



Liberty University Law Review

Volume 13
Issue 2 *Spring 2019*

Article 5

January 2019

Evidence-Based Medicine in Expert Testimony

Wells Mangrum

Richard Collin Mangrum

Follow this and additional works at: https://digitalcommons.liberty.edu/lu_law_review

Recommended Citation

Mangrum, Wells and Mangrum, Richard Collin (2019) "Evidence-Based Medicine in Expert Testimony," *Liberty University Law Review*. Vol. 13 : Iss. 2 , Article 5.

Available at: https://digitalcommons.liberty.edu/lu_law_review/vol13/iss2/5

This Article is brought to you for free and open access by the Liberty University School of Law at Scholars Crossing. It has been accepted for inclusion in Liberty University Law Review by an authorized editor of Scholars Crossing. For more information, please contact scholarlycommunications@liberty.edu.

ARTICLE

EVIDENCE-BASED MEDICINE IN EXPERT TESTIMONY

Dr. Wells Mangrum[†] and Professor Richard Collin Mangrum^{††}

I. INTRODUCTION

In early American common law, courts permitted experts to testify only if the opinion would be helpful to the trier of fact, narrowly interpreting that scope.¹ Some courts in the nineteenth century permitted expert testimony if it would be helpful on such things as the authorship of handwriting.² In the area of medical testimony, the Iowa Supreme Court in 1919 refused to permit a physician to explain that an x-ray or “skiagraph” depicted “a curvature of the spine,” holding, under the original writing rule, that the document speaks for itself.³ In 1923, the United States Court of Appeals for the District of Columbia, in *Frye v. United States*,⁴ became the first court to adopt a more analytical approach to expert testimony. This approach became known as the *Frye* standard of “general acceptance” for expert testimony:

† B.A. 2001; Harvard University; M.D. 2005, Mayo Medical School; Radiology Residency 2010, Duke University; Neuroradiology Fellowship 2011, Duke University.

†† Yossem Professor of Law, Creighton University School of Law; B.A. 1972, Harvard University; J.D. 1975, University of Utah School of Law; B.C.L., 1978, Oxford University; S.J.D., 1983, Harvard University School of Law.

1. Learned Hand, *Historical and Practical Considerations Regarding Expert Testimony*, 15 HARV. L. REV. 40 (1901).

2. *Marcy v. Barnes*, 82 Mass. 161 (1860).

It is a general rule, that the court is to determine in the first instance, upon the evidence produced, whether the witness offered is qualified by his peculiar skill, knowledge and experience in any particular art or employment, to testify as to his opinion as an expert; and unless the evidence upon which the determination to allow the witness to testify in that manner is reported, the decision is not open to revision in another court.

Id. at 164. See also *President of Quinsigamond Bank v. Hobbs*, 77 Mass. 250, 257 (1858) (“The authorities show that Southgate’s testimony, that in his opinion all the words in the note in suit were written at the same time, was within the legal province of an expert.”); Recent Case, *Evidence-Expert Testimony-Age of Handwriting*, 13 HARV. L. REV. 691 (1900) (“It seems a better doctrine to allow the trial judge to determine in each particular case whether the jury would be legitimately helped by an expert’s opinion.”).

3. *Lang v. Marshalltown*, 170 N.W. 463, 464-65 (Iowa 1919) (“[W]e are of the opinion that the court erred in permitting the doctor to testify to what appeared in the skiagraph” (quoting *Elzig v. Bales*, 112 N. W. 540, 541 (Iowa 1907))).

4. *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).

Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.⁵

Following the *Frye* case, most courts adopted the *Frye* standard of “general acceptability” for admitting expert testimony. Excepting “novel” expert testimony that had not acquired “general acceptability” amongst the relevant scientific community, courts were now permitted to admit expert testimony subject to the crucibles of cross-examination, rebuttal expert testimony, and closing arguments.⁶ However, the *Frye* standard began receiving criticism for providing a basis to exclude some expert testimony that may assist the trier of fact.⁷

In 1975, Congress enacted the Federal Rules of Evidence.⁸ Several of the newly-enacted rules addressed the issue of the admissibility of expert testimony. Federal Rule of Evidence Rule 702 broadly permitted expert testimony if the testimony would “help the trier of fact,”⁹ leaving unanswered whether the *Frye* standard of “general acceptability” remained as the threshold standard for what it meant to “help the trier of fact.”

Federal Rule of Evidence Rule 703 authorized experts, for the first time, to rely upon inadmissible evidence in forming their opinions “[i]f experts in the field would reasonably rely on those kinds of facts or data in forming an opinion on the subject.”¹⁰

5. *Id.* at 1014.

6. *Reed v. State*, 391 A.2d 364, 368 (Md. 1978) (“This criterion of ‘general acceptance’ in the scientific community has come to be the standard in almost all of the courts in the country which have considered the question of the admissibility of scientific evidence.”).

7. *People v. Kelly*, 549 P.2d 1240, 1244 (Cal. 1976) (“Some criticism has been directed at the *Frye* standard, primarily on the ground that the test is too conservative, often resulting in the prevention of the admission of relevant evidence.”).

8. Federal Rules of Evidence, Pub. L. 93-595, 88 Stat. 1926 (1975) (The United States Supreme Court adopted the Federal Rules of Civil Procedure in 1938 pursuant to the Rules Enabling Act, 28 U.S.C. Section 2072, in which Congress delegated to the judicial branch the authority to promulgate rules related to the procedures of the federal courts. The Supreme Court attempted to enact the Federal Rules of Evidence under The Rules Enabling Act, but Congress refused and instead enacted the Federal Rules of Evidence as a statute in 1975.).

9. FED. R. EVID. 702.

10. FED. R. EVID. 703.

Federal Rules of Evidence Rule 803(18) created an important hearsay exception especially designed for experts to include statements contained in a treatise, periodical, or pamphlet if:

(A) the statement is called to the attention of an expert witness on cross-examination or relied on by the expert on direct examination; and

(B) the publication is established as a reliable authority by the expert's admission or testimony, by another expert's testimony, or by judicial notice.

If admitted, the statement may be read into evidence but not received as an exhibit.¹¹

Notwithstanding the statutory changes related to expert testimony, the courts were confused about the standard for admissibility for expert testimony. Many federal courts interpreted Rule 702 as adopting the old common law *Frye* standard of "general acceptability."¹² Other courts and commentators suggested that the *Frye* standard had serious flaws and should only be one of many factors considered in admitting expert testimony.¹³ Some courts and commentators suggested that the *Frye* standard unduly restricted the admissibility of expert testimony.¹⁴ On the opposite end, some

11. FED. R. EVID. 803(18).

12. Andrea A. Moenssens, *Requiem for the "General Acceptance" Standard in Forensic Science*, in *LEGAL MEDICINE* 275, 279-80 (Cyril H. Wecht ed. 1982).

13. *United States v. Downing*, 753 F.2d 1224, 1237 (3d Cir. 1985).

[T]he *Frye* test suffers from serious flaws. The test has proved to be too malleable to provide the method for orderly and uniform decision-making envisioned by some of its proponents. Moreover, in its pristine form the general acceptance standard reflects a conservative approach to the admissibility of scientific evidence that is at odds with the spirit, if not the precise language, of the Federal Rules of Evidence. For these reasons, we conclude that "general acceptance in the particular field to which [a scientific technique] belongs," . . . should be rejected as an independent controlling standard of admissibility. Accordingly, we hold that a particular degree of acceptance of a scientific technique within the scientific community is neither a necessary nor a sufficient condition for admissibility; it is, however, one factor that a district court normally should consider in deciding whether to admit evidence based upon the technique.

Id. See also *State v. Chapple*, 660 P.2d 1208, 1218 (Ariz. 1983) (en banc); *United States v. Smith*, 736 F.2d 1103, 1105 (6th Cir. 1984) (per curiam); *People v. McDonald*, 690 P.2d 709, 723-24 (Cal. 1984) (en banc).

14. *United States v. Baller*, 519 F.2d 463, 466 (4th Cir. 1975) ("Unless an exaggerated popular opinion of the accuracy of a particular technique makes its use prejudicial or likely to

legal and scientific scholars began to express concern over “junk science”¹⁵ readily being admitted in trials under the too-permissive *Frye* standard.¹⁶ Other commentators suggested that admissibility under Rule 702 should be analyzed under the civil preponderance standard for civil cases and the beyond-a-reasonable-doubt standard for expert evidence offered against the accused.¹⁷ Uncertainty prevailed on what changes, if any, the newly enacted Federal Rules of Evidence required for expert testimony.

Everything changed dramatically in 1993 when the United States Supreme Court decided *Daubert v. Merrell Dow Pharmaceuticals, Inc.*¹⁸ In *Daubert*, parents, on behalf of infants, sued a pharmaceutical company to recover for birth defects allegedly caused by the mother’s ingestion of a “morning sickness” pill marketed under the name, “Bendectin.”¹⁹ The plaintiffs presented the testimony of eight qualified experts who each based their opinions that Bendectin presented a risk factor for birth defects based upon the unpublished “reanalysis” of previously-published human statistical studies.²⁰ The trial and appellate court excluded the expert testimony related to the proffered general causation theory on the basis that the theory of causation did not meet *Frye*’s “generally accepted” theory of admissibility.

Justice Blackmun delivered the opinion of the Court and decided several issues that have controlled the admissibility of expert testimony in the federal

mislead the jury, it is better to admit relevant scientific evidence in the same manner as other expert testimony and allow its weight to be attacked by cross-examination and refutation.”); see also *United States v. Sample*, 378 F. Supp. 44, 53 (E.D. Pa. 1974) (“The *Frye* test of general acceptance . . . precludes too much relevant evidence . . .”); *United States v. Stifel*, 433 F.2d 431, 438 (6th Cir. 1970) (“Every useful new development must have its first day in court. And court records are full of the conflicting opinions of doctors, engineers, and accountants, to name just a few of the legions of expert witnesses.”).

15. Kenneth Cheesbro, *Galileo’s Retort: Peter Huber’s Junk Scholarship*, 42 AM. U. L. REV. 1637, 1693-96 (1993).

16. Bert Black, *A Unified Theory of Scientific Evidence*, 56 FORDHAM L. REV. 595, 636-37; Roger S. Hanson, *James Alphonzo Frye is Sixty-Five Years Old; Should He Retire?*, 16 W. ST. U. L. REV. 357, 367-68 (1989); Frederick B. Lacey, *Scientific Evidence*, 24 JURIMETRICS 254, 265 (1984) (*Frye* jurisdictions will always lag behind advances in science); Joseph G. Petrosinelli, Note & Comment, *The Admissibility of DNA Typing: A New Methodology*, 79 GEO. L.J. 313, 317 (1990).

