Chapter 30A

Statistical Evidence in Products Liability Litigation

Nathan A. Schachtman

Nathan A. Schachtman, Esq., P.C.
Adjunct Faculty, Columbia Law School

§ 30A:1 Overview

Statistical issues have become important in several areas of the law. The 1960's saw an explosion in the use of statistical inference in discrimination law, deceptive advertising, and trademark infringement cases. The concurrent emergence of epidemiology and mass tort litigation in the 1970s introduced an era in which statistical models of causation would become necessary to the claims and defenses in products liability cases. Beyond epidemiology, statistical issues often dominate the proper interpretation of toxicologic, pharmicokinetic, neuropsychological, and other studies. Statistical issues also have arisen with respect to damages and life tables, and to issues of notice, adequacy, and typicality in class actions.

There are important resources on statistical evidence for the products liability lawyer. The Federal Judicial Center's *Reference Manual for Scientific Evidence* (the "*Reference Manual*")[1] is now in its third edition, as a joint production of the Center and the National Academies of Science. The chapter on statistical evidence[2] in the *Reference Manual* is one of the better chapters, and is an essential reference for statistical issues in the law. Other noteworthy resources for practitioners are the discussions of statistical evidence in two treatises, *Modern Scientific Evidence*,[3] and *The New Wigmore*.[4]

§ 30A:2 Litigation Context of Statistical Issues

Litigating statistical issues presents inherent difficulties for lawyers. The concepts of probability models and statistical inference are foreign to many lawyers and judges, and often require exposition that cannot

1. Federal Judicial Center, REFERENCE MANUAL ON SCIENTIFIC EVIDENCE (3d ed. 2011) [hereinafter Reference Manual], *available at* www.fjc.gov/public/pdf.nsf/lookup/SciMan3D01.pdf/$file/SciMan3D01.pdf.
2. D. Kaye & D. Freedman, *Reference Guide on Statistics, in* Reference Manual, *supra* note 1, at 211.
3. D. Faigman, et al., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY § 23 (2011).
4. D. Kaye, D. Bernstein, and J. Mnookin, THE NEW WIGMORE: A TREATISE ON EVIDENCE—EXPERT EVIDENCE (2d ed. 2011).

be reduced to trial lawyers' "sound bites." Statistical terminology can be misleading; phrases such as "statistical significance," "confidence interval," "effect size," and "random effects" do not have their ordinary language meanings. Courts and lawyers often stumble into misstating the meaning of statistical tests.

§ 30A:3 Qualification of Expert Witnesses Who Give Testimony on Statistical Issues

Most scientifically trained expert witnesses have taken at least one course in statistics. It is common for physicians, psychologists, biologists, epidemiologists, and others, to offer opinions about the interpretation of statistical tests and analyses; it is almost as common for such witnesses to offer incorrect and fallacious interpretations. Trial courts often avoid meaningful gatekeeping of statistical evidentiary opinions by adverting to proxies for reliability, such as qualifications, but proxies are themselves often unreliable indicators.[5] Many reviews of the peer review process suggest that publication is a poor proxy for assuming the reliability of statistical analyses in biomedical publications.[6]

§ 30A:4 Admissibility of Statistical Evidence—Rules 702 and 703

Many key decisions on statistical inference, especially in the field of discrimination law, pre-date the U.S. Supreme Court's decisions in *Daubert v. Merrell Dow Pharmaceuticals*,[7] *General Electric Co. v. Joiner*,[8] and *Kumho Tire Co. v. Carmichael*,[9] and the 2000 revision to Federal Rule of Evidence 702. For instance, in *Bazemore v. Friday*,[10] the defendant criticized plaintiffs' regression analysis on grounds that it omitted variables for major factors necessary to a fair, sensible model of salary. The Fourth Circuit had treated the omissions as fatal,

---

5. ATA Airlines, Inc. v. Fed. Express Corp., 665 F.3d 882, 888–89 (2011) (chastising trial court, and counsel, for failing to scrutinize invalid regression analysis). *See also* S. Greenland & C. Poole, *Problems in Common Interpretations of Statistics in Scientific Articles, Expert Reports, and Testimony*, 51 JURIMETRICS J. 113, 126 (2011) (describing misstatement of meaning of p-value by an expert witness, full professor of biostatistics, in *Neurontin* litigation).
6. D. Altman, *Poor-Quality Medical Research: What Can Journals Do?*, 287 J. AM. MED. ASS'N 2765 (2002).
7. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579 (1993).
8. Gen. Elec. Co. v. Joiner, 522 U.S. 136 (1997).
9. Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999).
10. Bazemore v. Friday, 751 F.2d 662, 672 (1984), *rev'd*, 478 U.S. 385 (1986).

but the Supreme Court excused the omissions and declared that "[n]ormally, failure to include variables will affect the analysis' probativeness, not its admissibility."[11] In a footnote, the Supreme Court acknowledged that there may be "some regressions so incomplete as to be inadmissible as irrelevant," but rejected the defendant's argument that plaintiffs' regression was fatally incomplete.[12] *Bazemore* has fostered a resistance to gatekeeping of some statistical evidence, which has survived *Daubert*. The better reasoned cases,[13] however, fully apply the principles of Rule 702 to statistical inference and analyses. Considering the prevalence of poorly conducted and interpreted statistical analyses in peer-reviewed publications, courts should stop relying upon peer review as a proxy for validity in statistical analysis.[14]

The warrant for examining the integrity of data relied upon by expert witnesses appears to be securely embedded in Rule 703, which rule has particular relevance to statistical testimony. Lawyers facing studies of dubious quality may need to press for discovery of underlying data and materials. In the *Viagra* vision loss multi-district litigation (MDL), the defendant sought and obtained discovery of underlying data from plaintiffs' expert witness's epidemiologic study of vision loss among patients using Viagra and similar medications.[15] Although the MDL court had struggled with inferential statistics, it understood the challenge based upon lack of data integrity, and reconsidered and granted defendant's motion to exclude the challenged expert witness.[16] The lawyering implications are fairly obvious. Statistical evidence requires counsel's special scrutiny to ensure compliance with the disclosure requirements of Federal Rule of Civil Procedure 26. Lawyers should request all computer runs,

---

11. 478 U.S. at 400.
12. *Id*. at 400 n.10.
13. ATA Airlines, Inc. v. Fed. Express Corp., 665 F.3d 882, 888–89 (2011) (Posner, J.) (reversing on grounds that plaintiff's regression analysis should never have been admitted); Munoz v. Orr, 200 F.3d 291 (5th Cir. 2000).
14. *See, e.g.*, D. Altman, *Poor-Quality Medical Research: What Can Journals Do?*, 287 J. AM. MED. ASS'N 2765 (2002).
15. *In re* Viagra Prods. Liab. Litig., 572 F. Supp. 2d 1071, 1090 (D. Minn. 2008).
16. *In re* Viagra Prods. Liab. Litig., 658 F. Supp. 2d 936, 945 (D. Minn. 2009). In other products litigation, the results from discovery of underlying data have been important even if less dramatic. In the welding litigation, access to underlying data from a study coordinated by plaintiffs' counsel led to that study's loss of strategic importance to the plaintiffs. *See In re* Welding Fume Prods. Liab. Litig., MDL 1535, 2005 WL 5417815 (N.D. Ohio Oct. 18, 2005) (upholding defendants' subpoena for documents and things from Dr. Racette author of study on welding and parkinsonism, conducted with assistance of plaintiffs' counsel); *id*., Mem. & Order (Mar. 6, 2006) (denying requested deposition, but requiring Dr. Racette to answer defendants' interrogatories).

programming[17] routines, and outputs, and they should zealously pursue witnesses' failure to maintain and produce data.

§ 30A:5 Significance Probability

§ 30A:5.1 *Definition of Significance Probability (The "p-value")*

Significance probability, the familiar p-value, is a conditional probability; it represents the probability that observed data would be as different from expected as they are, or more so, assuming that there really is no difference. This probability turns on the choice of probability model, an assumption of randomness in sampling, along with the assumption of the null hypothesis. Statisticians call an observed disparity "significant" when the p-value is small. The smaller the p-value, the stronger the evidence is against the null hypothesis. A p-value of no more than 5% is often used to make the call of significance, but frequently smaller p-values are used.

§ 30A:5.2 *The Transposition Fallacy*

The conditional nature of significance probability leads many judges and lawyers to confuse significance probability with the probability used to describe the burden of proof. Jurors in evaluating the entire evidentiary display of a trial are asked to determine the probability of the claim given the evidence. The order of claim given evidence, versus evidence given claim, in these probability statements makes a difference, as can be seen from a simple example. The probability that people have some training in the law given that they are reading this chapter is high, perhaps better than 90%. The probability that people are reading this chapter given that they have training in the law is, regrettably, low, perhaps approaching zero. Although we can hope that the next edition will do better, most persons with legal training will never read this chapter.

There is no easy way to derive the burden-of-proof probability from the significance probability. Confusing and transposing the conditions occurs so frequently that the error has acquired a name—the "transposition fallacy." Despite the best efforts of the *Reference Manual*, this fallacy remains all too prevalent in the reported judicial decisions and in lawyers' examinations and briefs.

