The Legal Sequelae of the 2016 American Statistical Association P-Value Statement

Nathan A. Schachtman
July 2019

In 2016, the American Statistical Association issued an unusual guidance document in which it attempted to redress its perception that p-values and statistical significance were widely misunderstood and misinterpreted. In addition to providing guidance on the meaning and use of attained significance probabilities, the ASA also encouraged the use of "other methods that emphasize estimation over testing," including Bayesian methods. Although the ASA guidance document warned against misuses of p-values, it did not warn of the potential for misapplication of these "other methods." The reaction of some segments of the legal community was prompt, both in interpreting the 2016 guidance as a rejection of p-values and significance testing, as well as an encouragement to use "other methods," for which the judiciary would have far less experience and acumen to detect invalid inferences. In this presentation, I will discuss how the ASA Statement was used rhetorically to justify causal claims that had been rejected by the FDA and scientific organizations, and to advance a "Bayesian hypothesis" test to support a claim that a meta-analysis showed that there was an 85 percent probability that testosterone replacement therapy caused either heart attack or stroke.

In the testosterone replacement therapy (TRT) MDL, the ASA Statement took on out-sized significance. The defense's Rule 702 motion to exclude plaintiffs' expert witnesses' causation opinions pointed to those witnesses' failure to rely upon studies that had shown statistically significant associations, as well as their failure to rely upon studies that had ruled out bias and confounding.[1] Plaintiffs' opposition relied heavily upon the ASA Statement, and insisted that the defense had over-emphasized the importance of study statistical

[1] Defendants' Mem. of Law in support of Motion to Exclude Plaintiffs' Expert Testimony on Causation, and for Summary Judgment, MDL 2545, 2017 WL 1104501, at *3, 7-8 (N.D. Ill. Feb. 20, 2017).

significance.[2] Despite the Plaintiffs' having made the Statement an exhibit to their brief, the MDL trial court seemed not to have appreciated the dispute; its opinion on the motions misstated the definition of confidence intervals and committed the transposition fallacy.[3] In a subsequent bellwether trial, plaintiffs' counsel extensively cross-examined the defense clinical cardiologist on the ASA Statement and other technical literature,[4] which he had never seen before, after the trial court had excluded the defense statistician as "cumulative."[5] Part of the defense motion for new trial was based upon the prejudice of not having a statistical expert witness available to address the ASA Statement.

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 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 begin to mature until the 1970s. Even then, the scientific community remained skeptical of, if not openly 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:

---

[2] Plaintiffs' Mem. of Law in Opp. to Motion of AbbVie Defendants to Exclude Plaintiffs' Expert Testimony on Causation, MDL 2545, 2017 WL 2889168, at *34 (N.D. Ill. Mar. 23, 2017).

[3] In re Testosterone Replacement Therapy Prods. Liab. Litig., MDL No. 2545, C.M.O. No. 46, 2017 WL 1833173 at *4 (N.D. Ill. May 8, 2017) ("Statisticians test for statistical significance to determine the likelihood that a study's findings are due to chance. *** According to conventional statistical practice, such a result … would be considered statistically significant if there is a 95% probability, also expressed as a 'p-value' of <0.05, that the observed association is not the product of chance.").

[4] Notes of Testimony of William French, M.D., vol. 10, at 2327-40, in Mitchell v. AbbVie, case no. 14 C 9178 (N.D. Ill. Apr. 23, 2018). The ASA Statement was not mentioned on direct examination.

[5] Mitchell v. AbbVie, MDL No. 2545, 2018 WL 1316724 (N.D. Ill. Mar. 14, 2018) (excluding Laurentius Marais). The MDL court had previously rejected a challenge to Marais on grounds that he was a statistician and not an epidemiologist. In re Testosterone Replacement Therapy Prods. Liab. Litig., MDL No. 2545, 2017 WL 1833173 (N.D. Ill. May 8, 2017).

