.

71. See FDA, *supra* note 5, at 19.

72. See FDA, *supra* note 51, at 18.

hone the precise nature of the safety risk.<sup>73</sup> Such post hoc analyses are generally frowned upon in the drug approval process due to risks of manufacturers mining data to achieve a higher rate of success for their products.<sup>74</sup>

In analyzing marketed drugs' safety, the FDA may gather information not only from manufacturers' studies, but also through its Adverse Event Reporting System (FAERS).<sup>75</sup> Reports made through FAERS may be submitted by doctors or pharmacists, but may also be created by layperson consumers.<sup>76</sup> The FDA "does not require that a causal relationship between a product and event be proven."<sup>77</sup> That being said, the FDA has noted that "a temporal relationship between medical product and adverse event [sic] . . . can make isolated reports conclusive as to a product-event association."<sup>78</sup> The FDA uses these FAERS reports to prescribe a variety of actions for a safety review of a drug, and, in some cases, may even use these reports to remove a product from the market.<sup>79</sup> Reported adverse effects sometimes form part of the basis of a recommended denial for a drug.<sup>80</sup> The FDA's broad and stringent requirements regarding FAERS reporting reflects how little it relies on statistical significance when considering the post-market safety of a drug,

---

73. See *id.* at 17. The FDA has noted that post hoc surveillance of a drug rarely results in a finding of causality between a drug and an adverse effect. See FDA, *supra* note 52, at 7. Despite that statement, such surveillance may produce a "prominent degree of suspicion," which may be considered sufficient for a regulatory decision. *Id.*

74. See FDA, *supra* note 5, at 19 (describing a report sufficient for drug approval as being one that yields a consistent conclusion of efficacy without using post hoc analyses); see also Malani et al., *supra* note 17, at 24 (suggesting that drug companies' data mining results in too many false positives).

75. See *FDA Adverse Event Reporting System (FAERS)*, FDA, <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm> (last updated Sept. 8, 2014).

76. *Id.*

77. *Id.*

78. FDA, *supra* note 52, at 7.

79. See *FDA Adverse Event Reporting System (FAERS)*, *supra* note 75; see also FDA, *supra* note 52, at 7.

80. See, e.g., *BASTINGS*, *supra* note 22, at 3. It is interesting to note that the FDA Commissioner's discretion to remove a drug from the marketplace over safety concerns does not require that a lack of safety be shown in any manufacturer's study. See *Fisher Bros. v. United States*, 46 F.3d 279, 285–86 (3d Cir. 1995). In practice, recalls are rare, and involuntary recalls even rarer, but statistical significance need not take an active role in a recall analysis. See 21 C.F.R. § 314.620(a) (2014) (describing requirements for a recall, including a catchall of "[o]ther evidence [that] demonstrates that the drug product is not shown to be safe or effective under its conditions of use").

and its weak reliance on such was explicitly acknowledged by the government in *Matrixx*.<sup>81</sup>

The *Matrixx* court held that the FDA's position—that statistical significance is not dispositive in the relevance of adverse effects of a drug—was convincing and extended the relevance of other factors into the world of securities fraud.<sup>82</sup> Foremost in the Court's reasoning was that adverse reports take many forms and that several factors the FDA considers in issuing safety-related decisions were material in investors' decisions to involve themselves with a drug manufacturer.<sup>83</sup> For example, the Court noted that reports discussed a plausibly-causal link between Zicam and anosmia and that “[c]onsumers likely would have viewed the risk associated with Zicam . . . as substantially outweighing its benefit.”<sup>84</sup> This bears a striking resemblance to the FDA's safety consideration involving the “seriousness of the event relative to the disease being treated.”<sup>85</sup> The Court also noted that it was important that *Matrixx* had “ignored reports linking Zicam and anosmia and claimed that zinc gluconate's safety was well-established, when it had evidence of a biological link between . . . zinc and anosmia.”<sup>86</sup> This, too, resembles the FDA's consideration of “biologic plausibility.”<sup>87</sup> Both the FDA and the courts, then, appear to weigh statistical significance similarly and similarly lightly when discussing a drug's safety.

## 2. The Statistical Realities of Adverse Effects and Public Concern for Safety Demand the Low Weight Afforded to Statistical Significance by the FDA

Congress gave the FDA wide berth in authorizing it to reject drugs based on questionable findings of safety.<sup>88</sup> Safety and public health have long been considered important government

---

81. See U.S. Brief, *supra* note 26, at 19 (“[The] FDA does not apply any single metric for determining when additional inquiry or action is necessary, and it certainly does not insist on ‘statistical significance.’”).

82. See *Matrixx Initiatives, Inc. v. Siracusano*, 131 S. Ct. 1309, 1313–14 (2011).

83. *Id.* at 1323.

84. *Id.*

85. See FDA, *supra* note 51, at 18.

86. See *Matrixx*, 131 S. Ct. at 1323.

87. See FDA, *supra* note 51, at 18.

88. See 21 U.S.C. § 355(d)(2), (d)(4) (2012) (stating that the Secretary may reject a drug if its study fails to affirmatively demonstrate its safety or lacks sufficient evidence to show its safety).

interests deserving of special treatment under the law.<sup>89</sup> Approving drugs for safe use should be policed just as strictly.