---

17. *See* Barnes v. Dist. of Columbia, 289 F.R.D. 1, 19-24 (D.D.C. 2012) (ordering production of underlying data and information because, "[i]n order for the [requesting party] to understand fully the . . . [r]eports, they need to have all the underlying data and information on how" the reports were prepared).

Consider a decision that affected several thousand Avandia product liability claims, aggregated for pretrial handling in a federal MDL. The federal judge, in ruling on defendant's Rule 702 motion, described a clinical trial known as DREAM, with a risk ratio greater than 1.0, with a p-value of 0.08, as follows:

> "[T]here was a trend towards an adverse outcome for Avandia users (e.g., in DREAM, the p-value was .08, which means that there is a 92% likelihood that the difference between the two groups was not the result of mere chance)."[18]

The 8% represents the significance probability of having obtained data at least as different from expected as observed, assuming that there really was no difference; the 8% does ***not*** represent the probability that there is no difference between observed and expected; nor does the complement of 8%,(1 – 0.08) or 92%, represent the probability that the observed difference was not the result of chance.

§ 30A:5.3 *Confusion Between Significance Probability and The Burden of Proof*

Some lawyers complain that the traditional significance level of $p < 5\%$ imposes a change in the burden of proof in a civil trial. To the extent this complaint invokes the transpositional fallacy, the complaint is demonstrably wrong. One MDL court, however, which committed the transpositional fallacy,[19] then fallaciously argued that the use of a critical value of less than 5% significance probability increased the "more likely than not" burden of proof upon civil litigants.[20] The court not only ignored the conditional nature of significance probability, but it over-interpreted the role of significance testing in arriving at a conclusion of causality. Significance probability addresses only the role of random error in a single sample estimate of a measure of interest. Significance probability does not quantify the potential role of bias or confounding, both of which must be ruled out in assessing causality. Furthermore, causal assessments require a

---

18. *In re* Avandia Mktg., Sales Practices & Prods. Liab. Litig., 2011 WL 13576, at *12 (E.D. Pa. Jan. 4, 2011). For an ever-increasing list of judicial decisions, regulations, and references that commit the transpositional fallacy, see N. Schachtman, Scientific Illiteracy Among the Judiciary (Feb. 29, 2012), *available at* http://schachtmanlaw.com/scientific-illiteracy-among-the-judiciary.
19. *In re* Ephedra Prods. Liab. Litig., 393 F. Supp. 2d 181, 191 (S.D.N.Y. 2005).
20. *Id*. at 188, 193. *See* MICHAEL O. FINKELSTEIN, BASIC CONCEPTS OF PROBABILITY AND STATISTICS IN THE LAW 65 (2009) (criticizing the *Ephedra* decision).

synthesis of information from multiple studies, of various kinds, and rarely can be reduced to a quantitative probability statement.

Even if the p-value from a single study could be turned into a probability statement of the correctness of the null hypothesis, there would be many other probabilities that would necessarily diminish that probability. Some of the other factors (which could be expressed as objective or subjective probabilities) include:

- accuracy of probability distribution used to model observed data
- accuracy of the data reporting
- data collection
- data categorization
- data cleaning
- data handling
- data analysis
- internal validity of the study
- external validity of the study
- credibility of study participants
- credibility of study researchers
- credibility of the study authors
- accuracy of the study authors' expression of their research
- accuracy of the editing process
- accuracy of the testifying expert witness's interpretation
- credibility of the testifying expert witness
- other available studies, and their respective data and analysis factors
- additional Bradford Hill factors for evaluating causality

If these largely independent factors each had a probability or accuracy of 95%, which would be remarkably high, the conjunction of their probabilities would put the probability that the result was the "true value" well below 50%. The apparent rigor of the usual cut-off of statistical significance is misleading; significance probability standards do not subvert the usual standards of proof in civil cases.[21]

---

21. *See also* S. Greenland, *Null Misinterpretation in Statistical Testing and Its Impact on Health Risk Assessment*, 53 PREVENTIVE MED. 225 (2011).

§ 30A:5.4 Hypothesis Testing

Type I error is the probability of observing a difference between observed and expected values when there is none. Statisticians refer to type I error as alpha (α). In order to avoid false-positive results, statisticians usually pre-specify that α not exceed 5%, before the assumption of no difference is rejected. If the significance probability exceeds 5%, the result is inconclusive; we cannot accept the null hypothesis of no difference; we merely do not reject it. This process is typically referred to as "hypothesis testing." The decision options presented are either reject the null hypothesis or fail to reject the null hypothesis.

Hypothesis testing assesses the play of chance ("random error") in the observed data. Such testing does not analyze the role of bias or confounding, which may well have a greater role than random error in influencing the observed data.

§ 30A:5.5 Confidence Intervals

A confidence interval is based upon the sample estimate from a study, with the range of specified standard error above and below the estimate, to provide a sense of the extent of random error, and the range of population values that are compatible with the sample estimate. Confidence intervals add considerably to the understanding of just the sample estimate with a p-value, although an interval can be constructed from both pieces of data. The level of "confidence" refers to the measure of standard error that surrounds the study's sample (or point) estimate. The coefficient of confidence is thus the complement of α (that is, 1−α). If a study's pre-specified cut-off for statistical significance was 5%, two tailed, then the corresponding 95% confidence interval may be derived by taking the sample estimate ± 1.96 standard errors, for data that are approximately normally distributed.

Like statistical significance, confidence intervals are subject to misunderstanding and misinterpretation. Judges and lawyers seem congenitally unable to avoid interpreting the 95% of the 95% confidence interval as a probability that the result is correct, or that the correct result is within the calculated interval.[22] Neither interpretation

---

22. *See, e.g.*, Turpin v. Merrell Dow Pharm., Inc., 959 F.2d 1349, 1353–54 & n.1 (6th Cir. 1992) (erroneously describing a 95% CI of 0.8 to 3.10, to mean that "random repetition of the study should produce, 95 percent of the time, a relative risk somewhere between 0.8 and 3.10"); DeLuca v. Merrell Dow Pharm., Inc., 791 F. Supp. 1042, 1046 (D.N.J. 1992) (incorrectly stating that a "95% confidence interval means that there is a 95% probability that the 'true' relative risk falls within the interval"), *aff'd*,

is correct.[23] The 95% probability involved is the frequency with which a large number of repeated samples, and corresponding intervals, will include the true population value. With respect to a single interval from a single study estimate before us, we cannot make a probability statement with respect to the true population value; to do so, would be to commit the transpositional fallacy in another form. Similarly, the confidence interval addresses only random error; it does not tell us anything about the presence, absence, or magnitude of bias and confounding.[24]

§ 30A:5.6 *Inappropriate Use of Statistical Significance*—Matrixx Initiatives, Inc. v. Siracusano

In *Matrixx Initiatives, Inc. v. Siracusano*,[25] the Supreme Court addressed a securities fraud class action against an over-the-counter pharmaceutical company for speaking to the market about its rosy financial projections, but failing to provide information in its possession about the hazards of the product, Zicam. The MDL trial court granted defendant's motion to dismiss for failure to state a claim, but the Ninth Circuit reversed and reinstated the complaint. The Supreme Court unanimously affirmed the Ninth Circuit.

The company's challenge to the pleadings was based upon the class's failure to plead statistical significance of the relationship between Zicam and the harm alleged—the loss of the sense of smell (anosmia). Matrixx argued that the required element of materiality implied that the alleged facts must support a scientific conclusion of causation. Matrixx interpreted statistical significance as the *sine qua non* of causality, and so the company argued that materiality can thus be reduced to a bright-line test that requires allegations of "statistical significance." The company interpreted the complaint to assert that the key non-disclosed information was a set of adverse event reports, and because these reports were not statistically significant, but merely

---

6 F.3d 778 (3d Cir. 1993); Eli Lilly & Co. v. Teva Pharm., USA, 2008 WL 2410420, at *24 (S.D. Ind. June 11, 2008) (stating incorrectly that "95% percent of the time, the true mean value will be contained within the lower and upper limits of the confidence interval range").

23. D. Kaye, *Is Proof of Statistical Significance Relevant?*, 61 WASH. L. REV. 1333, 1354 (1986) (suggesting that "confidence interval" should be renamed "frequency coefficient" in the courtroom to prevent confusion).

24. *See* Brock v. Merrill Dow Pharm., Inc., 874 F.2d 307, 311–12 (5th Cir. 1989) (incorrectly stating that the court need not resolve questions of bias and confounding because "the studies presented to us incorporate the possibility of these factors by the use of a confidence interval").

25. Matrixx Initiatives, Inc. v. Siracusano, 131 S. Ct. 1309 (2011).

isolated, hearsay, anecdotal reports,[26] then the class's complaint failed to state a cognizable claim.

The Supreme Court had little difficulty in dispatching the defense argument. Allegations of a causal relationship between Zicam and anosmia were not necessary to show the materiality of the non-disclosure. The company's bullish sales projections could falter from any adverse FDA action taken against the drug, on an evidentiary display that fell far short of showing causation.[27] The Court correctly noted that the FDA does not require proof of causation, but rather the agency may require warnings on the basis of reasonable evidence of an association.[28] The company's boastful sales predictions made while withholding information that would lead to regulatory intervention, although insufficient to support causal conclusions, easily met the materiality requirement for a securities fraud action under Rule 10(b)(5).