> "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."[6]

In the last two decades, however, 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 considerably in this period. Today, thousands of meta-analyses, of both observational and experimental studies, are published each year.[7] The suggestion that meta-analyses are rarely involved in litigation[8] is untenable. Meta-analyses have been involved in several dozens of pharmaceutical and other products litigations, several of which were federal or state MDL cases that encompassed many thousands of claims.[9]

Consensus Statements on Meta-Analyses

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 the conduct and reporting of meta-analyses of clinical trials, and of observational studies.[10] Recently, the U.S. Food and Drug Administration issued a draft guidance on the use of meta-analysis of

---

[6] 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).

[7] 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.

[8] D. FAIGMAN, et al., 3 MODERN SCI. EVIDENCE, § 23:14 n.10 (2018) (incorrectly stating that "[c]ourts infrequently confront expert opinions based on a meta-analysis").

[9] M. Finkelstein & B. Levin, *supra* note 127, at 124 & n.3. *See* § 29.11-5, *infra*.

[10] *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); D. Moher, et al., *Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement*, 151 ANN. INTERN. MED. 264 (2009). *See also In re* Zoloft Prods. Liab. Litig., MDL No. 2342, 26 F.Supp. 3d 449, 457 & n.25 (E.D. Pa. 2014) (criticizing and excluding expert witness who eschewed reliance upon a published meta-analysis that complied with MOOSE criteria, without adequately explaining the basis for her rejection).

randomized controlled clinical trials to evaluate medication safety outcomes.[11] Although well-conducted meta-analyses can advance our knowledge of biomedical phenomena, the threats to validity of meta-analyses are substantial, and worthy of greater recognition by litigants and courts.

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 or proportion). Sensitivity analyses within subsets of all studies can further help detect and evaluate 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.[12] 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.[13] Acknowledging that meta-analysis at issue might be invalid, the Third Circuit directed the trial court on remand to evaluate the methodological validity of the meta-analysis as performed.

The professional skepticism about meta-analysis in the 1970s 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

---

[11] FDA CDER, Meta-Analyses of Randomized Controlled Clinical Trials to Evaluate the Safety of Human Drugs or Biological Products – (Draft) Guidance for Industry (Nov. 2018).

[12] *In re* Paoli R.R. Yard PCB Litig., 706 F. Supp. 358, 373 (E.D. Pa. 1988).

[13] *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).

statistical procedure that claimed to make something out of nothing.[14] 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."[15]

The trial court's analogy to adding zeroes is mathematically incorrect. The Second Circuit reversed and remanded the colorectal cancer claim for trial.[16] Over a decade later, 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.[17] 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.[18] 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

---

[14] 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).

[15] *In re* Joint E. & S. Dist. Asbestos Litig., 827 F. Supp. 1014, 1042 (S.D.N.Y. 1993).

[16] *In re* Joint E. & S. Dist. Asbestos Litig., 52 F.3d 1124 (2d Cir. 1995).

[17] Institute of Medicine, ASBESTOS: SELECTED CANCERS 226 (Wash., D.C. 2006).

[18] *See, e.g.*, Baker v. Chevron USA, Inc., 680 F. Supp. 2d 865 (S.D. Ohio 2010) (addressing a meta-analysis on multiple myeloma 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. A similar error can be found in another benzene case in which plaintiffs' expert witnesses claimed that benzene caused myelofibrosis with myeloid metaplasia, based upon a meta-analysis of epidemiologic studies. LeBlanc v. Chevron USA Inc., No. 05–5485, 2009 WL 3837397 (E.D. La. Nov. 13, 2009), *aff'd*, 396 F. App'x 94 (5th Cir. 2010) (per curiam). Some of the studies in the meta-analysis did not sufficiently isolate the exposure of interest, or the outcome of interest, but the trial court added another, aberrant, reason for rejecting the expert witness's meta-analysis: the lack of statistical significance of the studies included in the meta-analysis. 2009 WL 3837397, at *3.

tissue diseases, to combat selective, partial presentation of evidence.[19] 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,[20] which they later published.[21] In the welding fume MDL, plaintiffs' expert witness offered a crude, non-quantified meta-analysis to argue that welding causes Parkinson's disease.[22] In rebuttal, a defense expert witness offered a quantitative meta-analysis, which provided strong evidence against plaintiffs' claim.[23] Although the MDL court excluded the defense expert's meta-analysis from the pre-trial Rule 702 hearing as untimely, the entire MDL was expeditiously resolved in the aftermath of the hearing.