The most obvious and compelling reason for withdrawing drug approval is the occurrence of a life-threatening illness associated with the drug's use. In cases involving life-threatening side effects, the FDA has not shied away from issuing public health advisories based on a low number of complaints.<sup>90</sup> In many cases, this results in a voluntary withdrawal of the drug from the market by its manufacturer.<sup>91</sup>

Other than life-threatening illnesses, however, the FDA has recognized a myriad of other reasons why a drug might unacceptably threaten public health.<sup>92</sup> Zicam, the drug in *Matrixx*, provides a prime example: Zicam was a cold remedy that potentially led to anosmia through a side effect of its main ingredient, which contained zinc.<sup>93</sup> That is, it treated a low-danger illness with a drug that had the potential to cause a severe side effect through a well-known biologically-indicated chemical process. This does not comport with the FDA's mission to protect consumers from unsafe drugs. This balance rapidly changes if Zicam had, for instance, been effective in fighting liver cancer. The FDA satisfies its statutory goals by preventing the loss of smell to consumers suffering from the common cold. On the other hand, preventing consumers from treating their liver cancer—even at the cost of their sense of smell—does not go towards fulfilling those goals.<sup>94</sup>

---

89. See, e.g., *Craig v. Boren*, 429 U.S. 190, 199–200 (1976) (“Clearly, the protection of public health and safety represents an important function of state and local governments.”); cf. *New Orleans Gas Light Co. v. Drainage Comm'n*, 197 U.S. 453, 460 (“The drainage of a city in the interest of its public health and welfare is one of the most important purposes for which the police power can be exercised.”).

90. See, e.g., Press Release, FDA, FDA Statement on the Voluntary Withdrawal of Raptiva from the U.S. Market (Apr. 8, 2009), available at <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm149561.htm> (issuing such an advisory after four patients using Raptiva for psoriasis developed multifocal leukoencephalopathy, three of whom died).

91. See, e.g., *id.*

92. See, e.g., FDA, *supra* note 51, at 18.

93. See *Matrixx Initiatives, Inc. v. Siracusano*, 131 S. Ct. 1309, 1312 (2011).

94. The FDA's guidance regarding safety analyses of a drug reflect this balance. See *supra* note 72 and accompanying text. Its regulations regarding whether or not to issue a recall, however, do not. See 21 C.F.R. § 7.41 (2014). This lack seems like it would bias the FDA's decision towards issuing a recall of drugs the intended effects of which are far more beneficial than their side effects are harmful, but the FDA's initial safety determination has a great im-

None of these reasons absolutely require the introduction of statistical significance to the immediate analysis.<sup>95</sup> Nor should a formal statistical significance showing be required when deciding whether a drug is unsafe, as such a requirement would take time, potentially allowing more injuries to develop from use of the drug. Statistical significance may, however, be useful in adjusting the weight of certain factors in the balancing test the FDA already uses. For example, if the incidence of a minor side effect, such as a migraine headache, were associated with a drug treating a minor illness, such as bronchitis, formal testing involving statistical significance analysis may be useful to help distinguish two things: a likelihood of causation and a more accurate indication of the number of sufferers of the side effect. Statistical significance does not, in and of itself, prove causation,<sup>96</sup> but researchers may reliably draw a strong inference of causation from it.<sup>97</sup> Even if significance is not reached in the study, a more formal study may assist the FDA in determining an approximate percentage of consumers who would suffer from the side effect, as these statistics are often skewed when only considering voluntary FAERS reports.<sup>98</sup> In these ways, statistical significance testing would assist the FDA in determining how much weight to put behind the relatively low-level side effect of a migraine headache.

Critics have argued that forcing a company to report all adverse events to its investors would be harmful to drug companies and, in turn, consumers everywhere.<sup>99</sup> It is easy to extrapolate this argument into the FDA regulatory realm by claiming that rampant FDA regulation based on non-causative reports would be detrimental to both drug companies and consumers. Requiring that drug developers only submit FAERS

---

pact on the decision as to whether to even consider a recall in the first place. *See id.* § 7.40.

95. Presumably, statistical significance has been involved in the empirical determination of zinc's toxicity. However, no independent study with statistical significance regarding the toxicity of zinc gluconate (Zicam's active ingredient) was required when the *Matrixx* court found that doctors' warnings to *Matrixx* regarding zinc's toxicity were relevant to *Matrixx*'s treatment of that information. *See Matrixx*, 131 S. Ct. at 1314.

96. *See* Brief for Statistics Experts, *supra* note 27, at 19–21.

97. *Id.*

98. *See FDA Adverse Event Reporting System (FAERS)*, *supra* note 75.