The holding is an unexceptional application of settled principles of securities fraud litigation in the context of claims against a pharmaceutical company with products liability cases pending. Because causation was not necessary to materiality, the company's claim that the alleged causation required allegations of statistical significance, turned out to be irrelevant to the Court's holding.

Matrixx Initiative's attempt to import Rule 702 principles of reliability into a motion to dismiss on the pleadings was seriously misguided. First, as noted, even assuming that statistical significance was necessary to causation, regulatory action did not require a showing of causality. Therefore, statistical significance was never necessary for the plaintiffs' case. Second, the company's argument that the adverse event reports at issue were "not statistically significant" was fallacious because adverse event reports, standing alone, could not be "statistically significant" or "insignificant."[29] The company would

---

26. *Id.* at 1313–14.
27. *See* FDA Regulations, 21 U.S.C. § 355(d), (e) (requiring drug sponsor to show adequate testing, labeling, safety, and efficacy); *see also* 21 C.F.R. § 201.57(e) (requiring warnings in labeling "as there is reasonable evidence of an association of a serious hazard with a drug; a causal relationship need not have been proved."); 21 C.F.R. § 803.3 (adverse event reports address events *possibly* related to the drug or the device); 21 C.F.R. § 803.16 (adverse event report is not an admission of causation).
28. *Matrixx Initiatives*, 131 S. Ct. at 1320.
29. *See* N. Schachtman, The Matrixx—A Comedy of Errors (Apr. 6, 2011), *available at* http://schachtmanlaw.com/the-matrixx-%E2%80%93-a-comedy-of-errors/; D. Kaye, *Trapped in the Matrixx: The U.S. Supreme Court and the Need for Statistical Significance*, BNA PROD. SAFETY & LIAB. RPTR. 1007 (Sept. 12, 2011).

need to know the expected base rate for anosmia among Zicam users, and it would need to frame the adverse event reports in terms of an observed rate, so that the expected and observed rates could be compared against an assumption of no difference. Third, the class plaintiffs had alleged considerably more than just the adverse events, and the allegations taken together might well deserve the attention of a reasonable investor.

Matrixx was a pleadings case. No evidence was ever offered; nor was there any ruling on the reliability or insufficiency of evidence of causation. A unanimous Supreme Court held that because FDA regulatory action does not require evidence that reliably supports a causal conclusion, pleading materiality for a securities fraud suit does not require an allegation of causation, and thus does not require an allegation of statistically significant evidence. The Court, however, went well beyond its holding to suggest that courts "frequently permit expert testimony on causation based on evidence other than statistical significance."[30] The Court's dictum was remarkable. As an initial matter, the Court failed to take notice of its own general causation decision in *Joiner*,[31] disapproving of expert witnesses' reliance upon studies that failed to reach statistically significant results.

Second, in stretching to identify cases that upheld findings of causation without statistical significance, the Court cited three cases, two of which involved differential diagnosis.[32] These two cases were irrelevant to the point because the differential diagnoses, really differential etiologies, were about specific causation. Such cases presuppose that the exposure or drug in question can cause the outcome, and proceed to inquire, by a process of elimination, what was the cause in an individual patient. Statistical significance never comes into play in such cases because general causation is already shown.[33]

Third, the remaining case cited by the Court, for the proposition that statistical significance was unnecessary, was a pre-*Daubert* case, *Wells v. Ortho Pharmaceutical Corp*.[34] *Wells* involved a claim that the use of spermicidal jelly contraceptive caused a child's birth defects. At least one study relied upon by the *Wells* plaintiffs reported a statistically significant increase in detected birth defects over the expected

---

30. *Matrixx Initiatives*, 131 S. Ct. at 1319.
31. Gen. Elec. Co. v. Joiner, 522 U.S. 136, 145 (1997).
32. *Matrixx Initiatives*, 131 S. Ct. at 1319, citing Best v. Lowe's Home Centers, Inc., 563 F.3d 171, 178 (6th Cir. 2009); Westberry v. Gislaved Gummi AB, 178 F.3d 257, 263–64 (4th Cir. 1999).
33. Raynor v. Merrell Pharm. Inc., 104 F.3d 1371, 1376 (D.C. Cir. 1997).
34. Wells v. Ortho Pharm. Corp., 788 F.2d 741, 744–745 (11th Cir. 1986).

rate.[35] *Wells*, therefore, cannot be an example of a case in which an expert witness properly opined about causation in the absence of a scientific study with statistical significance. Of course, finding statistical significance is just the beginning of assessing the causality of an association; the *Wells* case was and remains notorious for the expert witness's failure to rule out bias and confounding as the source of the apparent increased risk of birth defects, which is found in some but not all of the relied upon studies.

The *Wells* decision was received with severe criticism in the 1980s. Scientists and scholars noted that the decision failed to evaluate the entire evidentiary display, and failed to address study bias and confounding.[36] A few years later, another case in the same judicial district, against the same defendant, for the same product, with allegations of a birth defect, resulted in the grant of summary judgment.[37] *Wells* was tried to the bench, but it has come to epitomize "anything goes" in admitting expert witness testimony, and was thus overruled *sub silentio* by the Supreme Court's own decisions in *Daubert* and *Joiner*. Indeed, even these two decisions turned on a statute, Federal Rule of Evidence 702, which was subsequently and substantially revised to support gatekeeping.

[A] Sequelae of *Matrixx Initiatives*

To be sure, the Court stated that it "need not consider whether the expert testimony was properly admitted in [the three cases cited for dispensing with statistical significance], and we do not attempt to define here what constitutes reliable evidence of causation."[38] Furthermore, taken at face value, the Court's statement that "the premise that statistical significance is the only reliable indication of causation . . . is flawed,"[39] is unexceptionable. Surely, there are such cases; however, in modern products liability law, many causation puzzles are based upon the interpretation of rate-driven processes, measured using epidemiologic studies, involving a base-line risk and an observed higher or lower

---

35. Wells v. Ortho Pharm. Corp., 615 F. Supp. 262 (N.D. Ga. 1985), *aff'd, and rev'd in part on other grounds*, 788 F.2d 741 (11th Cir.), *cert. denied*, 479 U.S. 950 (1986).
36. *See, e.g.*, J. Mills & D. Alexander, *Teratogens and 'Litogens'*, 15 NEW ENGL. J. MED. 1234 (1986); S. Gross, *Expert Evidence*, 1991 WIS. L. REV. 1113, 1121–24 (1991) ("Unfortunately, Judge Shoob's decision is absolutely wrong. There is no scientifically credible evidence that Ortho-Gynol Contraceptive Jelly ever causes birth defects.").
37. Smith v. Ortho Pharm. Corp., 770 F. Supp. 1561 (N.D. Ga. 1991) (supposedly distinguishing *Wells* on the basis of more recent studies).
38. *Matrixx Initiatives*, 131 S. Ct. at 1319.
39. *Id*.

risk among a sample of an exposed population. In this context, some evaluation of the size of random error is, indeed, necessary.

Predictably, lower courts have been led astray by the dicta in *Matrixx Initiatives*. The MDL court in the *Chantix* litigation provides one such instance. Plaintiffs claimed that Chantix, which is medication that helps people stop smoking, causes suicide. Pfizer, the manufacturer, challenged plaintiffs' general causation expert witnesses for not meeting the standards of Federal Rule of Evidence 702, for various reasons, not the least of which was that the studies relied upon by plaintiffs' witnesses did not show statistical significance.[40] The *Chantix* MDL court, citing *Matrixx Initiatives* for a blanket rejection of the need to consider random error, denied the defendant's challenge.[41] The Supreme Court, however, never stated or implied such a blanket rejection of the importance of considering random error in evidence that was essentially statistical in nature.

Within two weeks of the *Chantix* decision, a similar erroneous interpretation of *Matrixx Initiatives* surfaced in MDL litigation over fenfluramine.[42] Rejecting a Rule 702 challenge to plaintiffs' expert witness's opinion, the MDL trial judge cited *Matrixx Initiatives* for the assertion that "*Daubert* does not require that an expert opinion regarding causation be based on statistical evidence in order to be reliable. * * * In fact, many courts have recognized that medical professionals often base their opinions on data other than statistical evidence from controlled clinical trials or epidemiological studies."[43] While some causation opinions might be perfectly appropriately based upon other than statistical evidence, the Supreme Court specifically disclaimed any comment upon Rule 702, in *Matrixx Initiatives*, which was a case about proper pleading in a securities fraud case, not about proper foundations for actual evidence of causation, at trial.[44]

[B] Is Statistical Significance Necessary?