Subsequently, the defense expert witness, with professional colleagues, published an expanded version of the meta-analysis.[24]

Meta-analyses of observational studies played a dispositive role in litigation of birth defects alleged to have been caused by maternal use of the anti-depressant sertraline. Plaintiffs' epidemiologist, Dr. Anick Bérard, ascribed the lack of statistical significance in many of the studies upon which she relied to insufficient statistical power.[25] In her Rule 26 report, Dr. Bérard relied upon a published meta-analysis that purported to aggregate data from studies to find a statistically significantly increased rate of certain birth defects. Upon being confronted at deposition with serious errors in this meta-analysis, including incorrect data input, transposition of numbers, and double counting studies,

---

[19] O. Wong, *A Critical Assessment of the Relationship between Silicone Breast Implants and Connective Tissue Diseases*, 23 REG'Y TOXICOL. & PHARMACOL. 74 (1996).

[20] B. Hulka, et al.,, *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.html. *See also* B. Hulka, et al., *Experience of a Scientific Panel Formed to Advise the Federal Judiciary on Silicone Breast Implants*, 342 NEW ENG. J. MED. 812 (2000).

[21] E. Janowsky, et al., *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).

[22] Deposition of Dr. Sanchez-Ramos, in Street v. Lincoln Elec. Co., Case No. 1:06-cv-17026, 2011 WL 6008514 (N.D. Ohio May 17, 2011).

[23] Deposition of Dr. Mortimer, in Street v. Lincoln Elec. Co., Case No. 1:06-cv-17026, 2011 WL 6008054 (N.D. Ohio June 29, 2011).

[24] J. Mortimer, et al., *Associations of Welding and Manganese Exposure with Parkinson's Disease: Review and Meta-Analysis*, 79 NEUROLOGY 1174 (2012).

[25] *In re* Zoloft Prods. Liab. Litig., MDL No. 2342, 26 F.Supp. 3d 449 (E.D. Pa. 2014) .

Dr. Bérard withdrew her reliance upon the meta-analysis.[26] At a subsequent Rule 702 hearing, Dr. Bérard attempted to support her "lack of power" theory by using a graphic representation of risk ratios from highly selected study end points, without the meta-analysis results. The trial court found this evasion unacceptable:

> "When epidemiologists hypothesize that there is a 'true' association which individual studies are underpowered to detect at a statistically significant level, the widely accepted approach to combining data from multiple studies—thus increasing the power to detect an association—is to conduct a systematic meta-analysis. Dr. Bérard did not address, in her report or her testimony, her reasons for relying upon her novel method of examining trends in odds ratios to test her hypothesis, rather than relying upon the well-established method generally relied upon by epidemiologists, nor did she provide objective, independent validation for her novel method."[27]

Dr. Bérard acknowledged that she was familiar with another published meta-analysis,[28] favorable to the defense, which complied with professional standards for such studies.[29] Although she was critical of this other meta-analysis for excluding some studies, Dr. Bérard failed to identify which studies were excluded and why the exclusions affected the validity of the meta-analysis.[30] The trial court found Dr. Bérard's disregard of the exonerative meta-analysis unpersuasive and excluded her proffered testimony. After permitting a substitute expert witness, who performed his own meta-analysis, the MDL Court conducted another Rule 702 hearing, and excluded the new witness, for

---

[26] S. Nikfar, *et al.*, *Increasing the Risk of Spontaneous Abortion and Major Malformations in Newborns Following Use of serotonin Reuptake Inhibitors During Pregnancy: A Systematic Review and Updated Meta-analysis*, 20 DARU J. PHARM. SCI. (2012).

[27] *Id.* at *18–19 (internal citations omitted).

[28] Nicholas Myles, *et al.*, *Systematic Meta-Analysis of Individual Selective Serotonin Reuptake Inhibitors Medications and Congenital Malformations*, 47 AUSTR. & NEW ZEALAND J. PSYCH. 1002 (2013).