99. *See generally* Brief for the Consumer Healthcare Products Association and the Council for Responsible Nutrition as Amici Curiae Supporting Petitioners, *Matrixx Initiatives, Inc. v. Siracusano*, 131 S. Ct. 1309 (2011) (No. 09-1156).

reports which have statistically significant causative connections would, however, ultimately harm drug companies and consumers more. Such a requirement would force manufacturers to undertake preemptive tests to accurately determine the statistical significance of an event, rather than simply flooding the FDA with unsubstantiated reported adverse events. Manufacturers are already required to submit studies to show the safety of their drugs, and the FDA may require studies on individual populations.<sup>100</sup> While this saves the FDA time and money, it forces drug companies to bear that financial burden.<sup>101</sup> In the case of an unsafe drug, some companies may undergo expensive testing, only to have their studies show unacceptable statistical significance of causation, resulting in regulatory action. Even in the case of manufacturers producing safe drugs required to perform a targeted safety study, some risk-averse manufacturers may decide to stop marketing their drug rather than undergo a safety study and run the risk of forced withdrawal.

Dismissing the need for statistical significance to withdraw a perceived unsafe drug is beneficial to both consumers and manufacturers. The FDA currently employs a multifactor test that is suitable in balancing concerns of causation and economy with consumer safety. This “low bar” to agency regulation regarding the approval of drugs over safety concerns is reinforced with the FDA’s stringent policies requiring substantial evidence of drug efficacy.

#### B. STATISTICAL SIGNIFICANCE IS AN IMPORTANT FACTOR IN DETERMINING EFFICACY

While statistical significance is largely meaningless for the FDA when examining a drug’s safety, it receives heavy scrutiny when examining a drug’s efficacy. Such a policy is in keeping with the FDA’s goal of protecting consumer’ safety by only allowing effective drugs to find their way into the marketplace.

---

100. See FDA, *supra* note 5, at 5 (discussing how a study may not be applicable to a certain population and would then require further studies targeting that population).

101. See Charles L. Hooper, *Pharmaceuticals: Economics and Regulation*, in THE CONCISE ENCYCLOPEDIA OF ECONOMICS (2d ed. 2008), available at <http://www.econlib.org/library/Enc/PharmaceuticalsEconomicsandRegulation.html> (“The path through the FDA’s review process is slow and expensive. The ten to fifteen years required to get a drug through the testing and approval process leaves little remaining time on a twenty-year patent.”).

### 1. Current State of Statistical Significance in Efficacy Determinations

Drug applications live and die by their ability to show statistical significance in their efficacy.<sup>102</sup> Drugs that are unable to achieve the pre-set p-value of 0.05 that the FDA gives for statistical significance are generally rejected.<sup>103</sup> Furthermore, the FDA has set out rigorous testing guidelines and prohibited certain types of analysis in order to more stringently police which drugs may be considered “effective” for their intended uses.<sup>104</sup>

The *Harkonen* court put similar weight into its analysis of the relevance of statistical significance to showing causation in convicting Harkonen of fraud.<sup>105</sup> Harkonen’s claims regarding the effectiveness of his drug, Actimmune, came primarily from sub-group post hoc analyses.<sup>106</sup> The results from these types of analyses are “notoriously unreliable” and, in most instances, must be independently confirmed by a future trial.<sup>107</sup> The

---

102. See generally CHEN, *supra* note 20 (using p-values to support conclusions of differences between treatment groups and results in an approval recommendation).

103. Drugs that purport to benefit a niche population suffering from a serious illness are notable exceptions to this rule. The FDA has been somewhat more flexible in allowing these drugs to come to market due to the lack of options relevant consumers have in those fields. FDA, PAVING THE WAY FOR PERSONALIZED MEDICINE: FDA’S ROLE IN A NEW ERA OF MEDICAL PRODUCT DEVELOPMENT 35 (Oct. 2013), available at <http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/PersonalizedMedicine/UCM372421.pdf>. This has not, however, always been the case, and the FDA has been oft-criticized for not being more flexible with its statistical requirements. See, e.g., Jay Barnes, *Perspective: The Right To Save Your Own Life*, NEWS TRIBUNE (June 8, 2014) <http://www.newstribune.com/news/2014/jun/08/perspective-right-save-your-own-life> (“While a stringent FDA approval process makes sense for non-life-threatening diseases and widespread sale of new drugs, it defies logic to forbid Americans who are about to die from taking drugs that might work.”); Erica Teichert, *FDA Clinical Trials Need More Flexibility, Lawmakers Say*, LAW360 (July 9, 2014, 2:15 PM ET), <http://www.law360.com/articles/555714/fda-clinical-trials-need-more-flexibility-lawmakers-say> (indicating that Congress wants the FDA to “allow alternative trial designs and emerging technologies to gain traction”).

104. See FDA, *supra* note 5, at 19 (stating that robust results should require no post hoc analysis).

105. See generally *United States v. Harkonen*, 510 F. App’x 633 (9th Cir. 2013). Despite what the petition for certiorari seems to imply, the court did not go so far as to say Harkonen definitely failed to establish causation, but that his press release was overreaching in its claims of effectiveness. See *Petition for Writ of Certiorari*, *supra* note 31, at 9–10; see generally *Harkonen*, 510 F. App’x 633.