Declaring that statistical significance is not a litmus test has become a judicial cliché.[45] The cliché is a half truth, swallowed whole

---

40. *In re* Chantix Prods. Liab. Litig., MDL 2092, 889 F. Supp. 2d 1272 (N.D. Ala. 2012).
41. *Id*. at 1286 (citing *Matrixx Initiatives*, 131 S. Ct. at 1319).
42. *See* Cheek v. Wyeth Pharm. Inc., 890 F. Supp. 2d 552, 561 (E.D. Pa. 2012).
43. *Id*. at *22 (citing *Matrixx Initiatives*, 131 S. Ct. at 1319, 1320).
44. *See also* Johns v. Bayer Corp., 2013 WL 1498965, *25 (S.D. Cal. 2013) (citing and quoting *Matrixx Initiatives*, for the proposition that statistical significance was unncessary for inferring causality, without noting that the proposition was dictum).
45. Cook v. Rockwell Int'l Corp., 580 F. Supp. 2d 1071, 1099 n.25 (D. Colo. 2006), *rev'd and remanded on other grounds*, 618 F.3d 1127 (10th Cir. 2010), *cert. denied*, __ U.S. __ , 2012 WL 2368857 (June 25, 2012).

in too many cases. Of course, statistical significance is not practical significance. For instance, a study may find that a medication reduces blood pressure 1mmHg compared with placebo, $p < 0.05$, but most clinicians would not be impressed. Statistical significance is not a bright-line test for many reasons, not the least of which is that p-value addresses only the role of random error. Even a pre-specification of α at no more than 5% can be too high when the study engages in multiple comparisons that dilute the protection against false-positive results.[46] Similarly, the inducements to publish results with $p < 0.05$ are great, and so publication bias obscures contrary research that does not see the light of day. Furthermore, no one would take the difference between a "statistically significant" p-value of 0.049, and a p-value of 0.051, to be particularly meaningful. Where courts have gone astray is equating the lack of an exact point of demarcation with a lack of importance to the assessment of random error. For instance, in the *Viagra* vision-loss litigation, the MDL court denied Rule 702 challenges to plaintiffs' expert witnesses, based in part upon their reliance on epidemiologic studies, the results of which lacked statistical significance.[47] In adopting this dismissive approach, the MDL court failed to consider the width of the available confidence intervals, or to make any consideration of the magnitude of random error in the studies involved.

§ 30A:6 Statistical Power

§ 30A:6.1 Definition of Statistical Power

A type II error results from the failure to reject the null hypothesis (usually of "no difference") when the null hypothesis is incorrect. In lay terms, a type II error is a "missed opportunity" to identify a statistically significant result. Statisticians refer to the probability of making a Type II error as β. Power, the complement of β (that is, $1 - \beta$), is the probability of obtaining a statistically significant result, at a given level of α, and for an expected variance, different from the null, of a specified magnitude.

Power is almost exclusively an *a priori* concept used to assess the appropriateness of sample size before a study is conducted. After a study has been conducted and the data have been analyzed, statisticians generally use the standard error, or an appropriately calculated

---

46. Reference Manual, *supra* note 1, at 256–57.
47. *In re* Viagra Prods. Liab. Litig., 572 F. Supp. 2d 1071, 1081, 1090 (D. Minn. 2008) (citing *In re* Phenylpropanolamine (PPA) Prods. Liab. Litig., MDL No. 1407, 289 F. Supp. 2d 1230, 1241 (W.D. Wash. 2003), as "persuasive authority" for the notion that the lack of statistical significance did not detract from the reliability of a study).

confidence interval, to assess the range of values reasonably compatible with the study's sample estimate.[48]

The power of a study depends upon several variables, including the size of the alternative hypothesis, the sample size, and the acceptable level of probability for false-positive findings, which level is reflected in the pre-specified p-value, α, at which level the study's findings would be interpreted as not likely consistent with the null hypothesis. The lower α is set, the lower will be the power of a test or a study, all other things being equal. Similarly, moving from a two-tailed to a one-tailed test of significance will increase power. Courts have acknowledged that both Type I and Type II errors, and the corresponding α and β, are important, but they have overlooked that Type II errors are usually less relevant to the litigation process.[49] A single study that failed to show a statistically significant difference in the outcome of interest does not support a conclusion that the outcome is not causally related to the exposure under study. In products liability litigation, the parties are typically not assigned a burden of proving the absence of causation.

A claim that a study has low power is meaningless unless both the alternative hypothesis and the level of significance are included in the statement.[50] Power can always be assessed as low by selecting an alternative hypothesis sufficiently close to the null. A study, using risk ratios, which has high power against an alternative hypothesis of 2.0, may have very low power against an alternative of 1.1. Because risk ratios greater than two are often used to attribute specific causation, measuring power of a study against an alternative hypothesis

---

48. David R. Cox & Christl Donnelly, PRINCIPLES OF APPLIED STATISTICS 25–26 (2011); S. Senn, *Power is Indeed Irrelevant in Interpreting Completed Studies*, 325 BRIT. MED. J. 1304 (2002); Jan Vandenbroucke, *et al.*, "Strengthening the reporting of observational studies in epidemiology (STROBE): Explanation and elaboration," 18 EPIDEMIOL. 805, 815 (2007) ("Do not bother readers with post hoc justifications for study size or retrospective power calculations"). *See* N. Schachtman, Power in the Courts—Part Two (Jan. 21, 2011), *available at* http://schachtmanlaw.com/power-in-the-courts-part-two/ (collecting references that reject power analysis to evaluate concluded studies). Philosopher of science Deborah Mayo has argued that power may be an appropriate measure of the severity of the test, such that we may infer the correctness of the null hypothesis. This use of power is very different from what plaintiffs' counsel hope to achieve by urging arguments based upon lack of power. *See* D. Mayo & A. Spanos, *Severe Testing as a Basic Concept in a Neyman–Pearson Philosophy of Induction*, 57 BRIT. J. PHIL. SCI. 323 (2006).

49. *See, e.g.,* DeLuca v. Merrell Dow Pharms., Inc., 911 F.2d 941, 948 (3d Cir. 1990) Poole

50. *See* Sander Greenland & Charles, *Problems in Common Interpretations of Statistics in Scientific Articles, Expert Reports, and Testimony*, 51 JURIMETRICS J. 113, 121-22 (2011)

of a doubling of risk might well be a reasonable approach in some cases.[51] Unless a court was willing to specify the level at which it would find the risk ratio unhelpful or not probative, power analyses of completed studies are not particularly useful.

Plaintiffs' counsel rightly complain when defendants claim that a study with a statistically "non-significant" risk ratio greater than 1.0 has no probative value. Although random error (or bias and confounding) may account for the increased risk, the risk may be real. If studies consistently show an increased risk, even though all the studies have reported p-values > 5%, meta-analytic approaches may very well help rule out chance as a likely explanation for the increased risk.[52] The complaint that a study, however, is underpowered, without more, does not help plaintiff establish an association; nor does the complaint establish that the study provides no useful information.

§ 30A:6.2 Cases Involving Statistical Power

In *Miller v. Pfizer Inc.*, plaintiffs claimed that Zoloft, an antidepressant, induced their son's suicide. None of the available clinical trials found a statistically significant increased risk of suicide, but plaintiffs contended that these trials lacked power to detect an increased risk. The court appointed two expert witnesses, an epidemiologist and a psychiatrist. The epidemiologist conducted power calculations, which showed that the available studies, and their confidence intervals, were sufficiently large to show a doubling of risk if such an association were present.[53] Adopting the expert witnesses' reports, the trial court excluded plaintiffs' causation witness,[54] and the Tenth Circuit affirmed.[55]

Plaintiffs' key claim in the *Avandia* litigation is that the medication, an oral anti-diabetic, causes heart attacks, even though none of the several dozen clinical trials found a statistically significant increased risk. Plaintiffs' expert witnesses argued that all the clinical trials of Avandia were "underpowered," and thus the failure to find an

---

51. In Miller v. Pfizer, 196 F. Supp. 2d 1062, 1079 (D. Kan. 2002), (acknowledging that most courts require a showing of risk ratio greater than two), *aff'd*, 356 F. 3d 1326 (10th Cir.), *cert. denied,* 543 U.S. 917 (2004), the trial court's Rule 706 expert witness calculated the power of a study to exceed 90% probability to detect a doubling of risk. Report of John Concato, M.D., 2001 WL 1793169, *9 (D.Kan. 2001).
52. For a discussion of meta-analysis, see *infra* section 30A:8.
53. Miller v. Pfizer Inc., 2001 WL 1793169 (D. Kan. Sept. 4, 2001) (reports of Dr. Concato and Dr. Davis).
54. Miller v. Pfizer Inc., 196 F. Supp. 2d 1062 (D. Kan. 2002).
55. Miller v. Pfizer Inc., 356 F.3d 1326 (10th Cir. 2004).

increased risk was a Type II (false-negative) error that resulted from the small size of the clinical trials. The *Avandia* MDL court, considering Rule 702 challenges to plaintiffs' expert witness opinions, accepted this argument:

> "If the sample size is too small to adequately assess whether the substance is associated with the outcome of interest, statisticians say that the study lacks the *power* necessary to test the hypothesis. Plaintiffs' experts argue, among other points, that the RCTs [randomized controlled trials] upon which GSK relies are all underpowered to study cardiac risks."[56]

The *Avandia* MDL court failed to realize that the power argument was empty without a specification of an alternative hypothesis. For instance, in one of the larger trials of Avandia, the risk ratio for heart attack was a statistically non-significant 1.14, with a 95% confidence interval that spanned 0.80 to 1.63.[57] This trial, standing alone, thus had excellent power against an alternative hypothesis that Avandia doubled the risk; such an alternative hypothesis would clearly be rejected based upon the RECORD trial. On the other hand, an alternative hypothesis of 1.2 would not be. The confidence interval, by giving a quantification of random error, conveys results reasonably compatible with the study estimate; the claim of "low power" against an unspecified alternative hypothesis, conveys nothing. More recently, several cases have demonstrated significant judicial confusion over the meaning and utility of statistical power. In a hormone therapy breast cancer case, the Eight Circuit confused power with β, and succumbed to plaintiff's expert witness's argument that he was justified in ignoring several large, well-conducted clinical trials and observational studies because they were "underpowered," without specifying the alternative hypothesis he was using to make his claim:

> "Statistical power is 'the probability of rejecting the null hypothesis in a statistical test when a particular alternative hypothesis happens to be true'. *Merriam—Webster Collegiate Dictionary* 973 (11th ed. 2003). In other words, it is the probability of observing false negatives. Power analysis can be used to calculate the likelihood of accurately measuring a risk that manifests itself at a given frequency in the general population based on the sample

---

56. *In re* Avandia Mktg., Sales Practices & Prods. Liab. Litig., 2011 WL 13576, at *2 (E.D. Pa. 2011) (emphasis in original).
57. P.D. Home, et al., *Rosiglitazone Evaluated for Cardiovascular Outcomes in Oral Agent Combination Therapy for Type 2 Diabetes (RECORD)*, 373 LANCET 2125 (2009).