[29] MOOSE Statement, *supra* note 130, at n.70.

[30] *In re* Zoloft, 26 F.Supp. 3d 449, 459-60 (E.D. Pa. 2014).

a variety of serious methodological flaws, including unexplained selectivity in what studies were included in his for-litigation meta-analysis.[31]

Competing Models for Meta-Analysis

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

The key steps for conducting 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 and excluded studies;
(4) Review the methods and results of all candidate studies;
(5) Abstract and summarize each included study's results;
(6) Assess the statistical and clinical variation (heterogeneity) between studies;
(7) Assess the suitability of the available studies for aggregation;
(8) Select the appropriate statistical model for meta-analysis:
(9) Apply statistical methods to produce a summary estimate of association;
(10) Report and interpret the findings;
(11) Conduct appropriate sensitivity analyses; and
(12) 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 weigh 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 have 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

[31] In re Zoloft Prods. Liab. Litig., 858 F.3d 787, 797 (3d Cir. 2017) (affirming exclusion of statistician who could not explain why he had included some and not other studies in his meta-analysis).

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

Recent Cases Involving Meta-Analyses

Expert witness testimony based upon meta-analysis has been proffered in numerous diverse pharmaceutical cases, including recent cases involving Accutane,[32] Aredia and Zometa,[33] Avandia,[34] Baycol,[35] benzodiazepine,[36] Celebrex,[37] Fosamax,[38] Gadolinium,[39] Neurontin,[40] Seroquel,[41] Thimerosal,[42] Trasylol,[43] Vioxx,[44] Zoloft,[45] and Zyprexa.[46] The various meta-analyses in these, and other, litigations raise many interesting, complex issues of validity.

A brief discussion of the role of meta-analysis in 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.[47] 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

---

[32] In re Accutane Litig., 234 N.J. 340, 191 A.3d 560 (2018).
[33] Deutsch v. Novartis Pharm. Corp., 768 F. Supp. 2d 420 (E.D.N.Y. 2011).
[34] *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).
[35] *In re* Baycol Prods. Litig., 532 F. Supp. 2d 1029 (D. Minn. 2007).
[36] Vinitski v. Adler, 69 Pa. D. & C. 4th 78, 2004 WL 2579288 (Phila. Cnty. Ct. Com. Pl. Sept. 17, 2004).
[37] *In re* Bextra & Celebrex Mktg. Sales Practices & Prod. Liab. Litig., 524 F. Supp. 2d 1166 (2007).
[38] *In re* Fosamax Prods. Liab. Litig., 645 F. Supp. 2d 164 (S.D.N.Y. 2009).
[39] *In re* Gadolinium-Based Contrast Agents Prod. Liab. Litig., 2010 WL 1796334 (N.D. Ohio May 4, 2010).
[40] *In re* Neurontin Mktg., Sales Practices & Prods. Liab. Litig., 612 F. Supp. 2d 116 (D. Mass. 2009).
[41] *In re* Seroquel Prods. Liab. Litig., 2009 WL 3806434, at *5 (M.D. Fla. June 18, 2009).
[42] Hennessey v. Sec'y Dep't Health & Human Servs., 2009 WL 1709053 (Fed. Cl. May 29, 2009).
[43] *In re* Trasylol Prods. Liab. Litig., 2010 WL 1489793 (S.D. Fla. Feb. 24, 2010).
[44] Merck & Co., Inc. v. Ernst, 296 S.W.3d 81 (Tex. Ct. App. 2009).
[45] Miller v. Pfizer, Inc., 356 F.3d 1326 (10th Cir. 2004).
[46] *In re* Zyprexa Prods. Liab. Litig., 489 F. Supp. 2d 230 (E.D.N.Y. 2007).
[47] 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).

are of unequal size. Indeed, 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.[48]

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 arms in several of the included clinical trials. Zero adverse events in both arms 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.[49] 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.[50]

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.[51]

The Bayesian Invitation

A now mostly concluded MDL showcased the use and misuse of meta-analysis in the context of claims that testosterone replacement therapies (TRT)

---

[48] *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).