106. See *Harkonen*, 510 F. App’x at 636.

107. *United States v. Harkonen*, No. C 08–00164 MHP, 2010 WL 2985257,

*Harkonen* court did not address what level of proof was required to support Harkonen's claim that his drug was effective for its intended use, but the government argued that the study's only meaningful p-value was for the primary endpoint, that the p-value was 0.5, and that such a value meant that "you can't draw any conclusions from this trial."<sup>108</sup> The government's position, then, is clearly spelled out: p-values are essential for finding causation in efficacy studies.

## 2. The Public Interest of Safety Extends to Efficacy and Demands a Strict Standard of Statistical Significance To Show Causation

The history of the FDA as an agency and the legislative history of the FDCA both indicate that Congress had the safety of consumers in mind when it gave the FDA the power to regulate drugs based on their effectiveness.<sup>109</sup> Congress recognized that no drug is truly safe unless it is also effective in its intended use.<sup>110</sup> Indeed, the opportunity cost of using a drug that purports to have a certain effect but is not effective would be the continuance, and possible worsening, of a disease.

Requiring a finding of efficacy, then, is similar to requiring relatively low levels of severity of side effects for a drug. Unlike in safety findings, however, there are few ways of inferring causation for efficacy outside of controlled studies.<sup>111</sup> One proposed method was to allow anecdotal evidence from physicians who regularly prescribed the drugs in question, but the hearings underlying the 1962 amendments to the FDCA marked a concern that impressions of physicians are "treacherous."<sup>112</sup>

---

at \*7 (N.D. Cal. Jul. 27, 2010). Further damning for Harkonen, in the court's eyes, was that he stated he would "cut that data and slice it until [he] got the kind of results [he was] looking for." *Harkonen*, 510 F. App'x at 636. The ability to "cut and slice" like this via post hoc analyses is exactly the kind of data manipulation the FDA seeks to avoid by prohibiting such data mining.

108. Petition for Writ of Certiorari, *supra* note 31, at 9. The *Harkonen* court declined to opine on this reasoning, presumably to avoid giving Harkonen any argument in its petition, which relies partly on *McAnnulty* to say that the court should not convict someone of fraud who communicates scientifically debatable facts. See Petition for Writ of Certiorari, *supra* note 31, at 10. See generally *Harkonen*, 510 F. App'x 633.

109. See *supra* notes 10–13 and accompanying text.

110. See *Janssen*, *supra* note 10.

111. See *supra* text accompanying notes 75–81. The reporting system is not used for pre-market approval of drugs. See *FDA Adverse Event Reporting System (FAERS)*, *supra* note 75.

112. *Weinberger v. Hynson, Westcott & Dunning*, 412 U.S. 609, 619 (1973).

Thus, anecdotal evidence, such as the FAERS reports in the context of safety, cannot carry significant weight in an efficacy analysis. This stance makes sense, as it creates perverse goals for manufacturers to incentivize physicians and consumers to report their positive results from using a drug.<sup>113</sup>

Other factors the FDA relies upon for safety analyses<sup>114</sup> also make little sense in the efficacy context. For instance, biologic plausibility is generally assumed, given the manufacturer researched and designed a drug to meet demand for a particular disease.<sup>115</sup> Evidence of a dose response is largely irrelevant, until the drug has the effect it purports to have.<sup>116</sup> Statistical significance tests have built-in endpoints, allowing such studies to effectively infer causation for claimed effects.<sup>117</sup>

For a finding of “efficacy,” both a useful effect and a strong inference of causation must be established.<sup>118</sup> These tests cost manufacturers—and thus, consumers—money, and prevent useful drugs from coming to the market quickly,<sup>119</sup> but attempt

---

113. Sometimes, the FDA does take consumers’ reports of effectiveness into account when deciding whether to roll back a full recall for a select group of consumers, but such an action is rare and is accompanied by further testing. See, e.g., Letter Regarding Lotronex from Janet Woodcock, Dir., Ctr. for Drug Evaluation & Research (Dec. 18, 2000), available at <http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm110883.htm>.

114. See *supra* note 72 and accompanying text.

115. Two likely exceptions come to mind. One is where a drug is developed for one use, such as pain relief, but frequently used for another use, such as lowering blood pressure (e.g., aspirin). In this situation, known as “off-label” use, a drug company must go through an additional round of testing for the new intended use if it wishes to market the drug in the new way. See “Off-Label” and Investigational Use of Marketed Drugs, Biologics, and Medical Devices – Information Sheet, FDA (June 25, 2014), <http://www.fda.gov/RegulatoryInformation/Guidances/ucm126486.htm>. Another situation in which the manufacturer may not know of a biologically plausible mechanism by which the drug functions is that in which the drug has been fraudulently developed without adequate scientific research and is actually ineffective. In both of these situations, the well-controlled testing requirements set by the FDA prevents a drug whose biological plausibility is in doubt from release without further testing.

116. For instance, it is certainly possible for a drug to have an effect that is “clinically insignificant.” See Cravens, *supra* note 24, at 593 n.179. A drug’s dosage response may show some causation, but that causation is irrelevant until it has a meaningful effect. *Id.*

117. See Ronald A. Thisted, What Is a P-Value? 5–6 (June 8, 1998) (unpublished manuscript), available at <http://galton.uchicago.edu/~thisted/Distribute/pvalue.pdf>.