17. Paul C. Giannelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States, A Half Century Later*, 80 COLUM. L. REV. 1197, 1248-50 (1980) (criminal defendant and civil litigants should be required to establish the validity of the scientific principle or technique by a preponderance of the evidence; the prosecution in a criminal trial should be required to prove validity beyond a reasonable doubt).

18. *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993).

19. *Id.* at 582.

20. *Id.* at 583.

courts and most state courts ever since.²¹ First, the Court clearly held that Rule 702 did not adopt *Frye*'s "generally accepted" standard for admissibility.²² Second, under a combination of the relevancy standard contained in Rule 402 and the expert standard contained in Rule 702, the Court determined that the trial court has a "gatekeeping" responsibility to "ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable."²³ The Court explained that

in the event the trial court concludes that the scintilla of evidence presented supporting a position is insufficient to allow a reasonable juror to conclude that the position more likely than not is true, the court remains free to direct a judgment, Fed.Rule Civ.Proc. 50(a), and likewise to grant summary judgment, Fed.Rule Civ.Proc. 56.²⁴

Third, the Court observed that "[t]he primary locus of this obligation is Rule 702, which clearly contemplates some degree of regulation of the subjects and theories about which an expert may testify."²⁵ Fourth, the Court stated that the reliability and admissibility of expert testimony depends upon whether the theories and methodologies relied upon by the expert in forming their opinion have been (1) scientifically "tested,"²⁶ (2) the testing has been subject of peer-reviewed critique and publication,²⁷ (3) the testing has

21. As of November 23, 2018, Westlaw indicates that *Daubert* has been cited 143,324 times.

22. *Daubert*, 509 U.S. at 589 n.6 ("[W]e hold that *Frye* has been superseded and base the discussion that follows on the content of the congressionally enacted Federal Rules of Evidence ...").

23. *Id.* at 589 n.7.

24. *Id.* at 596.

25. *Id.* at 589.

26. *Id.* at 593 ("Scientific methodology today is based on generating hypotheses and testing them to see if they can be falsified; indeed, this methodology is what distinguishes science from other fields of human inquiry." (citing ERIC GREEN & CHARLES NESSON, PROBLEMS, CASES, AND MATERIALS ON EVIDENCE 645 (1983))); see also CARL HEMPEL, PHILOSOPHY OF NATURAL SCIENCE 49 (1966) ("[T]he statements constituting a scientific explanation must be capable of empirical test."); KARL POPPER, CONJECTURES AND REFUTATIONS: THE GROWTH OF SCIENTIFIC KNOWLEDGE 48 (5th ed. 1989) ("[T]he criterion of the scientific status of a theory is its falsifiability, or refutability, or testability." (emphasis omitted)).

27. *Daubert*, 509 U.S. at 593-94.

Another pertinent consideration is whether the theory or technique has been subjected to peer review and publication. . . . The fact of publication (or lack thereof) in a peer reviewed journal thus will be a relevant, though not dispositive, consideration in assessing the scientific validity of a particular technique or methodology on which an opinion is premised.

produced an “known” “rate of error,”²⁸ and (4) the “general acceptability” of the theories and methodologies (a deferential reference to the earlier *Frye* standard) is relevant, but neither necessary nor sufficient.²⁹

Daubert assigned the trial court the role of “gatekeeper” for unreliable expert testimony, which dramatically altered the judge’s role with respect to the admissibility of expert testimony.³⁰ Rather than answering a single question, “is it novel scientific evidence,” the judge was given the responsibility to review *Daubert*’s four-part test of “reliability” with respect to the underlying theories³¹ and methodologies,³² if not the conclusions they generated.³³

Given *Daubert*’s sea change for admitting expert testimony, the next question was the standard of review for a trial court’s expert testimony decision. The Supreme Court, in *General Electric Co. v. Joiner*,³⁴ extended broad discretion to trial courts by recognizing an “abuse of discretion”

Id.

28. *Id.* at 594.

Additionally, in the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error, *see, e.g., United States v. Smith*, 869 F.2d 348, 353-54 ([7th Cir.] 1989) (surveying studies of the error rate of spectrographic voice identification technique), and the existence and maintenance of standards controlling the technique’s operation, *see United States v. Williams*, 583 F.2d 1194, 1198 ([2nd Cir.] 1978) (noting professional organization’s standard governing spectrographic analysis), cert. denied, 439 U.S. 1117, (1979).

Id. at 594.

29. *Id.* (“Finally, ‘general acceptance’ can yet have a bearing on the inquiry. A ‘reliability assessment does not require, although it does permit, explicit identification of a relevant scientific community and an express determination of a particular degree of acceptance within that community.” (quoting *United States v. Downing*, 753 F.2d 1224, 1238 (3d Cir. 1985))); *see also* JACK B. WEINSTEIN ET AL., 4 WEINSTEIN’S FEDERAL EVIDENCE ¶ 702.03, 702-41-42 (2d ed. 2018).

30. *Id.* at 597 (“We recognize that, in practice, a *gatekeeping role* for the judge, no matter how flexible, inevitably on occasion will prevent the jury from learning of authentic insights and innovations.” (emphasis added)).

31. *Id.* at 593 (“Another pertinent consideration is whether the theory or technique has been subjected to peer review and publication.”).

32. *Id.* (“Scientific methodology today is based on generating hypotheses and testing them to see if they can be falsified; indeed, this methodology is what distinguishes science from other fields of human inquiry.” (citing GREEN & NESSON, *supra* note 26, at 645; *see also* HEMPEL, *supra* note 26, at 49 (“[T]he statements constituting a scientific explanation must be capable of empirical test.”); POPPER, *supra* note 26, at 48 (“[T]he criterion of the scientific status of a theory is its falsifiability, or refutability, or testability.” (emphasis deleted))).

33. *Daubert*, 509 U.S. at 595 (“The focus, of course, must be solely on principles and methodology, not on the conclusions that they generate.”).

34. *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 139 (1997).

standard of review for all *Daubert* trial-judge determinations. The Court then extended the *Daubert* analysis to experienced-based expert testimony in *Kumho Tire Co., Ltd. v. Carmichael*.³⁵ The Court also held in *Weisgram v. Marley Co.*³⁶ that an appellate court may order an entry of judgment for the verdict loser when a trial court has excluded evidence under *Daubert*, reinforcing the responsibility of counsel at trial to ensure that a sufficient *Daubert* foundation has been provided. Finally, the Court decided in *Cavazos v. Smith*,³⁷ a shaken-baby case, that once a trial court conducts a proper *Daubert* analysis and admits expert testimony, the weight of that testimony is for the jury.³⁸

The *Daubert* paradigmatic change for expert testimony in the federal courts dramatically impacted the issue of admissibility of expert testimony in both federal and state courts, making the subject one of the most commonly reviewed issues on appeal.³⁹ Following the Supreme Court's decision in *Daubert*, many states abandoned the *Frye* test and adopted the *Daubert* analysis as controlling under state law.⁴⁰ Many of the states that did not expressly adopt the *Daubert* standard relied on the *Daubert* decision for guidance in determining their own standards for the admissibility of expert

35. *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 141 (1999).

36. *Weisgram v. Marley Co.*, 528 U.S. 440, 443 (2000).

37. *Cavazos v. Smith*, 565 U.S. 2 (2011)

38. *Id.* at 8.

In light of the evidence presented at trial, the Ninth Circuit plainly erred in concluding that the jury's verdict was irrational, let alone that it was unreasonable for the California Court of Appeal to think otherwise. . . . Doubts about whether Smith is in fact guilty are understandable. But it is not the job of this Court, and was not that of the Ninth Circuit, to decide whether the State's theory was correct. The jury decided that question, and its decision is supported by the record.

Id.

39. As of November 23, 2018, Westlaw indicates that *Daubert* has been cited 143,324 times.

40. See, e.g., *Motorola Inc. v. Murray*, 147 A.3d 751, 758-59 (D.C. 2016) ("We adopt Rule 702 to apply to the trial of this case and to any civil or criminal case in which the trial begins after the date of this opinion."); *Schaferman v. Agland Coop*, 631 N.W.2d 862 (Neb. 2001); *State v. Porter*, 698 A.2d 739, 742 (Conn. 1997) (adopting the *Daubert* standard); *Commonwealth v. Lanigan*, 641 N.E.2d 1342, 1349 (Mass. 1994) ("We accept the basic reasoning of the *Daubert* opinion because it is consistent with our test of demonstrated reliability."); *State v. Moore*, 885 P.2d 457, 471 (Mont. 1994), *abrogated by State v. Gollehon*, 906 P.2d 697 (Mont. 1995), and *Billings v. Bruce*, 965 P.2d 866 (Mont. 1998) ("We conclude that the guidelines set forth in *Daubert* are consistent with our previous holding in *Barmeyer* concerning the admission of expert testimony of novel scientific evidence, and we, therefore, adopt the *Daubert* standard for the admission of scientific expert testimony.").

testimony.⁴¹ The few states that have retained the *Frye* single-factor standard of general acceptability have generally modified that standard.⁴²

41. See, e.g., *People v. Shreck*, 22 P.3d 68, 78-79 (Colo. 2001).

[W]e conclude that CRE 702, rather than *Frye*, represents the appropriate standard for determining the admissibility of scientific evidence. We hold that under this standard, the focus of a trial court's inquiry should be on the reliability and relevance of the scientific evidence, and that such an inquiry requires a determination as to (1) the reliability of the scientific principles; (2) the qualifications of the witness; and (3) the usefulness of the testimony to the jury. We also hold that when a trial court applies CRE 702 to determine the reliability of scientific evidence, its inquiry should be broad in nature and consider the totality of the circumstances of each specific case.

Id.; *Ingram v. State*, 699 N.E.2d 261, 262 (Ind. 1998) (“In determining reliability, while various factors have been identified, there is no specific ‘test’ or set of ‘prongs’ which must be considered in order to satisfy Indiana Evidence Rule 702(b).” (quoting *McGrew v. State*, 682 N.E.2d 1289, 1292 (Ind. 1997)));

Such factors may include, but are not limited to: 1) whether the technique has been or can be empirically tested; 2) whether the technique has been subjected to peer review and publication; 3) the known or potential rate of error, as well as the existence and maintenance of standards controlling the technique's operation; and 4) general acceptance within the relevant scientific community.