[49] *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).

[50] 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).

[51] *See generally* M. Finkelstein & B. Levin, *supra* note 127 (exploring the validity issues in the Nissen meta-analysis); J. Stone, et al., *Was there really any evidence that rosiglitazone increased the risk of myocardial infarction or death from cardiovascular causes?* 25 PHARMACOEPIDEM. & DRUG SAFETY 223 (2015) ("no").

cause heart attack and stroke. In 2014, Public Citizen petitioned FDA for revisions to the TRT labeling, based upon two observational studies and a meta-analysis of clinical trials. The FDA's careful review of the petition found serious flaws in both the studies and the meta-analysis, which used an overly broad composite end point, and had questionably included clinically irrelevant trials to reach its summary estimate of association.[52] The FDA did, however, convene an Advisory Committee, which held hearings, which in turn led to revisions in the class labeling for TRT with respect to indication and a *possible* association with cardiovascular outcomes. And then the litigation came.

An initial round of Rule 702 motions in the TRT MDL included challenges to plaintiffs' expert witnesses, two of whom had used a published meta-analysis of clinical trial safety outcomes,[53] which had reported a slightly elevated, statistically non-significant risk ratio for a composite end point (heart attack and stroke) to conduct a "Bayesian hypothesis test." Using the Bayesian meta-analytic method described by John Carlin,[54] plaintiffs' witnesses asserted an 85% posterior probability, given the predicate meta-analysis, for the claim that there was some increased risk of the composite end point.

The defense Rule 702 motion challenged the plaintiffs' Bayesian analysis on several grounds, including the:

(1) the plaintiffs' witnesses' failure to publish their analysis;

(2) one of witnesses' having never published a significant Bayesian analysis previously;

(3) the absence of Bayesian analyses in the relevant studies on testosterone

(4) the rarity of Bayesian analyses in product liability cases;

(5) the witnesses' failure to state what the actual risk was, as opposed to the probability that it exceeded 1.0; and

---

[52] Letter of Janet Woodcock, Director of FDA's Center for Drug Evaluation and Research, to Sidney Wolfe, Director of Public Citizen's Health Research Group (July 16, 2014) (denying citizen petition for "black box" warning).

[53] S. Albert & J. Morley, *Testosterone therapy, association with age, initiation and mode of therapy with cardiovascular events: a systematic review*, 95 CLIN. ENDOCRINOL. 436 (2016).

[54] J. Carlin, *Meta-analysis for 2 x 2 tables: a Bayesian approach*, 11 STAT. MED. 141 (1992).

(6) the defense expert witness's calculation that the "Increased [cardiovascular] risk meets only a 70% level of evidence, which is far below the 95% level required."[55]

The MDL Court found the challenged expert witness's expertise in Bayesian statistics adequate, and the *Reference Manual*'s suggestion that such methods are subscribed to by a "well-established minority" to be sufficient to deny the defense challenge. The court also found that the challenge failed to articulate how Bayesian methods were subjective in a way that frequentist methods were not. Most important, the court adopted the plaintiffs' rebuttal that it was the defense that had conflated the 95% probability of the coefficient of confidence with a posterior probability.[56]

Unfortunately, some of the most important issues raised by the plaintiffs' Bayesian meta-analysis were not raised or addressed by the MDL TRT court. Bayes' Rule is a theorem that provides a posterior probability for a claim or proposition based upon a prior probability and the strength of the evidence at hand. Unlike frequentist statistics, which treat the population value (mean or risk ratio) as having a fixed, but unknown value, Bayesian analyses treat both prior and posterior probabilities as probability distributions. Every Bayesian analysis must start with a prior probability, and therein lies a serious methodological problem. In the TRT cases, the proffered Bayesian analysis, based upon Carlin's method, invoked a prior risk ratio of 1.0, which standing

---

[55] Defendants' Motion to Exclude Plaintiffs' Expert Testimony on the Issue of Causation, and for Summary Judgment, and Mem. of Law in Support, Case No. 1:14-CV-01748, MDL 2545, 2017 WL 1104501, at *69-70. (N.D. Ill. Feb. 20, 2017) (citing Reference Manual, *supra* n.4, at 259, for the proposition that "'subjective Bayesians are a well-established minority' of scientists whose methods 'have rarely been used in court.'"). *See also* Plaintiffs' Mem. of Law in Opp. to Motion of AbbVie Defendants to Exclude Plaintiffs' Expert Testimony on Causation, and for Summary Judgment, MDL 2545, Doc. No. 1753 (N.D. Ill. Mar. 23, 2017).