118. See *supra* notes 103–04 and accompanying text.

119. See Hooper, *supra* note 101 (discussing how the FDA’s evidentiary re-

to ensure consumers' safety. The FDA currently requires a p-value of 0.05 for efficacy studies of new drugs.<sup>120</sup> That value is something of a "magical standard" that attempts to balance consumers' interests in prohibiting drugs that cannot pass a safe threshold with the reality that absolute causation is impossible to prove without achieving an unrealistic p-value of zero.<sup>121</sup>

### C. THE EFFICIENCY OF A QUICK DRUG APPROVAL IS SOMETIMES MORE IMPORTANT THAN ESTABLISHING A STRONG INFERENCE OF CAUSATION

One of the factors that the FDA takes into account when assessing the safety of a drug is the "degree of benefit the product provides, including availability of other therapies."<sup>122</sup> When other therapies are unavailable, statistical significance testing may not be as crucial in the approval of a drug. For example, in the case of a drug which purports to cure a life-threatening disease for which there is no other treatment, a study reporting a p-value of 0.06, above the limit set by the FDA, should not be fatal to the drug's approval. In this case, the possibility that the drug would help the affected population outweighs the possibility that it might have no effect, as the alternative is that the affected population would have no treatment whatsoever.

Allowing such drugs to go to market without any statistical significance testing should, however, remain forbidden. Such a policy would incentivize "gold rushes" to niche markets for drug manufacturers, as the second-place manufacturers would be required to undergo testing while the first drug was marketed. This policy has obvious perverse incentives, including the release of a fraudulent drug that purports to have an effect with no reliable documentation of such an effect. Even for individuals with no other hopeful drugs, profiteering off of a miserable man's vain hopes of treatment is morally abhorrent and should be prohibited if there is no reason to believe a drug would have any effect. Furthermore, dropping any and all significance testing requirement would force ill consumers to do extensive re-

---

quirements for new drug approvals may take a sponsor a decade to achieve and cost waiting consumers their lives).

120. *In re Nuvelo, Inc. Secs. Litig.*, 2008 U.S. Dist. LEXIS 98441, at \*35 (N.D. Cal. Dec. 4, 2008).

121. *Cf. id.* at 35–36 (discussing how this value is allowed for two studies on the same drug, but how a lower p-value is required for a single study because of the lack of repetition in testing).

122. *See FDA, supra* note 51, at 18.

search in a flood of near-useless drugs to find the least-useless one—an expenditure of time, money, and opportunity of which a terminally ill patient has little.<sup>123</sup>

Requiring p-value testing at all slows the marketing approval process, though removing it completely would be unacceptable. Allowing higher p-values in single studies would aid in the speed with which a drug might be released while keeping some standards by which the FDA could judge the efficacy of a drug.

### III. POLICY CONCERNS DICTATE THAT P-VALUES SHOULD BE TREATED DIFFERENTLY IN DIFFERENT CONTEXTS

This Note has discussed the conflict that seems to arise from the opinions in *Harkonen* and *Matrixx*. It next analyzed the policy decisions that the FDA makes regarding safety and efficacy with regards to statistical significance. Next, in Section A, this Note compares the reasoning in *Harkonen* and *Matrixx* with the policy decisions the FDA has made regarding statistical significance and concludes that the courts have essentially adopted the FDA's reasoning, although couched in other arguments. Thus, while arguments rage over whether the *Matrixx* court or the *Harkonen* court got it right,<sup>124</sup> such arguments are misplaced because the courts are in accordance. In Section B, this Note forwards one field—personalized medicine—as one in which the FDA should explicitly discuss its policies regarding p-values and safety and efficacy. In Section C, this Note discusses the methods by which the FDA might promulgate its stance and the dangers in taking such a stance. The Note concludes in Section D by offering courts presumptions to hold in cases similar to those of *Harkonen* and *Matrixx* that focus on protecting public safety.

---

123. See Kathleen Doheny, *Fake Malaria Drugs Thwart Global Efforts To Treat Dangerous Diseases*, TAKEPART (Sept. 26, 2012), <http://www.takepart.com/article/2012/09/25/counterfeit-medications-thwart-global-efforts-treat-dangerous-diseases> (discussing how fake drugs with no testing requirements may contain some useful, active ingredients but ultimately harm consumers).

124. See Petition for Writ of Certiorari, *supra* note 31, at 1–2; see also Nathan A. Schachtman, *The Matrixx Motion in U.S. v. Harkonen* (Dec. 17, 2012), <http://schachtmanlaw.com/the-matrixx-motion-in-u-s-v-harkonen>.