Ingram, 699 N.E.2d at 262 n.5 (quoting *McGrew*, 682 N.E.2d at 1292)); *State v. Hungerford*, 697 A.2d 916, 922 (N.H. 1997).

Specifically, we considered important the presence of objective, quantifiable evaluation results, the existence of a “logical nexus” between the expert's observations and conclusions, the verifiability of any interpretive steps, and the likely difficulty of effective cross-examination of the expert. . . . Also helpful are the considerations enunciated by the United States Supreme Court in *Daubert*. In applying Federal Rule of Evidence 702, the *Daubert* Court discussed four considerations bearing upon the reliability and helpfulness of scientific evidence: (1) whether the theory or technique has been or can be tested; (2) whether the theory or technique has been subjected to peer review and publication; (3) the potential or known error rate; and (4) whether there is general acceptance of the theory or technique in the relevant scientific community.

Id. (internal citations omitted); *Higgs v. State*, 222 P.3d 648, 658-59 (Nev. 2010).

By not adopting the *Daubert* standard as a limitation on judges' considerations with respect to the admission of expert testimony, we give Nevada trial judges wide discretion, within the parameters of NRS 50.275, to fulfill their gatekeeping duties. We determine that the framework provided by NRS 50.275 sets a degree of regulation upon admitting expert witness testimony, without usurping the trial judge's gatekeeping function.

....

. . . In sum, *Daubert*, as any other case decided by the U.S. Supreme Court, is looked upon favorably by this court. We do not, however, adopt the *Daubert* standard as a limitation on the factors considered for admissibility of expert witness testimony. We hold that NRS 50.275 provides the standard for admissibility of expert witness testimony in Nevada.

Almost simultaneously, and largely independently of the *Daubert* revolution in the law of evidence, the medical teaching and practice communities began adopting a paradigm shift known as “evidence-based medicine.” The term “evidence-based medicine” was first coined in 1991, two years before *Daubert* was decided.⁴³ The “evidence-based medicine working group” coined the phrase and identified the “paradigm shift” for the practice and teaching of medicine as follows:

A new philosophy of medical practice and teaching has followed these methodological advances. This paradigm shift is manifested in a number of ways. A profusion of articles has been published instructing clinicians on how to access, evaluate, and interpret the medical literature. Proposals to apply the principles of clinical

Id.; *McDaniel v. CSX Transp., Inc.*, 955 S.W.2d 257, 265 (Tenn. 1997).

[W]e conclude that Tennessee’s adoption of Rules 702 and 703 in 1991 as part of the Rules of Evidence supersede the general acceptance test of *Frye*. In Tennessee, under the recent rules, a trial court must determine whether the evidence will substantially assist the trier of fact to determine a fact in issue and whether the facts and data underlying the evidence indicate a lack of trustworthiness. The rules together necessarily require a determination as to the scientific validity or reliability of the evidence. Simply put, unless the scientific evidence is valid, it will not substantially assist the trier of fact, nor will its underlying facts and data appear to be trustworthy, but there is no requirement in the rule that it be generally accepted.

Although we do not expressly adopt *Daubert*, the non-exclusive list of factors to determine reliability are useful in applying our Rules 702 and 703. A Tennessee trial court may consider in determining reliability: (1) whether scientific evidence has been tested and the methodology with which it has been tested; (2) whether the evidence has been subjected to peer review or publication; (3) whether a potential rate of error is known; (4) whether, as formerly required by *Frye*, the evidence is generally accepted in the scientific community; and (5) whether the expert’s research in the field has been conducted independent of litigation.

Id.

42. *Sargon Enters., Inc. v. Univ. of S. Cal.*, 288 P.3d 1237, 1252 (Cal. 2012) (Although previously following the *Frye* standard, the court explained that the trial court, in considering expert testimony, should “conduct[] a ‘circumscribed inquiry’ to ‘determine whether, as a matter of logic, the studies and other information cited by experts adequately support the conclusion that the expert’s general theory or technique is valid.’ The goal of trial court gatekeeping is simply to exclude ‘clearly invalid and unreliable’ expert opinion.” (internal citations omitted)).

43. Gordon Guyatt et al., *Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine*, 268 J. AM. MED. ASS’N 2420 (1992); see also Benjamin Djulbegovic & Gordon Guyatt, *Progress in Evidence-Based Medicine: A Quarter Century On*, 390 LANCET 415 (2017).

epidemiology to day-to-day clinical practice have been put forward. A number of major medical journals have adopted a more informative structured abstract format, which incorporates issues of methods and design into the portion of an article the reader sees first. The American College of Physicians has launched a journal, ACP Journal Club, that summarizes new publications of high relevance and methodological rigor. Textbooks that provide a rigorous review of available evidence, including a methods section describing both the methodological criteria used to systematically evaluate the validity of the clinical evidence and the quantitative techniques used for summarizing the evidence, have begun to appear. Practice guidelines based on rigorous methodological review of the available evidence are increasingly common. A final manifestation is the growing demand for courses and seminars that instruct physicians on how to make more effective use of the medical literature in their day-to-day patient care. We call the new paradigm "evidence-based medicine."⁴⁴

The similarity and timing of the paradigm shift to evidence-based medicine and the Court's paradigm shift in *Daubert* to assessing the reliability of the "principles and methodologies" is striking and compatible, but the overlapping coherence of the two separate domains has seldom been recognized by either profession.

This paper discusses how the paradigm shift of evidence-based-medicine and the tools developed in that genre provide an important basis at every stage of any expert's "health related" testimony under the *Daubert* standard of admissibility.⁴⁵

II. DISCOVERY OBLIGATIONS: FEDERAL RULES OF EVIDENCE RULE 26(B)(2)

A. *Qualification of the Expert*

The first foundational step for admitting any expert testimony under Federal Rule of Evidence Rule 702 is the qualification of the expert by education, training, or experience to give relevant expert testimony that "will assist the trier of fact."⁴⁶ For a medical malpractice claim, expert testimony is

44. Guyatt, *supra* note 43, at 2421.

45. Brief for Wisconsin Medical Society as Amicus Curiae Supporting Appellant, *Bayer ex rel. Petrucelli v. Dobbins*, 885 N.W.2d 173 (Wis. Ct. App. 2016) (No. 2015AP1470), 2015 WL 9596321.

46. *Diefenbach v. Sheridan Transp.*, 229 F.3d 27, 30 (1st Cir. 2000).

typically required as part of the prima facie case.⁴⁷ If the jurisdiction by statute or common law follows the “local practice” rule, an expert will have to have experience and training with respect to the local practice to be qualified to give testimony on the “local standard of care.”⁴⁸ Additionally, “[p]ractitioners in one specialty are not ordinarily competent to testify as experts on the standard of care applicable to another specialty.”⁴⁹ For example, an orthopedic surgeon would be qualified to testify about subjects within the ambit of the field of orthopedic surgery, but would not be qualified to testify about the standard of care for a nurse over which he had supervisory responsibility, due to the “wide variation between schools in both precepts and practices.”⁵⁰ In assessing qualification, the expert’s familiarity with the principles of evidence-based medicine is critical when addressing health-related opinions.

B. Principles or Theories of Reliability

Once an expert is qualified in a relevant area under investigation, the next *Daubert* question is the reliability of the underlying principles or theories upon which the expert relied. Many attorneys have a difficult time formulating the principle or theory upon which health-related questions are framed. Theories can be both general and specific with respect to both the differential diagnosis and etiology.

The most general theory question for health-related questions, should be:

It is well-settled that trial judges have broad discretionary powers in determining the qualification, and thus, admissibility, of expert witnesses. It is settled law in this circuit that [w]hether a witness is qualified to express an expert opinion is a matter left to the sound discretion of the trial judge. In the absence of clear error, as a matter of law, the trial judge’s decision will not be reversed.

Id. (quoting *Richmond Steel Inc. v. Puerto Rican Am. Ins. Co.*, 954 F.2d 19, 21 (1st Cir. 1992)) (internal quotations omitted).

47. *Blevens v. Halcomb*, 469 F.3d 692, 695 (8th Cir. 2006) (“Testimony that a physician failed to live up to some vague ‘standard of care,’ without elaboration as to the content of that standard, is insufficient to satisfy this burden.” (citing *Ladish v. Gordon*, 879 S.W.2d 623, 634 (Mo. Ct. App. 1994))).

48. *Jerden v. Amstutz*, 430 F.3d 1231, 1236 (9th Cir. 2005) (“The specific basis for Defendant’s objection to Dr. Gross’s testimony should have been asserted when Dr. Gross testified as to Defendant’s compliance with his standard of care without reciting that he had knowledge of the proper medical conduct within Defendant’s community.”).

49. *Nguyen v. IHC Health Servs., Inc.*, 232 P.3d 529, 535 (Utah Ct. App. 2010) (holding a general pediatric physician was not qualified to give testimony on breach of the standard of care of a critical care physician or to quantify the chance of survival absent the ventilator failure).

50. *De Adder v. Intermountain Healthcare, Inc.*, 308 P.3d 543, 549 (Utah Ct. App. 2013).

Q: What theory or principle did you rely upon in forming your opinion?

A: I relied upon the theory or principle of evidence-based medicine which helped inform both my general and specific causation theories of the case.⁵¹

The follow-up *Daubert* questions and answers related to this question should be:

Q: Has the theory of evidence-based medicine been tested?

A: Yes. Each of the studies relied upon by the expert have been subjected to peer-reviewed analysis under the principles of evidence-based medicine.⁵²

Q: Has the testing been subject to peer-reviewed critique?

A: Yes. The theory of evidence-based medicine, as a paradigm for expert practice, practice guidelines, and expert testimony, was first coined in 1992, and has since been subject to countless articles and papers in health-related areas.⁵³

Q: Has the theory been subject to peer-reviewed critique?

A: Yes, since 1992 medical research has been replete with discussion of, further elucidation, and refinement of the theory of evidence-based medicine. Any Google search of the theory of evidence-based medicine will demonstrate the depth of the scientific literature validating evidence-based medicine as the dominate theory of all health-related training and research.⁵⁴

Q: Is there an established rate of error?

A: Yes. The underlying premise of evidence-based medicine is that medical hypotheses should be tested and evaluated by empirically-based studies formulated to understand the appropriate confidence levels of the underlying medical assumptions or practice guidelines. The established rate of error in each instance is dependent upon the quantity and quality of the

51. *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 590 (1993) (“The adjective ‘scientific’ implies a grounding in the methods and procedures of science.”).

52. *Id.* at 593 (1993) (“Ordinarily, a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact will be whether it can be (and has been) tested.”).