> size used in a particular study. Such an analysis is distinguishable from determining which study among several is the most reliable for evaluating whether a correlative or even a causal relationship exists between two variables."[58]

The *Kuhn* court's formulation, "in other words," is incorrect. The court's further discussion of "accurately measuring" mischievously confuses one aspect of statistical power concerned with random variability, with study validity. The 8th Circuit's opinion never discusses or discloses what alternative hypothesis the plaintiff's expert witness had in mind when disavowing certain studies as underpowered.

Another MDL court, similarly lost its way under sway of claims that expert witnesses can never say anything more than epidemiologic studies fail to support an association, or that there is an "absence of evidence."[59] In *Cooley v. Lincoln Electic Co.*, the court held that a defense expert witness could not opine that the relevant epidemiologic studies ruled out any risk of Parkinson's disease from welding, or that the risk was at most infintesmally samll. The ruling ignored that most studies found risk ratios below 1.0, and some studies statistically significantly below 1.0. The ruling also ignored that all the relied upon studies provided confidence intervals, which provided a more appropriate assessment of the extent that individual studies were truly exonerative. A later meta-analysis on the issue demonstrated a highly statistically significant decreased risk.[60] Similarly, in the *Chantix* MDL, the court embraced a critique of epidemiologic studies based upon "power," without any consideration of what alternative hypothesis would be considered for assessing that power.[61]

§ 30A:7 Evidentiary Rule of Completeness

Statistical evidence is especially prone to being presented out of context. A point estimate, based upon a sample mean or proportion, if

---

58. Kuhn v. Wyeth, Inc., 686 F.3d 618, 622 n.5 (8th Cir. 2012), *rev'g, In re* Prempro Prods. Liab. Litig., 765 F. Supp. 2d 1113 (W.D. Ark. 2011).
59. Cooley v. Lincoln Elec. Co., 693 F. Supp. 2d 767 (N.D. Ohio 2010) (succumbing to similar arguments about power without any specification of alternative hypotheses, and ignoring confidence intervals).
60. J. Mortimer, A. Borenstein & L. Nelson, *Associations of Welding and Manganese Exposure with Parkinson Disease: Review and Meta-Analysis*, 79 NEUROLOGY 1174 (2012).
61. *In re* Chantix (Varenicline) Prod. Liab. Litig., 889 F. Supp. 2d 1272, 1291 (2012); *see also In re* Neurontin Mktg., Sales Practices & Prods. Liab. Litig., 612 F. Supp. 2d 116, 126 (D. Mass. 2009) ("Oftentimes, epidemiological studies lack the statistical power needed for definitive conclusions, either because they are small or the suspected adverse effect is particularly rare.") (failing to report any power analyses against a specific alternative hypothesis).

from a random, unbiased sample, may be the best estimate of the population mean or proportion, but still the estimate may suffer from significant imprecision because of the sample size. Accordingly, the presentation of a point estimate should be accompanied by a measure of its variability.[62] Similarly, trial courts should be vigilant against permitting a party to use the upper or the lower bound of a confidence interval to paint a misleading picture of "what the evidence shows."[63]

Codified and common law evidentiary principles provide that when a party offers a writing or statement, an adverse party may provide any other part of a writing or statement, which in fairness should be considered at the same time.[64] This rule is based upon the goal of avoiding the creation of misleading impressions that cannot adequately be addressed at a later time in the trial.[65] Statistical evidence deserves the same protection from misleading presentations, whether on direct or cross-examination.

§ 30A:8 Meta-Analysis

§ 30A:8.1 Definition and History of Meta-Analysis

Meta-analysis is a statistical procedure for aggregating data and statistics from individual studies into a single summary estimate of the population measure of interest. The first meta-analysis is typically attributed to Karl Pearson, *circa* 1904, who sought a method to overcome the limitations of small sample size and limited statistical power in individual studies. Statistical methods for meta-analysis, however, did not mature until the 1970s. Even then, the scientific community remained skeptical of, if not outrightly hostile to, meta-analysis. The hostility to meta-analysis, especially in the context of observational epidemiologic studies, was colorfully expressed, as late as the 1990s:

---

62. Federal Judicial Center, REFERENCE MANUAL ON SCIENTIFIC EVIDENCE at 117–18 (2d ed. 2000) ("[w]henever possible, an estimate should be accompanied by its standard error.").
63. Marder v. G.D. Searle & Co., 630 F. Supp. 1087 (D. Md. 1986) ("The upper range of the confidence intervals signify the outer realm of possibilities, and plaintiffs cannot reasonably rely on these numbers as evidence of the probability of a greater than two fold risk. Their argument reaches new heights of speculation and has no scientific basis."), *aff'd mem. on other grounds sub nom*. Wheelahan v. G.D.Searle & Co., 814 F.2d 655 (4th Cir. 1987) (per curiam).
64. Federal Rule of Evidence 106. *See also* United States v. Velasco, 953 F.2d 1467, 1475 (7th Cir.1992) (trial court should consider whether the omitted portion should be offered at the same time to ensure context, and a fair understanding of the meaning).
65. FED. R. EVID. 106, Advisory Committee Notes.

> "Meta-analysis begins with scientific studies. . . . [D]ata from these studies are then run through computer models of bewildering complexity which produce results of implausible precision."[66]

In the last two decades, meta-analysis has emerged as an important technique for addressing random variation in studies, as well as bias and confounding. The methodology of meta-analysis has advanced and matured considerably in this period. Today, thousands of meta-analyses, of both observational and experimental studies, are published each year.[67] The suggestion that meta-analyses are rarely involved in litigation[68] is untenable. Meta-analyses have been involved in close to two dozen pharmaceutical products litigations, several of which were federal MDL cases that encompassed thousands of claims.[69]

§ 30A:8.2 Consensus Statements

Perhaps because of a mistaken belief that meta-analyses are rarely involved in litigation, the *Reference Manual* and many legal treatises pay scant attention to the statistical techniques used to aggregate data from multiple studies. In addition to many textbooks on the subject, consensus guideline papers have been published for meta-analyses of clinical trials, and of observational studies.[70]

§ 30A:8.3 Use of Meta-Analysis in Litigation

Meta-analysis is particularly well suited to remedying the problem posed by multiple inconclusive studies. Aggregating data across studies can help determine whether the failure to achieve statistical significance is the result of small sample size or random error in the individual studies. Aggregation can also reduce the size of random error for the summary estimate of the population value (whether mean

---

66. S. Shapiro, *Meta-Analysis/Smeta-Analysis*, 140 AM. J. EPIDEM. 771, 777 (1994). *See also* A. Feinstein, *Meta-Analysis: Statistical Alchemy for the 21st Century*, 48 J. CLIN. EPIDEM. 71 (1995).
67. M. Finkelstein & B. Levin, *Meta-Analysis of "Sparse" Data: Perspectives from the Avandia Cases*, 52 JURIMETRICS J. 123 (2012) [hereinafter M. Finkelstein & B. Levin], at 124 & n.3.
68. D. FAIGMAN, ET AL., 3 MODERN SCI. EVIDENCE, § 23:14 n.10 (2011) (incorrectly stating that "[c]ourts infrequently confront expert opinions based on a meta-analysis").
69. M. Finkelstein & B. Levin, *supra* note 67, at 124 & n.3.
70. *See* D. Stroup, *et al.*, *Meta-analysis of Observational Studies in Epidemiology: A Proposal for Reporting*, 283 J. AM. MED. ASS'N 2008 (2000) (MOOSE statement); D. Moher, *et al.*, *Improving the Quality of Reports of Meta-analyses of Randomised Controlled Trials: The QUOROM Statement*, 354 LANCET 1896 (1999).

or proportion). Sensitivity analyses within a set of all studies can further help detect associations, biases, and confounding.

The initial deployment of meta-analyses in litigation was met with hostility from bench and bar, which, over time, has turned into uncritical acceptance, regardless of the individual meta-analysis' merit or validity. In litigation over exposure to PCBs (polychlorinated biphenyl), plaintiffs' expert witness offered an unpublished meta-analysis of health outcomes among exposed workers. The trial court upheld defendants' challenge to the proffered testimony on grounds that meta-analysis was a novel technique, and that this particular meta-analysis had not been published.[71] On appeal, the Third Circuit reversed and held that meta-analysis was not novel, and that lack of peer-review was not an automatic disqualification.[72] Acknowledging that meta-analysis could be performed poorly, using invalid methods, the Third Circuit directed the trial court on remand to evaluate the validity of the meta-analysis as performed.