A. *MATRIX*X AND *HARKONEN* STAND FOR TWO SIDES OF THE SAME COIN

The issue in *Matrixx* was whether or not *Matrixx* had made statements that were misleading as to material facts.<sup>125</sup> The Supreme Court in *Matrixx* first considered the merits of *Matrixx*'s contention that "adverse event reports that do not reveal a statistically significant increased risk of adverse effects from product use are not material information."<sup>126</sup> Relying on experts' opinions, including the FDA's, the Court focused on the broad nature of information the FDA considers relevant and the lower standards with which the FDA treats causality and statistical significance in post-market surveillance.<sup>127</sup> The Court stopped short of requiring that all adverse events reported to a company be disclosed to investors,<sup>128</sup> but noted that the *type* of information that *Matrixx* received should have tipped it off to a change in the "total mix" of information provided to it.<sup>129</sup>

The *Matrixx* Court found that this "total mix" was shifted because of reports from "more than 10" patients who had lost their smell from three medical professionals and researchers and due to *Matrixx*'s knowledge of a presentation made at a medical conference about a potential causal link between *Zicam* and anosmia.<sup>130</sup> The Court noted that that presentation included a case study which suggested "a temporal relationship between *Zicam* use and anosmia."<sup>131</sup>

The *Matrixx* Court relied heavily on the criteria that the FDA uses in conjunction with the FAERS system to investigate potentially harmful drugs. The Court, as well as the FDA, looked to the number of similar complaints of the same side effects (i.e., consistency across reports) as well as other numerous factors, including the temporality of the drug use and implicated side effects.<sup>132</sup>

---

125. *Matrixx Initiatives, Inc. v. Siracusano*, 131 S. Ct. 1309, 1318 (2011).

126. *Id.* (internal quotation marks omitted).

127. *See id.* at 1320.

128. *See id.* at 1321.

129. *Id.* at 1322.

130. *Id.*

131. *Id.*

132. These criteria cited by the court are similar to those given by the FDA as considerations when looking at the safety of a drug. *See FDA, supra* note 51 at 6–7.

The court in *Harkonen* focused on the materiality of evidence regarding not safety and side effects, but efficacy.<sup>133</sup> The court found against Harkonen based mainly on his knowledge of the likelihood of his drug's rejection, his willingness to manipulate data to get the results he wanted, and InterMune's biostatistics expert's testimony that the company overstated the conclusiveness of the results.<sup>134</sup> While the court did not explicitly rely on the FDA's statistical significance standards for drug approval regarding efficacy, it was swayed by Harkonen's behavior, especially when he said that he wanted to take the data from the drug study and cut and slice it until he got the results (the p-value) he was looking for.<sup>135</sup> The FDA's requirement for statistical significance seems to have been the driving factor for Harkonen's behavior.

Because this was a jury trial, we may never know what the tipping point of the testimony was, but it is clear from the record that the jury heard testimony from InterMune's researchers stating that "[t]he indices didn't show any difference whatsoever, and the p-values were very high showing no evidence whatsoever."<sup>136</sup> That researcher also informed Harkonen that they "had no evidence of an effect on the primary efficacy endpoint."<sup>137</sup> The researcher suggested further tests to follow up on a post hoc analysis that showed possible survival benefits.<sup>138</sup> Thus, *some* evidence (i.e., post hoc analysis) existed to show the glimmer of some benefits. The FDA, however, does not support the use of post hoc analysis in efficacy testing, as it is inherently unreliable.<sup>139</sup> When Harkonen represented this glimmer of success in survival benefits as being supported by data, then, the jury found him guilty of fraud.<sup>140</sup> This strict reliance on reliable statistical significance by the jury, upheld by the Ninth Circuit, mirrors the FDA's approach to efficacy and statistical significance.

*Harkonen* and *Matrixx* both represent the FDA's and our society's views that, as matters of public policy, efficacy should

---

133. See *United States v. Harkonen*, 510 F. App'x 633, 636–37 (9th Cir. 2013).

134. *Id.* at 637–38.

135. *Id.* at 636.

136. *United States v. Harkonen*, No. C 08–00164 MHP, 2010 WL 2985257, at \*7 (N.D. Cal. Jul. 27, 2010).

137. *Id.*

138. *Id.*

139. See FDA, *supra* note 5, at 19.

140. *Harkonen*, 510 F. App'x at 636–37.

be held to a stricter standard to show causality (requiring statistical significance), while a showing of a product's danger (imputing its safety) may be satisfied with far less evidence. These decisions reflect the view that the safety of a consumer is paramount in drug transactions. The cases are not at odds, but rather are pointing to the same conclusion: companies face a high burden when bringing drugs to market, as there is a presumption of inefficacy in place to protect consumers; and companies face a high burden once deleterious effects have been linked to their product, as there is a presumption of causation that companies must disprove. The Supreme Court seems to agree that no further interpretation is needed, as Harkonen's petition for certiorari was recently denied.<sup>141</sup>

With that in mind, both the FDA and society at large recognize times when these standards should be bent. As discussed above, for instance, the FDA has recognized that consumer demand for potentially unsafe, potentially ineffective drugs may be justified in situations where there is no other hope.<sup>142</sup> What about other scenarios when, for example, statistical significance testing may be next to impossible, or post hoc analysis is the only kind of analysis available?