53. *Id.*

54. *Id.* at 593 (“Another pertinent consideration is whether the theory or technique has been subjected to peer review and publication.”).

studies relied upon in forming the specifically applicable medical-related opinion.⁵⁵

Q: Is the theory of “evidence-based medicine” generally accepted in the relevant health-related community?

A: Yes, evidence-based medicine has become the norm of medical education, practice training, and practice guidelines, as can be demonstrated by reference to any medical curriculum.⁵⁶

The more specific diagnosis and etiology theories would be more fact-specific.

C. Methodologies

After asking about the principles or theories underling the expert’s opinion, the next question on the *Daubert* checklist is whether the methodologies relied upon by the expert in forming his or her opinion satisfy the same *Daubert* reliability standards of “testability,” “peer reviewed critique,” “established rate of error,” and “general acceptability”?

Several of the most “generally accepted” methodologies for evaluating health-related conditions include SOAP, differential diagnosis, and differential etiology.

Under the SOAP methodology, the physician (1) considers the *subjective* complaints of the patient, (2) investigates the probabilities by *objective* tests, (3) *analyzes* the information through a differential diagnosis, and (4) adopts a *plan* of treatment.⁵⁷

Under the well-established methodologies of differential diagnosis and differential etiology, the health investigator considers all the alternatives suggested by the subjective complaints and objective tests, analyzes the information provided, rules out the least-likely diagnosis or cause, and arrives at a plan or conclusion.⁵⁸

In performing the ruling in and ruling out of the differential diagnosis or etiology “methodologies,” the answer should be case-specific, depending

55. *Id.* at 594 (“Additionally, in the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error . . .”).

56. *Id.* (“Finally, ‘general acceptance’ can yet have a bearing on the inquiry.”).

57. LYNN S. BICKLEY, *BATES’ GUIDE TO PHYSICAL EXAMINATION AND HISTORY TAKING* (12th ed. 2017).

58. *King v. Burlington N. Santa Fe Ry. Co.*, 277 Neb. 203, 237, 762 N.W.2d 24, 50 (2009) (“If an expert’s general causation opinion is admissible to show that a suspected agent should be ruled in as a possible cause of the plaintiff’s disease, the court must next determine whether the expert performed a reliable differential etiology.”).

upon the form of the case study relied upon and its ranking on the evidence-based medicine pyramid.

The evidence-based methodology for ruling in and ruling out the possibilities to perform the differential diagnosis or etiology has four steps: (1) finding the appropriate research study, (2) searching the literature, (3) ranking the evidence-based levels of reliability, and (4) critiquing the study using the principles of evidence-based medicine.⁵⁹

1. The First Methodological Step: Finding the Appropriate Research Study that Formulated a Relevant Question

Medical researchers first formulate the clinical problem being investigated.⁶⁰ Expert witnesses seeking evidence-based support for their opinions, investigate the “fit” of the extant studies on the same issue. If no study fits precisely under the question being investigated, then the closest study should be examined with full recognition that an “unfair factual extrapolation”⁶¹ argument may be fairly made. The Federal Advisory Committee Notes on Rule 702 suggest that whether the relevant study is “litigation-dependent” should be considered when deciding admissibility.⁶²

As a methodology for formulating the research question, both the researcher and the expert must first define precisely whom the question is about by asking, “How would I describe a group of patients similar to this one?” Demographic features such as age, gender, and race describe the patient group. Medical risk factors and conditions also define the patient group. Second, the researcher or expert should describe the intervention/maneuver that will be performed or has been performed on the patient population. Third, if necessary, the researcher or expert should provide or investigate a comparison maneuver. For example, the intervention can be the administration of a drug or diagnostic radiologic exam and the comparison can be a placebo or an alternative diagnostic exam. Fourth, the researcher or expert should define the outcome such as reduced mortality or improved quality of life.

59. TRISHA GREENHALGH, HOW TO READ A PAPER: THE BASICS OF EVIDENCE-BASED MEDICINE 39 (5th ed. 2014).

60. *Id.*

61. *Daubert*, 509 U.S. at 591 (“‘Fit’ is not always obvious, and scientific validity for one purpose is not necessarily scientific validity for other, unrelated purposes.”).

62. FED. R. EVID. 702 advisory committee’s note (In 2000, the Advisory Committee Notes to Rule 702 identified as a bias factor related to the issue of admissibility and impeachment, the “(1) whether” the testimony relates to “matters growing naturally and directly out of research they have conducted independent of the litigation,” or developed “expressly for purposes of testifying.”).

These four steps formulating the question under investigation are commonly known by the “PICO” mnemonic: patient group, intervention, comparison and outcome.⁶³ Foundation for the PICO methodology can be established through the same *Daubert* checklist:

Q: Has the PICO methodology for formulating a medical research question been tested?

A: Yes. Each of the studies relied upon by the expert were the product of a PICO analysis consistent with the premises of evidence-based medicine.

Q: Has the specific PICO study been subject to peer-reviewed critique?

A: Yes. This study was published in a reputable scientific journal.

Q: Is the journal a peer-reviewed journal?

A: Yes.

Q: Did the study discuss any probability analysis for the results?

A: Yes. The study had a probability analysis published as part of the research.

Q: Are the conclusions of the study generally-accepted in the relevant scientific community?

A: Yes.

The “reliability” or “general acceptability” of the study depends, in part, upon whether these steps have been adequately taken. Other considerations include the quantity and quality of the study. Finally, the question in litigation is whether the study is sufficiently analogous to the question being litigated that the expert can fairly extrapolate from the study and the theory of general causation to the facts of the case or the issue of specific causation.

2. Second Step: Searching the Literature

After precisely formulating the clinical problem much like the original researcher, the expert, consistent with evidence-based medicine, should perform the appropriate literature search.⁶⁴ Federal Rule of Evidence Rule 703 enables an expert to rely on inadmissible evidence, such as research studies, if the evidence is reasonably relied upon by experts in the field.⁶⁵ Also, Federal Rule of Evidence Rule 803(18) permits an expert to read to the trier of fact the content of “learned treatises” if they are reasonably relied upon by

63. Lisa P. Lavelle et al., *Evidence-Based Practice of Radiology*, 35 RADIOGRAPHICS 1802, 1804 (2015); Marie Staunton, *Evidence-Based Radiology: Steps 1 and 2—Asking Answerable Questions and Searching for Evidence*, 242 RADIOLOGY 23, 24 (2007).

64. *Id.* at 25.

65. FED. R. EVID. 703.

experts in the field.⁶⁶ These evidentiary rules separately affirm the admissibility of “reliable” excerpts from “learned treatises” and specify the method the information will be communicated to the trier of fact, both as a basis for expert testimony and as independent evidence orally presented to the jury.

Google Scholar and PubMed search the web for useful primary scientific literature containing both the studies and the peer reviewed critique referenced by the Supreme Court in *Daubert*. For quick clinical searches, Google Scholar returns more relevant articles and provides greater access to free full-text articles.⁶⁷ PubMed is more valuable than Google Scholar when performing a more thorough primary literature review.⁶⁸ The secondary literature summarizes the primary literature often through systematic reviews, a highly ranked form of evidence under the evidence-based medicine paradigm.

The Cochrane and DARE databases search systematic reviews. These reviews have a higher level of evidence and, thus, have more appeal in the courtroom. A thorough literature review can be a time-consuming process and can be outside of the expertise of many physicians. For this reason, it may be worthwhile to employ a health librarian to provide a professional-literature search. The question of whether an expert did an adequate literature review is an important question for either direct or cross-examination.

Search Engine	URL	Contents
Google Scholar	https://scholar.google.com	Primary Literature
PubMed	www.ncbi.nlm.nih.gov/pubmed/	Primary Literature
Cochrane Database	http://www.thecochranelibrary.com	Systematic Reviews
DARE Database	http://www.crd.york.ac.uk/crdweb/	Systematic Reviews
TRIP Database	https://tripdatabase.com	Primary and Secondary Literature

66. FED. R. EVID. 803(18).

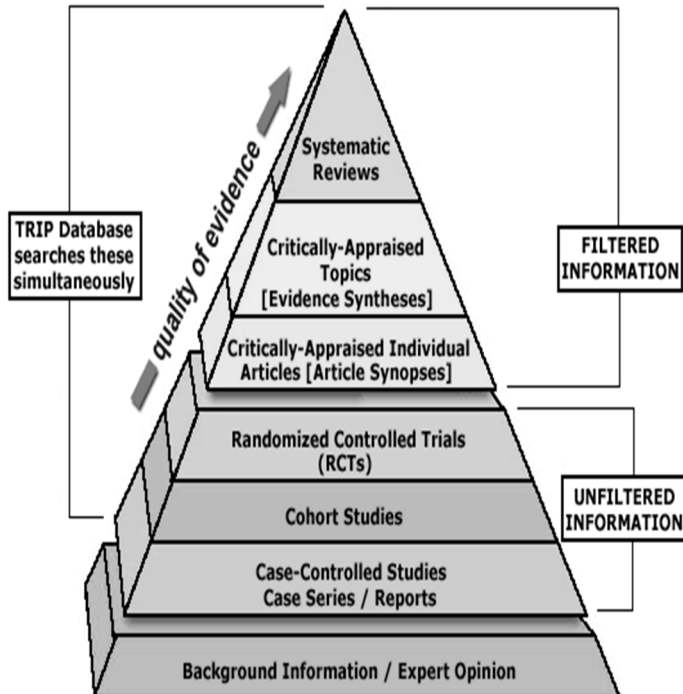
67. Salimah Shariff et al., *Retrieving Clinical Evidence: A Comparison of PubMed and Google Scholar for Quick Clinical Searches*, 15 J. MED. INTERNET RES. E164 (2013); Austin v. Am. Ass'n of Neurological Surgeons, 253 F.3d 967, 971 (7th Cir. 2001) (“[T]here is an abundance of up-to-date relevant literature easily retrievable from the World Wide Web.”).

68. Michael Anders & Dennis Evans, *Comparison of PubMed and Google Scholar Literature Searches*, 55 RESPIRATORY CARE 578, 582(2010); Wichor M. Bramer et al., *The Comparative Recall of Google Scholar Versus PubMed in Identical Searches for Biomedical Systematic Reviews*, 2 SYSTEMATIC REVIEWS 115 (2013), <https://doi.org/10.1186/2046-4053-2-115>.