The professional skepticism about meta-analysis was reflected in some of the early judicial assessments of meta-analysis. In the 1980s and early 1990s, some trial judges erroneously dismissed meta-analysis as a flawed statistical procedure that claimed to make something out of nothing.[73] In one of many colorectal cancer asbestos cases, one trial court correctly sensed that plaintiffs' expert witness's back-of-the-envelope, non-quantitative meta-analysis was invalid, but went too far in condemning the entire meta-analytic approach:

> "no matter how many studies yield a positive but statistically insignificant SMR [standardized mortality ratio] for colorectal cancer, the results remain statistically insignificant. Just as adding a series of zeros together yields yet another zero as the product, adding a series of positive but statistically insignificant SMRs together does not produce a statistically significant pattern."[74]

The trial court's analogy to adding zeroes is mathematically incorrect. The Second Circuit reversed the entry of summary judgment, and remanded the colorectal cancer claim for trial.[75] Over a decade later,

---

71. *In re* Paoli R.R. Yard PCB Litig., 706 F. Supp. 358, 373 (E.D. Pa. 1988).
72. *In re* Paoli R.R. Yard PCB Litig., 916 F.2d 829, 856–57 (3d Cir. 1990), *cert. denied,* 499 U.S. 961 (1991); *see also* Hines v. Consol. Rail Corp., 926 F.2d 262, 273 (3d Cir. 1991).
73. Allen v. Int'l Bus. Mach. Corp., No. 94-264-LON, 1997 U.S. Dist. LEXIS 8016, at *71–74 (suggesting that meta-analysis of observational studies was controversial among epidemiologists).
74. *In re* Joint E. & S. Dist. Asbestos Litig., 827 F. Supp. 1014, 1042 (S.D.N.Y. 1993).
75. *In re* Joint E. & S. Dist. Asbestos Litig., 52 F.3d 1124 (2d Cir. 1995).

with many more accumulated studies and data, the Institute of Medicine found, using appropriate meta-analytic techniques, that the evidence for asbestos plaintiffs' colorectal cancer claims was insufficient to show causation.[76] Courts continue to go astray with the erroneous belief that multiple studies, all without statistically significant results, cannot yield a statistically significant summary estimate of increased risk.[77] As noted, the occurrence of multiple inconclusive studies is one of the most important reasons to conduct a meta-analysis.

In the silicone breast implant litigation, one defense expert witness prepared and published a meta-analysis of studies of breast implants and connective tissue diseases, to combat selective, partial presentation of evidence.[78] The MDL court appointed a panel of four distinguished scientists to serve as court-appointed experts. All four rejected the plaintiffs' claims; two of them conducted meta-analyses of the available studies,[79] which they later published.[80] In the welding fume litigation, plaintiffs' expert witness offered a crude, non-quantified meta-analysis to argue that welding causes Parkinson's disease.[81] In rebuttal, one of the defense expert witnesses offered a quantitative meta-analysis, which provided strong evidence against plaintiffs' claim.[82] Although the MDL court excluded the defense expert's meta-analysis from the pre-trial

---

76. Institute of Medicine, Asbestos: Selected Cancers 226 (Wash. D.C. 2006).
77. *See, e.g.*, Baker v. Chevron USA, Inc., 680 F. Supp. 2d 865 (S.D. Ohio 2010) (addressing a meta-analysis on multiple myeloma outcomes in studies of benzene-exposed workers). There were many sound objections to this meta-analysis, but the suggestion that multiple studies without statistical significance could not yield a summary estimate of risk with statistical significance was not one of them.
78. O. Wong, *A Critical Assessment of the Relationship between Silicone Breast Implants and Connective Tissue Diseases*, 23 REGULATORY TOXICOL. & PHARMACOL. 74 (1996).
79. B. Hulka, B. Diamond, N. Kerkvliet & P. Tugwell, *Silicone Breast Implants in Relation to Connective Tissue Diseases and Immunologic Dysfunction: A Report by a National Science Panel to the Hon. Sam Pointer Jr.*, MDL 926 (Nov. 30, 1998), *available at* www.fjc.gov/BREIMLIT/SCIENCE/summary.htm. *See also* B. Hulka, N. Kerkvliet & P. Tugwell, *Experience of a Scientific Panel Formed to Advise the Federal Judiciary on Silicone Breast Implants*, 342 NEW ENG. J. MED. 812 (2000).
80. E. Janowsky, L. Kupper & B. Hulka, *Meta-Analyses of the Relation between Silicone Breast Implants and the Risk of Connective-Tissue Diseases*, 342 NEW ENG. J. MED. 781 (2000); P. Tugwell, et al., *Do Silicone Breast Implants Cause Rheumatologic Disorders?*, 44 ARTHRITIS & RHEUM. 2477 (2001).
81. Deposition of Dr. Juan Sanchez-Ramos, Street v. Lincoln Elec. Co., Case No. 1:06-cv-17026, 2011 WL 6008514 (N.D. Ohio May 17, 2011).
82. Deposition of Dr. James Mortimer, Street v. Lincoln Elec. Co., Case No. 1:06-cv-17026, 2011 WL 6008054 (N.D. Ohio June 29, 2011).

Rule 702 hearing as untimely, plaintiffs' counsel soon thereafter initiated settlement discussions of the entire set of MDL cases. Subsequently, the defense expert witness, with his professional colleagues, published an expanded version of the meta-analysis.[83]

§ 30A:8.4 Competing Models for Meta-Analysis

Meta-analyses typically weight the included studies by the inverse of study variance to arrive at a summary estimate of association. There are many variations on this methodological theme, and many threats to study validity, however, discussion of these is beyond the scope of this chapter.

The key preliminary steps for a meta-analysis include:

(1) State the clinical question and purpose of the meta-analysis;

(2) State inclusionary and exclusionary criteria for selecting the studies;

(3) Identify all eligible studies;

(4) Review the methods and results of all candidate studies;

(5) Abstract and summarize each included study's results;

(6) Apply statistical methods to produce a summary estimate of association;

(7) Assess the variation (heterogeneity) between studies;

(8) Report and interpret the findings; and

(9) Plan for additional research and for updating the meta-analysis.

There are competing statistical models employed in meta-analysis, depending upon whether the included studies purport to estimate a single population value, or whether different population values are estimated by the individual studies. The former type of meta-analysis statistical model is referred to as "fixed effect"; the latter, "random effects." Fixed-effect models weight included studies only by the studies' internal variance. Random-effects models include a term for between-study variance, with the result that large studies may have less weight than they would in a fixed-effect model. In products liability cases, where different studies will likely be examining samples of workers with different levels of exposure, or different dosages of

---

83. J. Mortimer, A. Borenstein & L. Nelson, *Associations of Welding and Manganese Exposure with Parkinson Disease: Review and Meta-Analysis*, 79 NEUROLOGY 1174 (2012).

medication use, in the presence of an hypothesized dose response, random effects models will typically have greater validity.

§ 30A:8.5 Recent Cases Involving Meta-Analyses

Expert witness testimony based upon meta-analysis has been proffered in numerous pharmaceutical cases, including recent cases involving Aredia and Zometa,[84] Avandia,[85] Baycol,[86] benzodiazepine,[87] Celebrex,[88] Fosamax,[89] Gadolinium,[90] Neurontin,[91] Seroquel,[92] Thimerosal,[93] Trasylol,[94] Vioxx,[95] Zoloft,[96] and Zyprexa.[97] The various meta-analyses in these, and other, litigations raise many interesting, complex issues of validity, discussion of which is beyond the scope of this chapter.

A brief discussion of the *Avandia* litigation may illustrate some of the complexities that courts have faced or evaded. The avalanche of filings against the sponsor of Avandia began shortly after the publication of a meta-analysis, in a well-respected clinical journal, by a well-known clinician, Dr. Steven Nissen.[98] Nissen's meta-analysis became the pillar of support for plaintiffs' expert witnesses, and also a target for scholarly criticism. Nissen had used a particular statistical approach—the Peto method for fixed-effect modeling—known to be biased when the arms of clinical trials are of unequal size. Indeed,

---

84. Deutsch v. Novartis Pharm. Corp., 768 F. Supp. 2d 420 (E.D.N.Y. 2011).
85. *In re* Avandia Mktg., Sales Practices & Prods. Liab. Litig., 2011 WL 13576, at *12 (E.D. Pa. 2011); Avon Pension Fund v. GlaxoSmithKline PLC, 343 F. App'x 671 (2d Cir. 2009).
86. *In re* Baycol Prods. Litig., 532 F. Supp. 2d 1029 (D. Minn. 2007).
87. Vinitski v. Adler, 69 Pa. D. & C. 4th 78, 2004 WL 2579288 (Phila. Cnty. Ct. Common Pleas Sept. 17, 2004).
88. *In re* Bextra & Celebrex Mktg. Sales Practices & Prod. Liab. Litig., 524 F. Supp. 2d 1166 (2007).
89. *In re* Fosamax Prods. Liab. Litig., 645 F. Supp. 2d 164 (S.D.N.Y. 2009).
90. *In re* Gadolinium-Based Contrast Agents Prod. Liab. Litig., 2010 WL 1796334 (N.D. Ohio May 4, 2010).
91. *In re* Neurontin Mktg., Sales Practices & Prods. Liab. Litig., 612 F. Supp. 2d 116 (D. Mass. 2009).
92. *In re* Seroquel Prods. Liab. Litig., 2009 WL 3806434, at *5 (M.D. Fla. June 18, 2009).
93. Hennessey v. Sec'y Dep't Health & Human Servs., 2009 WL 1709053 (Fed. Cl. May 29, 2009).
94. *In re* Trasylol Prods. Liab. Litig., 2010 WL 1489793 (S.D. Fla. Feb. 24, 2010).
95. Merck & Co., Inc. v. Ernst, 296 S.W.3d 81 (Tex. Ct. App. 2009).
96. Miller v. Pfizer, Inc., 356 F.3d 1326 (10th Cir. 2004).
97. *In re* Zyprexa Products Liab. Litig., 489 F. Supp. 2d 230 (E.D.N.Y. 2007).
98. S. Nissen & K. Wolski, *Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death From Cardiovascular Causes*, 356 NEW ENG. J. MED. 2457 (2007); Erratum, 357 NEW ENG. J. MED. 100 (2007).