#### B. STATISTICAL SIGNIFICANCE SHOULD BE WEIGHTED LIGHTLY IN THE APPROVAL OF PERSONALIZED MEDICINE

Personalized medicine is an up-and-coming field of treatment in which drugs are narrowly tailored to specific populations, or even individuals.<sup>143</sup> For example, one drug might aim at reducing blood pressure in only African-American diabetic patients because it targets a specific characteristic unique to that population. The ultimate aim of personalized medicine is to create treatment plans unique to an individual, tailoring a drug and its dosage not just to the disease, but to the disease and patient.<sup>144</sup>

The problem with requiring a showing of statistical significance for approval in such scenarios lies in the target populations' small sample sizes. The FDA has warned that tests with small sample sizes often lack the power to show significant clinical effect.<sup>145</sup> As treatments shift to the individual level, it be-

---

141. Harkonen v. United States, 134 S. Ct. 824 (2013) (mem.).

142. See *supra* note 103 and accompanying text.

143. See FDA, *supra* note 103, at 2.

144. See *id.*

145. *Id.* at 34.

comes impossible for drug companies to design studies that may predict effectiveness in a patient—instead of extrapolating those effects from other studies—without testing each individual.<sup>146</sup>

Individualized medicine, then, marks a boundary at which statistical significance is no longer useful in the way that the FDA currently uses it to approve drugs.<sup>147</sup> For the purposes of individualized medicine, post hoc analyses currently not allowed in approval of drugs may be useful in approving personalized drugs, simply because there is little to no other option to establish statistical significance.

While post hoc analyses are in danger of abuse by manufacturers, this policy concern is greatly lessened in the sphere of personalized medicine because of the great differential in potential profits a company may recoup from a personalized medicine approval.<sup>148</sup> For example, a late-stage pancreatic cancer medicine that is found to be ineffective to the population at large but is found to be effective in a post hoc analysis for 0.5 percent of the population which carries a certain recessive genetic trait should probably be approved for that sub-group. The manufacturer's incentive to manipulate the data may still be present, but the reward, being much smaller, makes doing so less tempting. That analysis, when considered with other factors like the availability of other effective medicine and the severity of the illness, should take priority in approving individualized medicine.

### C. GUIDANCE: ADVANTAGES AND PITFALLS

Issuing a guidance document to delineate the purpose of p-values may, at first, seem like an attractive proposition. The FDA regularly issues guidance documents to aid drug and device developers in conforming to the FDA's requirements.<sup>149</sup> In such documents, however, the FDA has been careful not to deal in absolutes when discussing the use of p-values in determining

---

146. *Id.* at 30.

147. The approval process generally is not particularly useful for personalized medicine, which, by its nature, produces drugs that require approval only once an individual requires them; a near-record time for approval is roughly 3.6 months. *Id.* at 35. Despite that, it is clear that some sort of approval process is still necessary. *See* Doheny, *supra* note 123.

148. *Cf.* FDA, *supra* note 103, at 2 (discussing how personalized medicine markets to a small population or to an individual).

149. *See generally* *Clinical Trials Guidance Documents*, FDA (Dec. 24, 2014), <http://www.fda.gov/RegulatoryInformation/Guidances/ucm122046.htm>.

drug safety<sup>150</sup> and efficacy.<sup>151</sup> This inclusive approach comports with its stance on safety, but seems to run afoul of its seemingly religious adherence to statistical significance in showing efficacy. Instead of providing bright-line rules for the use of p-values in efficacy, which would doubtlessly be useful for the efficiency of the drug approval system, the FDA has historically elected to be flexible in its approach.<sup>152</sup>

### 1. Safety Above All Else

The FDA has openly stated that p-values may sometimes be valueless in the determination of a drug's safety: in its guidance on statistical principles, the Agency determined that p-values may oftentimes be considerably imprecise when dealing with low-frequency occurrences, such as in the context of side effects.<sup>153</sup> At the same time, the FDA states that p-values can be useful in evaluating safety claims, particularly with large amounts of laboratory data.<sup>154</sup>

What the Agency's guidance does not state is as important as what it does. For example, while the guidance discusses how to treat p-values—useful or useless, depending on the situation—it does not discuss all the methods of identifying safety problems with a developing drug, though it does try to provide helpful hints to researchers seeking to construct an informative trial.<sup>155</sup> This all-inclusive approach in its guidance is consistent with the FDA's overall approach to safety.

### 2. Efficacy and Flexibility

On the other hand, the lack of firm guidance with regards to statistical significance and efficacy is, at first, more difficult to explain. While the FDA normally requires two “adequate and well-controlled individual trials,”<sup>156</sup> it has noted a number of exceptions that may allow a single drug study to support an effec-

---

150. See Harmonisation, *supra* note 49, at 49,596.

151. Cf. FDA *supra* note 5, at 15 (noting that while a “statistically very persuasive finding” may be enough to allow for a single trial, rather than the standard double trial, such findings have occasionally been shown to be false positives).

152. See *id.* at 3 (“[The] FDA has been flexible within the limits imposed by the congressional scheme, broadly interpreting the statutory requirements to the extent possible where the data on a particular drug were convincing.”).