Once an article of interest is discovered, both a researcher and an expert in review should read the articles that have cited the lead article. These cited articles may agree or disagree with the findings. Support or criticism would be relevant to direct or cross-examination. Citation chaining can be performed easily with Google Scholar. Once an article is found, click on the “cited by” feature to find the citation chain.

3. Third Step: Ranking the “Evidence-based” Levels of Reliability

The third step in evidence-based medicine is to lexically order the epistemological strength of health-related decision-making and practice guidelines to the strength of the supporting empirically-based research. The lexical ordering of the studies, often depicted by a “pyramid of reliability,” ranks in descending order of reliability empirically-based meta-analyses, systematic reviews, randomized controlled trials, case-control studies, anecdotal experiences or individualized case-studies, animal studies, and, finally, in vitro studies. The lexical ordering has been memorialized by various versions of the below depicted evidence-based medicine pyramid:



The Oxford Center for Evidence-Based Medicine created a heuristic device, the “Levels of Evidence,” that gives a quick assessment of a research study’s level of trustworthiness, irrespective of the journal of origin.⁶⁹ Expert witnesses should always consider the “Levels of Evidence” associated with any research relied upon. Research studies with a study design ranked at the top of the ranking system are held to be more reliable than those at the bottom of the ranking system. Originally, one universal rating system was created to rank all types of studies. However, experience taught that such a simplified rating system overgeneralizes. For example, the study type that best determines the effectiveness of a new drug would not be equally effective in assessing the quality of a diagnostic test. To adjust to this complexity, the current ranking system of Level of Evidence varies depending on the research question being asked (see attached tables).

Table 1 of 2.⁷⁰

Question	Step1 (Level1*)	Step2 (Level2*)	Step3 (Level3*)
How common is the problem?	Local and current random sample surveys (or censuses)	Systematic review of surveys that allow matching to local circumstances**	Local non-random sample**
Is this diagnostic or monitoring test accurate? (Diagnosis)	Systematic review of cross sectional studies with consistently applied reference standard and blinding	Individual cross sectional studies with consistently applied reference standard and blinding	Non-consecutive studies, or studies without consistently applied reference standards**
What will happen if we do not add a therapy? (Prognosis)	Systematic review of inception cohort studies	Inception cohort studies	Cohort study or control arm of randomized trial*

69. Jonathan D. Dodd, *Evidence-Based Practice in Radiology: Steps 3 and 4—Appraise and Apply Diagnostic Radiology Literature*, 242 *RADIOLOGY* 342, 345-46 (2007); *OCEBM Levels of Evidence*, OXFORD CTR. FOR EVIDENCE-BASED MED. (May 1, 2016), <https://www.cebm.net/2016/05/ocebml-levels-of-evidence/>.

70. The table can be viewed in its PDF format at <http://www.cebm.net/wp-content/uploads/2014/06/CEBM-Levels-of-Evidence-2.1.pdf>.

Question	Step1 (Level1*)	Step2 (Level2*)	Step3 (Level3*)
Does this intervention help? (Treatment Benefits)	Systematic review of randomized trials or <i>n</i> -of-1 trials	Randomized trial or observational study with dramatic effect	Non-randomized controlled cohort/follow-up study**
What are the COMMON harms? (Treatment Harms)	Systematic review of randomized trials, systematic review of nested case-control studies, <i>n</i> - of 1 trial with the patient you are raising the question about, or observational study with dramatic effect	Individual randomized trial or (exceptionally) observational study with dramatic effect	Non-randomized controlled cohort/follow-up study (post-marketing surveillance) provided there are sufficient numbers to rule out a common harm. (For long-term harms the duration of follow-up must be sufficient.)*
What are the RARE harms? (Treatment Harms)	Systematic review of randomized trials or <i>n</i> -of-1 trial	Randomized trial or (exceptionally) observational study with dramatic effect	
Is this (early detection) test worthwhile? (Screening)	Systematic review of randomized trials	Randomized trial	Non-randomized controlled cohort/follow-up study**

Table 2 of 2.⁷¹

Question	Step 4 (Level 4*)	Step 5 (Level 5)
How common is the problem?	Case-series**	n/a
Is this diagnostic or monitoring test accurate? (Diagnosis)	Case-control studies, or “poor or non-independent reference standard**	Mechanism-based reasoning
What will happen if we do not add a therapy? (Prognosis)	Case-series or case-control studies, or poor quality prognostic cohort study**	n/a
Does this intervention help? (Treatment Benefits)	Case-series, case-control studies, or historically controlled studies**	Mechanism-based reasoning
What are the COMMON harms? (Treatment Harms)	Case-series, case-control, or historically controlled studies**	Mechanism-based reasoning
What are the RARE harms? (Treatment Harms)		
Is this (early detection) test worthwhile? (Screening)	Case-series, case-control, or historically controlled studies**	Mechanism-based reasoning

*Level may be graded down on the basis of study quality, imprecision, indirectness (study PICO does not match questions PICO), because of inconsistency between studies, or because the absolute effect size is very small; Level may be graded up if there is a large or very large effect size.

**As always, a systematic review is generally better than an individual study.⁷²

The physician or expert uses the clinical question to find the appropriate row in Oxford’s table. The physician or expert then searches for articles that receive the level 1 score. If the question is about the efficacy of a therapy, then a systematic review of a randomized controlled trial is preferred. If one wants to demonstrate the validity of a diagnostic test, then use a systematic review

71. The table can be viewed in its PDF format at <http://www.cebm.net/wp-content/uploads/2014/06/CEBM-Levels-of-Evidence-2.1.pdf>.

72. *OCEBM Levels of Evidence*, *supra* note 69.

of a cross-sectional survey. For causation, use a systematic review of cohort studies. If no level 1 evidence is found, then the physician or expert proceeds down the row to level 2, then level 3, and so on. The Oxford's Evidence Based Medicine Table lexically orders the reliability of studies from least reliable to most reliable.

a. Lowest Levels of Evidence: Anecdotal Expert Opinions or Case Reports

Science values the opinion of an expert, whether it is based on mechanistic reasoning or based on clinical experience, for this opinion often serves as the basis to start an empiric investigation. However, an expert's "experiential" opinion without evidence-based research studies to validate or refute, is ordered the least reliable of all expert opinions, especially where confirming or refuting studies are available, as in the *Austin* case below. A brief written by the Wisconsin Medical Society to the court argues:

[T]he Society asks this Court to recognize that medical opinions that are supported solely by unsystematic clinical observations presumptively fail to cross the *Daubert* reliability threshold. These include both a physician's subjective beliefs based solely on their personal credentials and experience as well as medical literature identified as case reports. While such evidence may be properly part of physicians' decision-making process, it lacks an objective methodology on which physicians, and in turn courts, can rely. This is not to say that physician experience and clinical observations have no place in expert testimony. Training and experience plays a large role in medical analysis and decision making, and that training and experience can be incorporated into expert testimony. However, from a medical perspective, conclusions based solely on experience and anecdote are regarded with suspect in their application to other patients and circumstances.⁷³

In *Austin v. American Association of Neurological Surgeons*,⁷⁴ the claimant, Dr. Austin, sued his medical society for suspending him for "irresponsible expert testimony."⁷⁵ In denying Dr. Austin's claim, the United States Court of Appeals for the Seventh Circuit explained,

73. Brief for Wisconsin Medical Society as Amicus Curiae Supporting Appellant, Bayer *ex rel.* Petrucelli v. Dobbins, 885 N.W.2d 173 (Wis. Ct. App. 2016) (No. 2015AP1470), 2015 WL 9596321, at *9.

74. *Austin v. Am. Ass'n of Neurological Surgeons*, 253 F.3d 967 (7th Cir. 2001).

75. *Id.* at 971.

[T]here is little doubt that his [anecdotally-based] testimony was irresponsible and that it violated a number of sensible-seeming provisions [Including the provision that a medical society member] should provide the court with accurate and documentable opinions on the matters at hand.⁷⁶

What makes this case interesting is that the court performed its own evidence-based medicine literature search on the topic that would have placed the anecdotal testimony in an improved context. The court chastised the expert witnesses for not doing the literature search that would have enhanced the reliability of the expert testimony:

Oddly, apart from Cloward's article, and the Watkins article of unknown provenance (unknown to the lawyers, that is), no literature on anterior cervical fusion or injuries to the recurrent laryngeal nerve was presented either to the Association's hearing board or to the district court, although some additional literature had been presented at the malpractice trial and there is an abundance of up-to-date relevant literature easily retrievable from the World Wide Web. There we discover in a cursory search that permanent damage to the recurrent laryngeal nerve is a known though fortunately rare complication of anterior cervical fusion (a 1982 study found only 52 cases of paralysis to the recurrent laryngeal nerve in 70,000 such operations—.07 percent) against which the patients should be warned.⁷⁷

In *Berk v. St. Vincent's Hosp. and Medical Center*,⁷⁸ a patient claimed that Defendant's negligence resulted in a delay in diagnosis of a post-operative septic knee joint resulting in exacerbation of knee joint destruction. The plaintiff's expert witness, Depuy, based his testimony exclusively on his own extensive clinical experience of having performed over two-thousand arthroscopic surgeries. In excluding this experienced-based anecdotal testimony as too low on the evidence-based medicine pyramid, the court explained:

Depuy's conclusion, which appears to be based on no scientific support other than his own personal experience of not having encountered instances of fluid draining from knees of patients on whom he has operated, bears none of the hallmarks of reliability necessary for it to be considered admissible under *Daubert* and

76. *Id.* (internal quotations omitted).

77. *Austin*, 253 F.3d at 970-71.

78. *Berk v. St. Vincent's Hosp. & Med. Ctr.*, 380 F. Supp. 2d 334 (S.D.N.Y. 2005).

Rule 702 [O]f the basic four criteria on which a *Daubert* reliability analysis typically rests, two are particularly relevant in cases in which the expert testimony is based on personal experience: the rate of error of the experience-based *methodology* and “whether such a method is generally accepted in the scientific community.” An anecdotal account of one expert’s experience, however extensive or impressive the numbers it encompasses, does not by itself equate to a methodology, let alone one generally accepted by the relevant professional community.⁷⁹

b. Second Lowest Level of Reliability: Case Reports and Case Series

The Oxford’s Evidence Based Medicine Table ranks a case report or a case series as the second lowest level of reliability for evidence-based medicine. A case report gives a detailed description of the symptoms, signs, diagnosis, treatment and follow-up of an individual patient. The value of a case report is in describing a rare or previously unreported event. This report might sensitize readers and facilitate detection of similar cases leading to the reporting of a case series, a study that organizes multiple case reports around a central theme.⁸⁰ A case series can be used to generate hypotheses for more rigorous research studies, but alone is nothing more than the reporting of an unreliable anecdotal experience.