[56] In re Testosterone Replacement Therapy Prods. Liab. Litig., MDL No. 2545, C.M.O. No. 46, 2017 WL 1833173 at *11-12 (N.D. Ill. May 8, 2017). The MDL court reiterated its views in response to a subsequent motion contra plaintiffs' statistical expert witness, Martin Wells. MDL No. 2545, 2018 WL 4030585, *8 (N.D.Ill. Aug. 23, 2018). The use of Bayesian statistics is indeed uncommon in civil litigation, outside of paternity cases. For a singular exception from Canadian tort law, *see* Goodman v. Viljoen, 2011 ONSC 821 (CanLII), *aff'd*, 2012 ONCA 896 (CanLII), *leave appeal den'd*, S. Ct. Canada No. 35230 (July 11, 2013). In regulatory affairs, the FDA has promulgated a guidance for recommended practices of Bayesian analyses in medical device trials. FDA Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials (February 5, 2010); 75 Fed. Reg. 6209 (February 8, 2010). Bayesian methods are rarely the primary pre-specified approaches in clinical trials for medication approvals.

alone might seem perfectly fair. The chosen variance around 1.0, which made up the prior probability distribution, however, was extremely wide and flat, essentially encompassing no risk at the low end, and absolute risk, at the high end. The assumptions of the plaintiffs' meta-analysis were, therefore, completely unrealistic and counterfactual.

Furthermore, the proffered Bayesian meta-analysis in the TRT cases was based upon Carlin's method, which assumed a "hierarchical normal model," which he described as reasonable "as long as the studies are large and observed counts are not too small."[57] In the dataset used by plaintiffs' expert witnesses, however, virtually all the studies had very low event counts, often zero or one, in either the TRT or placebo arm, or both. Carlin acknowledged that it was difficult to assess the validity of the normal model, and emphasized that "[a] study of the sensitivity of conclusions to the choice of prior would be important."[58] Subsequent simulation studies have shown that so-called "vague" or "non-informative" priors, such as were used by plaintiffs' expert witnesses, can exercise an "unintentionally large degree of influence on any inferences."[59] The plaintiffs' litigation expert witnesses offered no tests of the validity of Carlin's method in the context of meta-analyzing clinical trials for sparse safety outcomes.

The Bayesian analysis in the TRT cases also suffered from two additional, more readily appreciated problems not raised in the MDL court's 2017 opinion. First, the advancing of a causal claim with an 85% posterior probability was bound to be confused with the plaintiffs' burden of proof of greater than 50%, notwithstanding that the calculated posterior probability did not take into account uncertainty from bias and other non-random errors in the aggregated clinical trial data. Second, the posterior probability was based upon a composite end point that combined heart attack and stroke. As a later deposition of one of the plaintiffs' Bayesian protagonists showed, had the Carlin method been applied to just the heart attack summary point estimate,

---

[57] Carlin, *supra* at note 174, at 157.

[58] *Id*.

[59] P. Lambert, et al., *How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS*," 24 STATS. MED. 2401, 2402. *See also* A. Gelman, *Prior distributions for variance parameters in hierarchical models*, 1 BAYESIAN ANALYSIS 515 (2006); E. Pullenayegum, *An informed reference prior for between-study heterogeneity in meta-analyses of binary outcomes*, 30 STAT. MED. 3082 (2010).

then the posterior probability that TRT causes heart attack would have been less than 50%, and greater than 50% that TRT does not cause heart attack.[60]

[60] Deposition of Martin Wells, in Martin v. Actavis, Inc., case no. 15-cv-4292, 2018 WL 7350886 (N.D. Ill. Apr. 2, 2018).