using any fixed-effect approach is problematic when the dosages used in the included trials vary substantially, and the included populations ranged from non-diabetic to pre-diabetic to serious, chronic diabetic patients. No one reasonably expected that there would be one measure of risk across such diverse patient populations. Furthermore, Nissen violated his own inclusionary criteria by omitting trials that favored the safety of Avandia. When other investigators attempted to replicate Nissen's meta-analysis, they found that more appropriate methodological choices caused the statistical significance in Nissen's meta-analysis to evaporate.[99]

One of the problems with the Nissen meta-analysis was that many of the included trials had zero events, that is, there were no heart attacks in either the Avandia or the comparator groups. Zero adverse events would seem to be a good thing. The rate of heart attack was equal and low in both groups of these clinical trials. The meta-analytic method used by Nissen, however, excluded zero-event trials from consideration.[100] A meta-analysis based upon the risk difference, which gave some weight to the zero-event trials, showed that random error had not been excluded as an explanation for the increased number of heart attacks among patients on Avandia.[101]

The complexities of the competing meta-analyses of heart attacks among patients on Avandia did not gain recognition in the *Avandia* MDL. Proxies for reliability—the fact of peer review publication, in a well-respected journal, by a well-credentialed investigator—trumped analysis of validity and reliability.[102]

§ 30A:9 Statistical Inference in Securities Fraud Cases Against Pharmaceutical Manufacturers

Pharmaceutical manufacturers are particularly vulnerable to securities fraud claims arising from the manufacturers' pronouncements about safety or efficacy, the evidence for which is often statistical in

---

99. *See, e.g.*, G. Diamond, et al., *Uncertain Effects of Rosiglitazone on the Risk for Myocardial Infarction and Cardiovascular Death*, 147 ANN. INTERN. MED. 578 (2007).

100. *In re* Neurontin Mktg., Sales Practices & Prods. Liab. Litig., 612 F. Supp. 2d 116, 126 (D. Mass. 2009) (MDL 1629) (discussing risk difference in meta-analysis as an appropriate method for sparse data).

101. L. Tian, et al., *Exact and Efficient Inference Procedure for Meta-Analysis and its Application to the Analysis of Independent 2 X 2 Tables With All Available Data But Without Artificial Continuity Correction*, 10 BIOSTATISTICS 275 (2009).

102. *See generally* M. Finkelstein & B. Levin, *Meta-Analysis of "Sparse" Data: Perspectives from the Avandia Cases*, 52 JURIMETRICS J. 123 (2012) (exploring the validity issues in the Nissen meta-analysis).

nature. Safety claims may involve complex data sets, both from observational studies and clinical trials. Efficacy claims are typically based upon clinical trial data.

Publicly traded manufacturers may find themselves caught between competing securities regulations. In evaluating safety or efficacy data, manufacturers will often consult with an outside science advisory board, or report to regulatory agencies. Securities regulations specify that any disclosure of confidential inside information to an outsider triggers an obligation of prompt public disclosure of that information.[103] Companies also routinely seek to keep investors informed of research and marketing developments. Generally, manufacturers will make their public disclosures through widely circulated press releases.[104] Not surprisingly, disgruntled investors may challenge the accuracy of the press releases, when the product or drug turns out to be less efficacious or more harmful than represented in the press release. These challenges, brought under the securities laws, often are maintained in parallel to product liability actions, sometimes in the same multi-district litigation.

Securities laws require accurate disclosure of all material information.[105] Rule 10b-5 of the Securities Exchange Commission (SEC) prohibits any person from making "any untrue statement of material fact" or from omitting "a material fact necessary in order to make the statements made, in light of the circumstances under which they were made, not misleading."[106]

A prima facie case of securities fraud requires that plaintiff allege and establish, among other things, a material misrepresentation or omission.[107]

---

103. Security Exchange Comm'n Regulation FD, 17 C.F.R. § 243.100 (requiring prompt public disclosure of any confidential, material inside information after disclosed to non-insiders).

104. Selective Disclosure and Insider Trading, Securities Act Release No. 7881, Fed. Sec. L. Rep. (CCH) ¶ 86,319 (Aug. 15, 2000) ("As a general matter, acceptable methods of public disclosure for purposes of Regulation FD will include press releases distributed through a widely circulated news or wire service . . . .").

105. Section 10(b) of the Exchange Act of 1934 prohibits any person "[t]o use or employ, in connection with the purchase or sale of any security . . . any manipulative or deceptive device or contrivance in contravention of such rules and regulations as the [Securities and Exchange Commission] may prescribe." 15 U.S.C. § 78j(b).

106. 17 C.F.R. § 240.10b-5.

107. Stoneridge Inv. Partners LLC v. Scientific-Atlanta, 552 U.S. 148, 157 (2008) ("(1) a material misrepresentation or omission []; (2) scienter; (3) a connection between the misrepresentation or omission and the purchase or sale of a security; (4) reliance upon the misrepresentation or omission; (5) economic loss; and (6) loss causation.").

The obligations to speak and to speak accurately have opened manufacturers to second guessing in their analyses of safety and efficacy data. In most securities fraud cases, courts have given manufacturers a wide berth by rejecting differences in opinions about the proper interpretation of studies as demonstrating fraud under the securities regulations.[108] This latitude has been given both in judgment of what test procedures to use, as well as in how best to interpret data.[109] In *Padnes v. Scios Nova Inc.*, the manufacturer was testing a drug for treatment of acute kidney failure. Scios Nova issued a press release after its phase II trial, to announce a statistically significant reduction in patients' need for dialysis. When the early phase III results failed to confirm this result, plaintiffs sued Scios Nova for not disclosing the lack of statistically significant outcomes on other measures of kidney function, as well as for its interpretation of dialysis results as statistically significant.[110] The trial court dismissed the complaint.[111]

Several securities fraud cases have turned on investor dissatisfaction on how companies interpreted clinical trial subgroup data. In *Noble Asset Management v. Allos Therapeutics, Inc.*,[112] the company issued a press release, noting no statistically significant increase overall in survival advantage from a drug for breast cancer, but also noting a statistically significant increased survival in a non-prespecified subgroup of patients with metastatic breast cancer.[113] The plaintiff investors claimed that the company should have disclosed that the FDA was unlikely to approve an indication based upon an ad hoc subgroup analysis, but the trial court rejected this claim because FDA policy on drug approvals is public and well known.[114] The plaintiffs

---

108. *In re* Medimmune, Inc. Sec. Litig., 873 F. Supp. 953, 965 (D. Md. 1995). The biological product at issue in this case was Respivir, a polyclonal antibody product, which "significantly" reduced frequency of hospitalization for respiratory syncytial virus (RSV). Plaintiffs alleged "flaws" in study design, but the trial court appeared to interpret the statistical significance to mean that Respivir was "unquestionably efficacious." *Id.* at 967.

109. *See, e.g.*, Padnes v. Scios Nova Inc., No. C 95-1693 MHP, 1996 WL 539711, at *5 (N.D. Cal. Sept. 18, 1996) (Patel, J.) [cited herein as Padnes]. *See also* DeMarco v. DePoTech Corp., 149 F. Supp. 2d 1212, 1225 (S.D. Cal. 2001) ("Although plaintiffs have established a legitimate difference in opinion as to the proper statistical analysis, they have hardly stated a securities fraud claim."); *In re* Aldor Corp. Sec. Litig., 616 F. Supp. 2d 551, 568 n.15 (E.D. Pa. 2009) (allegations as to how data should have been analyzed do not support claims for false or misleading statements).

110. *Padnes*, 1996 WL 539711, at *2.

111. *Id.* at *10.

112. Noble Asset Mgmt. v. Allos Therapeutics, Inc., 2005 WL 4161977 (D. Colo. Oct. 20, 2005).

113. *Id.* at *1.