Case reports and case series have many limitations that result in their low ranking on the tier of Levels of Evidence. First, these reports and series are not chosen from representative populations and consequently cannot generate information on incidence or prevalence of disease. Second, case studies have no control groups; a large case series can imply a causal relationship, but without a control, these case series cannot be relied upon to determine a causal relationship.⁸¹

This low estimation of case studies holds for both evidence-based medicine and judicial analysis of case studies. For example, in *Siharath v. Sandoz Pharmaceuticals Corporation*,⁸² a post-partum woman suffered a hemorrhagic stroke after taking a prescription drug manufactured by the defendant.⁸³ The plaintiff tried to prove causality through case reports. In granting a summary judgment, the court explained that Plaintiff’s

79. *Id.* at 354 (internal citation omitted).

80. Trygve Nissen & Rolf Wynn, *The Clinical Case Report: A Review of its Merits and Limitations*, 7 BMC RES. NOTES 264 (2014).

81. *Id.*

82. *Siharath v. Sandoz Pharm. Corp.*, 131 F. Supp. 2d 1347 (N.D. Ga. 2001).

83. *Id.* at 1349, *aff’d sub nom.*; *Rider v. Sandoz Pharm. Corp.*, 295 F.3d 1194 (11th Cir. 2002).

experts (1) have failed to provide any evidence, either published or unpublished, that Parlodel® increases one's risk of stroke; (2) rely on uncontrolled and unreliable spontaneous reports and anecdotal case reports as the basis for their opinions; and (3) cannot show that their opinions have an acceptable error rate or are otherwise generally accepted.⁸⁴

The court further explained that Plaintiff did not satisfy their burden of proof:

No evidence has been offered of an increase in postpartum strokes after the drug was approved for suppression of lactation; no evidence has been offered of a decrease in postpartum strokes after the approval for suppression of lactation was withdrawn. The absence of epidemiological support raises the question of whether the causation opinions of Plaintiffs' experts are merely speculative and not based on scientific knowledge.⁸⁵

The court further dissected the case reports the Plaintiff relied upon as an unfair extrapolation:

The [alleged] adverse drug reports . . . lack the requisite quantity, nature and content. From 1980 to 1994, millions of women took Parlodel®. The modest number of case reports associating the drug with stroke or even postpartum hypertension is not what would be expected if there was a significant increased risk. Only one report exists that links Parlodel® to a stroke, and in that case the patient suffered from an underlying condition that itself can cause stroke. No other patient in any case reports suffered any form of stroke. The other patients instead suffered non-cerebral effects such as hypertension and myocardial infarction. Many of the case reports cited involved patients who were not postpartum. One case report involved a patient who was dechallenged but continued to suffer from hypertension for another four to five days. In short, Plaintiffs' [expert has] not pointed to a single case report involving a postpartum woman who suffered a hemorrhagic stroke. Accordingly, even if case reports could be used to establish general causation, any reasonable observer would have to conclude that they are insufficient to do so in this case. The case reports simply lack the quantity, nature and content that

84. *Siharath*, 131 F. Supp. 2d at 1352.

85. *Id.* at 1358 (internal citations omitted).

Dr. Dukes himself claims is necessary for case reports to provide reliable scientific information about causation.⁸⁶

Finally, the court made the point that even Plaintiff's expert on cross-examination conceded that case reports can never establish general causation:

Both of Plaintiffs' experts who testified at the *Daubert* hearing recognize the severe limitations of case reports and differential diagnosis in establishing general causation. Dr. Kulig admitted the limitations in the following exchange:

Q: As a matter of scientific methodology, Dr. Kulig, case reports do not establish general causation and you would never attempt to do so, true?

A: True.

Q: And as a matter of scientific methodology, Dr. Kulig, case series do not establish general causation and you would never attempt to do so, true?

A: True.

Q: And as a matter of scientific methodology, Dr. Kulig, differential diagnosis as applied to a specific patient cannot establish general causation?

A: In and of itself, I wouldn't establish it, but now you're getting closer.

Case reports can establish only specific causation. Testimony regarding specific causation, however, is irrelevant unless general causation is established. Accordingly, given the limits of case reports in establishing general causation, as recognized by Plaintiffs' experts, the Court must conclude that Plaintiffs' reliance upon case reports as a substitute for epidemiology cannot withstand the scrutiny that *Daubert* requires.⁸⁷

86. *Id.* at 1361.

87. *Id.* at 1363 (internal citations omitted).

Thus, case studies are generally inadequate sources for establishing general causation under evidence-based medicine.

c. Third Level or Mid-Tier Levels of Evidence: Observational Studies (Case-Control Studies)

A case-control study compares a group of patients with a disease with another group of patients without a disease to determine whether a potential causal attribute is more associated with the disease group than the control group. A clarifying example occurred in history when Drs. Doll and Hill studied whether smoking causes lung cancer. For every newly admitted patient with lung cancer, they selected another patient of the same age and gender but with a sickness other than lung cancer. They then asked each group whether or not they smoked. The results revealed a statistically significant relationship between smoking and lung cancer.⁸⁸

As implied by their relatively low status of “mid-tiered” on Oxford’s Level of Evidence, case-control studies have limitations. One of the biggest limitations of case-control studies is that they can show a correlation but they cannot prove causation.⁸⁹ For this reason, Dr. Doll quickly followed his case-control study with a more reliable case-cohort study, the natural progression of evidence-based research.⁹⁰ Cohort studies compare groups who have been “exposed” to an agent at issue, with groups who have not been exposed.⁹¹ The study compares each group’s rate of disease with or without exposure.⁹² Case-control studies, in comparison, compare a group with a disease with a group without the disease and investigate past exposures to determine whether an association exists between the exposure and the incidence of the disease.⁹³ However, an association, by itself, cannot establish causation.⁹⁴ These observational studies only show the “degree of statistical relationship between two or more events or variables. Events are said to be associated when they occur more or less frequently together than one would expect by chance.”⁹⁵ An “association” can be quantified by statistical analysis into

88. Richard Doll & A. Bradford Hill, *Smoking and Carcinoma of the Lung*, 2 BRIT. MED. J. 739, 740, 746 (1950).

89. TRISHA GREENHALGH, *HOW TO READ A PAPER: THE BASICS OF EVIDENCE-BASED MEDICINE* 39 (5th ed. 2014).

90. Richard Doll & A. Bradford Hill, *Lung Cancer and Other Causes of Death in Relation to Smoking; A Second Report on the Mortality of British Doctors*, 2 BRIT. MED. J. 1071 (1956).

91. FED. JUDICIAL CTR., *REFERENCE MANUAL ON SCIENTIFIC EVIDENCE* 340 (2d ed. 2000).

92. *Id.*

93. *Id.* at 342.

94. *Id.* at 374.

95. *Id.* at 387.

“relative risk.”⁹⁶ “Relative risk” is “defined as the ratio of the incidence rate (often referred to as incidence) of disease in exposed individual to the incidence rate in unexposed individuals.”⁹⁷ If the “relative risk” is 1, “the risk in exposed individuals is the same as the risk in unexposed individuals.”⁹⁸ Thus if the relative risk is 1, then “[t]here is no association between exposure to the agent and disease.”⁹⁹ If the relative risk is greater than 1, “[t]here is a positive association between exposure to the agent and the disease, which could be causal.”¹⁰⁰ From a probability perspective, “[t]he higher the relative risk, the greater the likelihood that the relationship is causal.”¹⁰¹

Under the court’s “gatekeeping” responsibility under *Daubert*,¹⁰² an expert must express an opinion to a reasonable degree of probability, a burden that may keep the jury from hearing useful information about association. If the “probability” burden were translated directly to the issue of “relative risk,” then the relative risk would have to be greater than 2.0 for any expert to express an opinion to a reasonable degree of probability. This is because if the relative risk is 2.0, “the agent is responsible for an equal number of cases of disease as all other background causes.”¹⁰³ This finding from a probability perspective “implies a 50% likelihood that an exposed individual’s disease was caused by the agent.”¹⁰⁴ Because the Rule 104(a) burden for the admissibility of evidence under the *Daubert* gatekeeping responsibility is “reasonable degree of probability,” an expert must express two forms of “causation” opinions at “probability” level.¹⁰⁵ The first causation opinion is that the agent has the capacity of causing the disease or problem at a probability level. This “general causation” theory, can be contrasted to the

96. *Id.* at 348.

97. REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, *supra* note 91, at 348.

98. *Id.* at 349.

99. *Id.*

100. *Id.*

101. *Id.* at 376.

102. *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 597 (1993) (“We recognize that, in practice, a *gatekeeping* role for the judge, no matter how flexible, inevitably on occasion will prevent the jury from learning of authentic insights and innovations.” (emphasis added)).

103. REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, *supra* note 91, at 384.

104. *Id.* at 384.

105. *King v. Burlington N. Santa Fe Ry. Co.*, 277 Neb. 203, 212–13 (2009). (“To prevail, a plaintiff must show both general and specific causation. But a court should first consider whether a party has presented admissible general causation evidence before considering the issue of admissible specific causation evidence.”).

second causation opinion, that the agent was the “actual cause,” where other possible causes of this particular disease can be reasonably ruled out.¹⁰⁶

Courts considering the “general,”¹⁰⁷ “generic,”¹⁰⁸ or “capable of causing”¹⁰⁹ theory of causal connection between the subject agent generally have not adopted a consensus “statistical” standard of probability for this relationship. Some courts have reversed jury verdicts where the supporting statistical studies establishing general causation were lacking.¹¹⁰ Some commentators have recommended¹¹¹ and some courts have required proof of a relative risk of 2.0 or greater to prove general causation.¹¹² Other courts have not required

106. *Hilao v. Estate of Marcos*, 103 F.3d 767, 788 (9th Cir. 1996) (Rymer, J., dissenting in part and concurring in part) (contrasting “generic causation—that the defendant was responsible for a tort which had the capacity to cause the harm alleged—with individual proximate cause and individual damage”).

107. *Raynor v. Merrell Pharm., Inc.*, 104 F.3d 1371, 1376 (D.C. Cir. 1997) (“[T]estimony on specific causation ha[s] legitimacy only as follow-up to [general causation]”); *In re Bextra & Celebrex Mktg. Sales Practices & Prod. Liab. Litig.*, 524 F. Supp. 2d 1166, 1172 (N.D. Cal. 2007) (“General or generic causation means ‘whether the substance at issue had the capacity to cause the harm alleged.’”) (citing *In re Hanford Nuclear Reservation Litig.*, 292 F.3d 1124, 1133 (9th Cir. 2002) (“General, or ‘generic’ causation has been defined by courts to mean whether the substance at issue had the capacity to cause the harm alleged . . .”).