153. See Harmonisation, *supra* note 49, at 49,596.

154. *Id.*

155. *Id.* at 49,595–96.

156. *Id.*

tiveness claim.<sup>157</sup> While one of these exceptions is a particularly persuasive p-value, it is crucial to note that the “Agency has not comprehensively described the situations in which a single adequate and well-controlled study might be considered adequate support for an effectiveness claim,” and that not all of the exceptions include a persuasive p-value.<sup>158</sup> The FDA recognizes that, while statistical significance is an exceptionally useful argument for a drug’s efficacy in a single trial, rather than a double, it need not be a necessity in a drug’s approval. Statistical significance remains, for the most part, a de facto requirement, even if not de jure.<sup>159</sup>

As discussed above, the FDA has already begun to enunciate flexible standards with regards to personalized medicine.<sup>160</sup> The FDA’s firm *practical* stance on requiring statistical significance for drug testing has been effective in screening potentially ineffective drugs from the market, but its loose *guidelines* have allowed it to react dynamically to advancements of science and the dawn of the age of personalized medicine. Calcifying the Agency’s official stance on p-values with regards to efficacy may stifle further scientific development.

#### D. JUDICIAL TAKEAWAY AND PRESUMPTIONS OF SIGNIFICANCE

Courts should be encouraged to look at the public policy behind the FDA’s decisions to treat statistical significance as it does—nearly a necessity in the realm of efficacy, barring some extraordinary circumstances; and as useful, but not dispositive in the realm of safety. Clear guidance from the FDA regarding statistical significance and claims of a drug’s efficacy is not available, nor will it likely be forthcoming due to concerns over the FDA’s flexibility in dealing with future drug development advances. Guidance from the FDA regarding which factors, including p-values, may be dispositive in showing a safety concern will necessarily be vague so as to allow flexibility in finding flaws in treatments through a variety of pathways. Through its decisions and guidance regarding statistical significance in clinical drug trials, however, the FDA has focused on a central public policy goal that courts ought to notice: safety of drugs is paramount.

---

157. FDA, *supra* note 5, at 13–15.

158. *Id.* at 12.

159. See *supra* notes 19–22 and accompanying text.

160. See *supra* note 103; Part III.B.

As such, the courts should favor public safety by adopting presumptions against the safety and efficacy of drugs. If a drug's efficacy is challenged, it should face the high bar of statistical significance currently—if not explicitly—adopted in FDA policy. If a drug's safety is challenged, it should face a similarly high bar of disproving the allegations of harm, or else be forced to disclose those allegations. These presumptions favoring public safety are inherent in the conclusions drawn by the *Matrixx* and *Harkonen* courts and should be used expansively, especially now that the Supreme Court has declined to comment on the supposed “discord.” Some exceptions will certainly arise, such as in the case of personalized medicine, where stringent bars regarding statistical significance may be nigh impossible to overcome. In such cases, courts should look to balance consumers' concerns, such as the lack of alternative treatments for their illnesses, against the threat to their safety, and lower the requisite causative bar accordingly.

Developers of drugs like Actimmune should not be allowed to tell their investors that the drug is effective if the unadorned p-value shows a lack of statistical significance. A lack of statistical significance indicates a lack of efficacy and, thus, an unsafe investment of time and effort for both the investor and the consumer. Developers of drugs like Zicam should not be allowed to leave objectively credible reports that their drug is harming their consumers unreported, even if those events constitute low statistical significance, because a sure showing of causation is of secondary concern in a situation where consumers are being harmed. An analysis similar to that which the FDA undertakes for safety concerns<sup>161</sup> or for efficacy<sup>162</sup> is appropriate in the presence of extraordinary factors. Such analyses should be relatively simple for courts to perform in cases replete with experts in statistics and drug development.

### CONCLUSION

The courts in *Matrixx* and *Harkonen* treated statistical significance differently because of the different contexts in which each was presented: *Matrixx* presented an issue of safety and *Harkonen* one of efficacy. The FDA treats statistical significance with similar context-sensitivity, though it does so implicitly. In the context of safety, statistical significance is relatively

---

161. See FDA, *supra* note 51, at 18.

162. See *supra* Part II.B.

unimportant to the goal of preventing dangerous drugs from entering and remaining on the market, as adverse effects caused by the drug may often manifest at statistically-insignificant but extremely personally-significant levels. In the context of efficacy, statistical significance is of utmost importance because it provides a barrier over which only probably-effective drugs are allowed; stringent statistical cutoffs are required to fulfill the goal of safety and reduce opportunity costs for consumers.

Although statistical significance may play a role in any approval, the FDA uses a flexible approval process for the burgeoning field of personalized medicine, where large sample sizes—and, thus, reliable tests for statistical significance—are impossible. The Agency instead allows responsible post hoc data analysis to serve as acceptable showings of statistical significance for drugs. Courts should follow the FDA's lead in showing flexibility in determining the importance of statistical significance where a showing of such would be impossible. In applying any balancing test, however, courts should keep in mind that public safety is, and should be, the government's primary concern in regulating drugs.