114. *Id.* at *6–7.

also complained that the press release referred to statistically significant results from a Cox multiple regression analysis rather than the log-rank (non-parametric survival) analysis required by FDA. The trial court rejected this claim as well, opining that the analysis was not misleading when the company correctly reported the raw data and the results of the Cox multiple regression analysis.[115]

Two recent appellate decisions emphasize the courts' unwillingness to scrutinize the contested statistical methodology that underlies plaintiffs' claims of misrepresentation. In *In re Rigel Pharmaceuticals, Inc. Securities Litigation*, the plaintiff investors were dissatisfied, not with reporting of subgroups, but rather with the failure of the company to report geographic subgroup results, as well as its use of allegedly improper statistical tests and its failure to account for multiple comparisons.[116]

The Ninth Circuit affirmed the dismissal of a complaint. The appellate court held that allegations of "statistically false p-values" were not sufficient; plaintiffs must allege facts that explain why the difference between two statements "is not merely the difference between two permissible judgments, but rather the results of a falsehood."[117] Alleging that a company should have used a different statistical method to analyze the data from its clinical trial is not sufficient to raise an issue of factual falsity under the securities fraud statute and regulations.[118] The Court explained that the burden was on plaintiffs to plead and prove that the difference between two statistical statements "is not merely the difference between two permissible judgments, but rather the result of a falsehood."[119] The Court characterized the plaintiffs' allegations to be about judgments of which statistical tests or methods are appropriate, and not about false statements. Furthermore, the Court emphasized that the company's statistical method was called for in the trial protocol, and was selected before the data were unblinded and provided to the company.[120]

---

115. *Id*. at *11.
116. 2010 WL 8816155 (N.D. Cal. Aug. 24, 2010).
117. 697 F.3d 869, 877 (9th Cir. 2012) (internal citations omitted), *aff'g* 2010 WL 8816155 (N.D. Cal. Aug. 24, 2010).
118. *Id*. at 877–78.
119. *Id*. at 878.
120. *Id*. ("Because there are many ways to statistically analyze data, it is necessary to choose the statistical methodology before seeing the data that is collected during the trial; otherwise someone can manipulate the unblinded data to obtain a favorable result."), citing and attempting to distinguish United States v. Harkonen, 2010 WL 2985257, at *4 (N.D. Cal. July 27, 2010).

In *Kleinman v. Elan Corp.*,[121] the Second Circuit affirmed the dismissal of a securities fraud class action against two pharmaceutical joint venturers, which issued a challenged press release on interim phase II clinical trial results for bapineuzumab, a drug for mild- to moderate-Alzheimer's disease. The press release at issue announced "top line" findings and promised a full review at an upcoming international conference.[122] According to the release, the clinical trial data did not show a statistically significant benefit on the primary efficacy end point, but "[p]ost-hoc analyses did show statistically significant and clinically meaningful benefits in important subgroups."[123]

The plaintiffs in Kleinman complained that the clinical trial had started with crucial imbalances between drug and placebo arms, thus indicating a failure in randomization, and that the positive results had come from impermissible post-hoc subgroup analyses.[124] The appellate court appeared not to take the randomization issue seriously, and rejected the notion that statements can be false when they represent a defendant company's reasonable interpretation of the data, even when the interpretation later turns out to be shown to be false:[125]

> "At bottom, Kleinman simply has a problem with using post-hoc analyses as a methodology in pharmaceutical studies. Kleinman cites commentators who liken post-hoc analyses to moving the goal posts or shooting an arrow into the wall and then drawing a target around it. Nonetheless, when it is clear that a post-hoc analysis is being used, it is understood that those results are less significant and should have less impact on investors. Our job is not to evaluate the use of post-hoc analysis in the scientific community; the FDA has already done so."

In *United States v. Harkonen*,[126] the government turned the law of statistical analyses in securities fraud on its head, when it prosecuted

---

121. Kleinman v. Elan Corp., 706 F.3d 145 (2d Cir. 2013).
122. *Id*. at 149.
123. *Id*. at 149–50 (also noting that the press release provided a "preliminary analysis," which might be less favorable upon further analysis).
124. *Id*. at 150.
125. *Id*. at 154–55 & 155 n.11 (citing and quoting FDA Center for Drug Evaluation and Research, *E9 Statistical Principles for Clinical Trials*, 63 Fed. Reg. 49583, 49595 (Sept. 16, 1998), that post-hoc analyses are exploratory and "unlikely" to be accepted as support of efficacy.)
126. United States v. Harkonen, 2010 WL 2985257 (N.D. Calif. 2010) ((Patel, J.) (denying defendant's post–trial motions to dismiss the indictment, for acquittal, or for a new trial). Sometimes judges are looking for bright lines in the wrong places).

a pharmaceutical company executive for his role in issuing a press release on clinical trial data. The jury acquitted Dr. Harkonen on a charge of misbranding,[127] but convicted on a single count of wire fraud.[128] Dr. Harkonen's crime? Bad statistical practice. The government conceded that the data represented in the press release were accurate, as were the calculated p-values. The chargeable offense lay in Dr. Harkonen's describing the clinical trial results as "demonstrating a survival benefit" of the biological product (interferon γ-1b) in a clinical trial subgroup of patients with mild- to moderate-idiopathic pulmonary fibrosis. The p-value for the subgroup was 0.004, with an effect size of 70% reduction in mortality. The subgroup, however, was not prespecified, and was not clearly labeled as a post-hoc analysis. The trial had not achieved statistical significant on its primary end point.

In prosecuting Dr. Harkonen, the government offered no expert witness opinion. Instead, it relied upon a member of the clinical trial's data safety monitoring board, who advanced a strict, orthodox view that if the primary end point of a trial "failed," then the data could not be relied upon to infer any meaningful causal connection within secondary end points, let alone non-prespecified end points. The prespecified survival secondary end point showed a 40 percent reduction in mortality, $p = 0.08$ (which shrank to 0.055 on an intent-to-treat analysis). The press release also relied upon a previous small clinical trial that showed a benefit in survival at five years, with the therapy group at 77.8%, compared with 16.7% in the control groups, $p = 0.009$.

The trial court accepted the government's claim that p-values less than 0.05 were something of "magic numbers,"[129] and rejected post-trial motions for accquittal. Dr. Harkonen's use of "demonstrate" to describe a therapeutic benefit was, in the trial court's view, fraudulent, because of the lack of "statistical significance" on the primary end point, and the multiple testing with respect to the secondary survival end point. The Ninth Circuit affirmed the judgment of conviction in an unpublished per curiam opinion.[130]

In contrast to the criminal wire fraud prosecution, the civil fraud actions against Dr. Harkonen and the company were dismissed.[131] The prosecution and the judgment in *United States v. Harkonen* is at

---

127. 21 U.S.C. §§ 331(k), 333(a)(2), 352(a).
128. 18 U.S.C. § 1343.
129. United States v. Harkonen, 2010 WL 2985257, at *5 (N.D. Calif. 2010).
130. United States v. Harkonen, 2013 WL 782354 (9th Cir. 2013).
131. *In re* Actimmune Mktg. Litig., 2010 WL 3463491 (N.D. Cal. Sept. 1, 2010), *aff'd*, 464 F. App'x 651 (9th Cir. 2011).

odds with the latitude afforded companies in securities fraud cases. Furthermore, the case cannot be fairly squared with the position that the government took as an amicus curiae in *Matrixx Initiatives, Inc. v. Siracusano*,[132] where the Solicitor General's office, along with counsel for the Food and Drug Division of the Department of Health & Human Services, in their zeal to assist plaintiffs on claims against an over-the-counter pharmaceutical manufacturer, disclaimed the necessity, or even the importance, of statistical significance:[133]

> "[w]hile statistical significance provides some indication about the validity of a correlation between a product and a harm, a determination that certain data are not statistically significant … does not refute an inference of causation."

Suddenly, when prosecuting an unpopular pharmaceutical company executive, the government's flexible approach changed. Government duplicity was a much greater problem than statistical multiplicity in *Harkonen*.[134]

§ 30A:10 Conclusion

In products liability cases, invalid statistical reasoning leads to bad legal reasoning. Lawyers should avoid relying upon, and citing language in, judicial cases for the meaning of statistical terms. Buy and use a good statistics book!

The widespread difficulty experienced by judges in evaluating statistical evidence suggests that the Federal Judicial Center's educational mission might require greater support if our courts will have the ability to adjudicate twenty-first century disputes. Of course, courts

---

132. Matrixx Initiatives, Inc. v. Siracusano, 131 S. Ct. 1309 (2011).

133. Brief for the United States as Amicus Curiae Supporting Respondents, in Matrixx Initiatives, Inc. v. Siracusano, 2010 WL 4624148, at *14 (Nov. 12, 2010).

134. Dr. Harkonen is expected to petition the Supreme Court for certiorari on statutory and constitutional grounds. *See* Alex Kozinski & Stuart Banner, "Who's Afraid of Commercial Speech?" 76 VA. L. REV. 627, 635 (1990) ("[T]here are many varieties of noncommercial speech that are just as objective as paradigmatic commercial speech and yet receive full first amendment protection. Scientific speech is the most obvious; much scientific expression can easily be labeled true or false, but we would be shocked at the suggestion that it is therefore entitled to a lesser degree of protection. If you want, you can proclaim that the sun revolves around the earth, that the earth is flat, that there is no such thing as nitrogen, that flounder smoke cigars, that you have fused atomic nuclei in your bathtub—you can spout any nonsense you want, and the government can't stop you.").

have the option of appointing expert witnesses (court-appointed expert witnesses), or technical advisors, even when the parties do not urge these approaches. Proposals for science courts, once in vogue, deserve further consideration in the light of an empirical analysis of the judicial use of statistical and scientific evidence.