108. *Sterling v. Velsicol Chem. Corp.*, 855 F.2d 1188, 1200 (6th Cir. 1988) (“This enabled the court to determine a kind of generic causation—whether the combination of the chemical contaminants and the plaintiffs’ exposure to them had the capacity to cause the harm alleged.”).

109. *Id.* at 1199–200 (“[T]he district court concluded that Velsicol’s chemicals and the duration of the plaintiffs’ exposure to them were *capable* of causing the types of injuries alleged . . .”); *Bonner v. ISP Techs., Inc.*, 259 F.3d 924, 928 (8th Cir. 2001) (“To prove causation in a toxic tort case, a plaintiff must show both that the alleged toxin is capable of causing injuries like that suffered by the plaintiff in human beings subjected to the same level of exposure as the plaintiff, and that the toxin was the cause of the plaintiff’s injury.”).

110. See, e.g., *Brock v. Merrell Dow Pharm., Inc.*, 874 F.2d 307, 312 (5th Cir. 1989), *amended*, 884 F.2d 167 (5th Cir. 1989), *cert. denied*, 494 U.S. 1046 (1999) (“The plaintiffs did not offer one statistically significant (one whose confidence interval did not include 1.0) study that concludes that Bendectin is a human teratogen. No published epidemiological study has found a statistically significant increased risk between exposure to Bendectin and birth defects.”).

111. Russel S. Carruth & Bernard D. Goldstein, *Relative Risk Greater than Two in Proof of Causation in Toxic Tort Litigation*, 41 JURIMETRICS J. 195 (2001).

112. *In re Breast Implant Litig.*, 11 F. Supp. 2d 1217, 1227 (D. Colo. 1998) (“None of these studies reports a statistically significant elevation of risk of rheumatic or connective tissue disease, either classic or atypical, over 2.0.”); *Sanderson v. Int’l Flavors & Fragrances, Inc.*, 950 F. Supp. 981, 1000 (C.D. Cal. 1996) (“Since Thrasher’s probability estimate is not founded upon epidemiological studies showing a relative risk of greater than two, or some other evidence that would lend a scientific foundation to the assertion that fragrances more likely

“a mathematically precise table” equating levels of exposure with levels of harm to establish general causation.”¹¹³

*King v. Burlington N. Santa Fe Railway Company*¹¹⁴ provides an example of the discussion of relative risk, the issue of general causation and probability analysis. In *King* the question was whether the deceased’s work-related exposure to benzene found in diesel fuel caused the multiple myeloma cancer from which he died. One of the issues presented was whether the plaintiff had to establish a relative risk of greater than two to survive a Daubert challenge on the issue of general causation. Addressing this issue, the court presented a scholarly discussion of the relationship between relative risk and proof of general causation. In declining to set a minimum threshold for relative risk, the court made several points. First, the court observed: “[w]hile important, a positive association presents only one piece of the causation puzzle.”¹¹⁵ Second, “[o]nce an association has been found between exposure to an agent and development of a disease, researchers consider whether the association reflects a true cause-effect relationship.”¹¹⁶ Third, while “[e]pidemiologists use causation to mean that an increase in the incidence of disease among the exposed subjects would not have occurred had they not been exposed to the agent . . .”¹¹⁷ determining causation differs from the objective inquiry into relative risk.” Fourth, “[a]n assessment of a causal relationship is not a scientific methodology as that term is used to describe logic (like a syllogism) and analytic methods. Instead, it involves subjective judgment.”¹¹⁸ Finally, the court noted that “[e]xperts consider several factors under different sets of criteria that can point to causation. Relative risk presents only one factor that they consider.”¹¹⁹ Based upon these principles, the court in *King* observed, “we believe that requiring a study to show a relative risk of 2.0 or greater is

than not caused plaintiff’s injuries, it does not constitute a valid scientific connection to the pertinent inquiry of causation.” (internal citations omitted)).

113. *Bednar v. Bassett Furniture Mfg. Co.*, 147 F.3d 737, 740 (8th Cir. 1998) (“The Bednars did not need to produce ‘a mathematically precise table equating levels of exposure with levels of harm’ in order to show Marian’s level of exposure to gaseous formaldehyde, but only ‘evidence from which a reasonable person could conclude that [the] defendant’s emission has probably caused’ the harm about which they complain.”).

114. *King v. Burlington N. Santa Fe Ry. Co.*, 762 N.W.2d 24 (Neb. 2009).

115. *Id.* at 39.

116. *Id.* (citing the REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, *supra* note 91, at 374).

117. *Id.*

118. *Id.*

119. *Id.*

too restrictive when the expert relies on the study to support an opinion on general causation.”¹²⁰ Instead, the court concluded:

So we decline to set a minimum threshold for relative risk, or any other statistical measurement, above the minimum requirement that the study show a relative risk greater than 1.0. We agree that “it would be far preferable for the district court to instruct the jury on statistical significance and then let the jury decide whether many studies over the 1.0 mark have any significance in combination.” In short, the significance of epidemiological studies with weak positive associations is a question of weight, not admissibility.¹²¹

On the separate issue of “specific” or “actual” cause, the court explained, “[i]f an expert’s general causation opinion is admissible to show that a suspected agent should be ruled in as a possible cause of the plaintiff’s disease, the court must next determine whether the expert performed a reliable differential etiology.”¹²² The court explained that first, “[t]o perform an adequate differential etiology, a medical expert must first compile a comprehensive list of hypotheses that might explain the set of salient clinical findings under consideration.”¹²³ Second, “the expert engages in a process of elimination, based on the evidence, to reach a conclusion regarding the most likely cause of the disease.”¹²⁴

Apart from the issue of the “relative risk” findings of a study for determining general causation, misclassification of “cases” can be a potentially non-obvious cause for bias in case-control studies.¹²⁵ To avoid this problem, studies need to clearly define *what* is a “case” and *when* that individual becomes a “case.” The importance of this potential bias emerged in the debate about the pertussis vaccine. Between 1976 and 1979 a large case-control study was performed in Britain called the National Childhood Encephalopathy Study (“NCES”).¹²⁶ The case group in this study consisted of

120. *Burlington N.*, 762 N.W.2d at 46.

121. *Id.* at 46-47 (citing *In re Joint E. & S. Dist. Asbestos Lit.*, 52 F.3d 1124, 1134 (2d Cir. 1995)).

122. *Id.* at 50.

123. *Id.*

124. *Id.*

125. Karen T. Copeland et al., *Bias Due to Misclassification in the Estimation of Relative Risk*, 105 AM. J. EPIDEMIOLOGY 488 (1977), <https://doi.org/10.1093/oxfordjournals.aje.a112408>.

126. David L. Miller & Euan McDonald Ross, *National Childhood Encephalopathy Study: An Interim Report*, 2 BRIT. MED. J. 992 (1978).

all children under three years of age admitted to British hospitals with the diagnosis of an acute serious neurological illness. The study matched the sex, age, and area of residence of the case group to two separate control groups. The study found seven post-DTP vaccination patients with severe neurologic deficits. Extrapolation of this number revealed a rate of permanent neurologic of 1:330,000.¹²⁷

A British court thoroughly examined the NCES study in 1988.¹²⁸ The judge ruled that the NCES study misclassified three of the seven post-DTP vaccination patients as having severe neurologic deficits. The judge concluded that two of the patients had a viral illness and one was a case of Reye's syndrome. After this reevaluation of the results of the study, there was no significant increase of encephalopathy in the vaccinated patients.

The NCES study and the follow-up case illustrate two points. First, a case-control study needs to be careful about the definition of what is termed a "case". The NCES provided data on 5.4 million child-years of observation, but the judge's reclassification of three patients changed the result from statistically significant to not.¹²⁹ Second, in the setting of a trial where medical literature is cited, it makes sense to intensely scrutinize and challenge the literature rather than simply taking a study at face value.

d. Fourth Level of Reliability under the Oxford Standards: The Cohort Study

Cohort studies rank higher under the Oxford standards than case-controlled studies. Cohort studies differ from case-control studies by being prospective or forward looking. In a cohort study, two or more groups of people are selected based upon the differences in their exposure to a particular agent. These groups are followed over time (prospectively) to determine whether they develop a particular disease or another outcome.

An example cohort study is a follow-up lung cancer study performed by Drs. Doll and Hill. Doctors Doll and Hill followed their case-control smoking study with what they deemed to be a more reliable cohort study. They sent questionnaires to 40,000 men and women in the British Medical Register. They separated the smokers and non-smokers into different groups and then

127. David L. Miller et al., *Pertussis Immunisation and Serious Acute Neurological Illness in Children*, 282 *BRIT. MED. J.* 1595 (1981); Simon Shorvon & Anne Berg, *Pertussis Vaccination and Epilepsy—An Erratic History, New Research and the Mismatch Between Science and Social Policy*, 49 *EPILEPSIA* 219 (2008).

128. *Loveday v. Renton* (1990) 1 *Med. L.R.* 117 (Q.B.).

129. Simon Shorvon & Anne Berg, *Pertussis Vaccination and Epilepsy—An Erratic History, New Research and the Mismatch Between Science and Social Policy*, 49 *EPILEPSIA* 219, 220 (2008).

followed up with them in different intervals to determine whether any died, and if they did die, from what cause. They found smoking to be statistically significantly associated with lung cancer-induced mortality.¹³⁰

In general, evidence-based medicine favors cohort studies over case-control studies, however, neither of these observational epidemiological studies can prove causation. They can only show correlation. To prove causation, researchers and the courts often rely on criteria termed the “Bradford Hill” factors for causation.¹³¹

The Bradford Hill factors provide another example of a generally accepted methodology for determining the causal relationship between a source and disease or outcome. Such factors may be admissible under the Federal Rules of Evidence Rule 803(17) as protocol generally followed by people within a particular occupation.¹³² Once admitted, the expert can walk down the steps and explain how the methodology provided a basis for a differential etiology, another well-established methodology that can be validated by expert testimony as generally relied upon by medical experts in the field. Courts have judicially noticed the methodologies of differential diagnosis and etiology as reliable methodologies for determining issues of diagnosis and causation.¹³³

The nine Bradford Hill factors are (1) temporal relationship, (2) strength of the association, (3) dose-response relationship, (4) replication of the findings, (5) biological plausibility, (6) consideration of alternative explanations, (7) cessation of exposure, (8) specificity of the association, and (9) consistency with other knowledge.¹³⁴ To apply this to the case of smoking and lung cancer, if the study shows a dose-response relationship of smoking and lung cancer, then that would serve as supporting evidence for causality.

130. Richard Doll & A. Bradford Hill, *The Mortality of Doctors in Relation to Their Smoking Habits*, 1 BRIT. MED. J. 1451, 1454, 1455 (1954).

131. King v. Burlington N. Santa Fe Ry. Co., 762 N.W.2d 24 (Neb. 2009) (citing FED. JUDICIAL CTR., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 340 (2d ed. 2000) and MICHAEL D. GREEN ET AL., *Reference Guide on Epidemiology*, in FED. JUDICIAL CTR. ET AL., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 335, 335-36 (3d ed. 2000)); Michael Höfler, *The Bradford Hill Considerations on Causality: A Counterfactual Perspective*, 2 EMERGING THEMES EPIDEMIOLOGY 11 (2005); A. Bradford Hill, *The Environment and Disease: Association or Causation?*, 58 PROCS. ROYAL SOC'Y MED. 295 (1965).

132. FED. R. EVID. 803(17) (“Market quotations, lists, directories, or other compilations that are generally relied on by the public or by persons in particular occupations.”).

133. *In re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717, 758 (3d Cir. 1994); *Coastal Tankships, U.S.A., Inc. v. Anderson*, 87 S.W.3d 591 (Tex. App. 2002); *Carlson v. Okerstrom*, 675 N.W.2d 89 (Neb. 2004); *Schafersman v. Agland Coop.*, 631 N.W.2d 862 (Neb. 2001).

134. *Burlington N.*, 762 N.W.2d at 40 (citing MICHAEL D. GREEN ET AL., *Reference Guide on Epidemiology*, in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, *supra* note 131, at 375-76.).

The more Bradford Hill factors that apply to the situation, the greater the probability of causality.¹³⁵

The issue of bias in a cohort study is always important to review. The two groups selected in the cohort study need to be as close as possible to being identical, with the sole exception being their exposure to the agent being studied. The greater the differences between the groups, the greater the risk of bias.¹³⁶

e. Fifth Level: Cross-Sectional Studies

Diagnostic studies often utilize the cross-sectional method. The diagnostic test being studied (the “index test”) and a reference test are administered to a given patient population. The results of the tests are compared to determine the accuracy of the index test. The studied patient population can be selected via either the case-control method or a consecutive series method. In the case-control selection, the researcher selects case patients with a known disease and compares them to control patients who are known not to have the disease. This artificial selection of a patient pool is performed because the methodology is cheap and fast. For this reason, it is often used in the initial evaluation of a diagnostic test or in the evaluation of a rare disease. However, just because a test has good accuracy in the artificial environment of a case-control study does not necessarily mean that the test will be accurate in the clinical setting. To best test for the accuracy of a diagnostic test in the clinical setting, a consecutive series method is used. In this method, patients suspected of having the disease, but in whom disease status is unknown, are given both the index test and the reference test. Ideally, these patients are consecutively chosen from the appropriate clinical setting.¹³⁷

The patient group composition can dramatically affect the study results. A study which only includes very sick patients and perfectly healthy patients will make a test look better than it is. This is called spectrum bias.¹³⁸ The prevalence of disease, severity of disease, and presence of comorbid conditions can all have a big effect. For example, a mammogram on a screening population will have different sensitivity and specificity for disease than a mammogram on a patient population with a palpable lump. The physician or expert using the study needs to ascertain whether the patient population in the study is similar to the clinical patient or patient in the legal trial.

135. Höfler, *supra* note 131, at 11.

136. Copeland, *supra* note 124.

137. *Id.*

138. J. Andre Knottnerus & Jean W. Muris, *Assessment of the Accuracy of Diagnostic Tests: The Cross-Sectional Study*, 56 J. CLINICAL EPIDEMIOLOGY 1118 (2003).

The clinical relevance of a diagnostic test's accuracy must always be assessed. If the test result is highly accurate but the information does not change patient outcome, then the increased accuracy may be irrelevant.

The reference test can be a source of bias. In "incorporation bias," the index test result is incorporated into the reference test. This can artificially increase the accuracy of the index test. To avoid this, interpreters of the reference test should be blinded to the result of the index test and vice versa.¹³⁹

f. Sixth Level: Randomized Controlled Trials

In a randomized controlled trial ("RCT"), patients are randomly assigned to the treatment group or the control group. Both groups are followed for a pre-specified time period and analyzed in terms of a specific outcome at the outset of the study. The randomization of patients reduces the risk of selection bias.

Case-control and cohort studies are observational studies in the sense that the researcher need not directly change the behavior of the participants. On the other hand, a RCT does precisely determine the treatment plan/behavior of its participants. This is advantageous in reducing bias, but it can also make a RCT ethically problematic such as when assessing the toxicity of an agent, such as smoking.

An RCT can also be prone to a number of biases. Selection bias can still occur in a RCT if there is imperfect randomization or if there is a failure to randomize all eligible patients. Performance bias is a systematic difference in the care provided to the randomized groups, apart from the intervention being studied. Exclusion bias is systematic differences in withdrawals from the trial. Detection bias occurs in systematic differences of outcome assessment, this is particularly common when there is a failure to blind the assessors to the randomization status of patients.¹⁴⁰ Another bias occurs when those who fail to complete a clinical trial are ignored. In general, this bias tends to be in favor of the intervention. For this reason, studies with a low follow-up rate are generally considered untrustworthy and studies are generally analyzed on an "intent-to-treat" basis.¹⁴¹ Although not a source of bias, a limitation of a RCT can be the clinical relevance of the measured outcome. The outcome being measured may be quantifiable, and thus easy to study, but that does not mean that it has practical clinical value. Despite

139. *Id.*

140. TRISHA GREENHALGH, HOW TO READ A PAPER: THE BASICS OF EVIDENCE-BASED MEDICINE 50 (5th ed. 2014).

141. *Id.* at 56.

these listed limitations, an RCT is favored over observational studies because a properly performed RCT can eliminate many of the sources of potential bias in research.

g. Highest Tier Evidence: Secondary Studies, Systematic Reviews, and Meta-Analyses and Practice Guidelines

Primary research studies, such as observational studies and randomized controlled trials, collect data on individual patients. Secondary studies, such as meta-analyses and systematic reviews, aggregate information from multiple primary studies. A systematic review begins by asking a well-defined clinical question. Then a systematic search of the literature is performed, using a well-defined and reproducible search strategy, to find all primary research articles that ask the specified clinical question. The researcher then closely scrutinizes the methods of the aggregated studies. If the methods fit the strict and explicit criteria, then that study's results need to be included in the analysis, regardless of the results of that study. A meta-analysis begins with a systematic review but then goes further by applying statistics to quantify the results of the systematic review.

There are several advantages of a systematic review and a meta-analysis over a primary study. First, by grouping smaller primary studies, a meta-analysis may change a trend found in the primary studies to a statistically significant finding. Second, meta-analyses can help resolve contradictory findings among the different primary studies. Finally, the results of a meta-analysis are more robust, more generalizable, and more likely to be true.¹⁴²

However, meta-analyses/systematic reviews are not without weaknesses. They can replicate and magnify flaws in the original primary studies. Additionally, meta-analysis can yield false results if the underlying studies are done on different patient populations or use different methodologies, which frequently occurs.¹⁴³

Medical societies create committees of respected physicians to review the scientific literature and create evidence-based medicine practice guidelines.¹⁴⁴ These guidelines are published to assist practitioners in making daily clinical decisions. Committee members who design the guidelines often

142. John P.A. Ioannidis, *Why Most Published Research Findings Are False*, 2 PLOS MED. e124 (2005).

143. *King v. Burlington N. Santa Fe Ry. Co.*, 762 N.W.2d 24, 48 (Neb. 2009); Charles H. Hennekens & David Demets, *The Need for Large-Scale Randomized Evidence Without Undue Emphasis on Small Trials, Meta-analyses, or Subgroup Analyses*, 302 J. AM. MED. ASS'N 2361 (2009).

144. See *Practice Management, Quality and Informatics*, AM. COLL. OF RADIOLOGY, <http://www.acr.org/Quality-Safety> (last visited Nov. 28, 2018).

debate at length about the wording of the guidelines. As a result, the final publication often represents a compromise between differing physician opinions.

Physicians and experts generally view practice guideline recommendations as a high level of evidence, comparable to systematic reviews and meta-analyses.¹⁴⁵ For example, in *Bayer ex rel. Petrucelli v. Dobbins*,¹⁴⁶ the Wisconsin Court of Appeals, in considering a malpractice claim denying a causal theory of “maternal forces,” reversed a trial court’s exclusion of a compendium of studies released in 2014 by the American College of Obstetricians and Gynecologists (“ACOG”) on “Neonatal Brachial Plexus Palsy” even though the purpose of the compendium was

[t]o review and summarize the current state of the scientific knowledge, as set forth in the peer-reviewed and relevant historical literature, about the mechanisms which may result in neonatal brachial plexus palsy. The purpose of conducting such review is to produce a report which will succinctly summarize the relevant research on the pathophysiology of neonatal brachial plexus palsy.¹⁴⁷

But practice guidelines and systematic reviews are not infallible. Unfortunately, studies have shown that some practice guidelines have serious shortcomings.¹⁴⁸ And at times practice guidelines held by differing medical societies can have widely disparate recommendations despite the fact that the differing recommendations from the societies are all based on the same primary studies.¹⁴⁹

h. Limitations of the Oxford Levels of Evidence

The Oxford Center does not intend for its ranking system to be the final say in the assessment of the quality of a study. Exceptions exist where research studies exceed or fall behind the expectations of the model of the Levels of Evidence. Consequently, the use of this heuristic requires a “healthy dose of skepticism and judgement . . . to appraise evidence and apply it to

145. Brief for Wisconsin Medical Society as Amicus Curiae Supporting Appellant, *Bayer ex rel. Petrucelli v. Dobbins*, 885 N.W.2d 173 (Wis. Ct. App. 2016) (No. 2015AP1470), 2015 WL 9596321; Lisa P. Lavelle et al., *Evidence-Based Practice of Radiology*, 35 RADIOGRAPHICS 1802 (2015).

146. *Bayer ex rel. Petrucelli v. Dobbins*, 885 N.W.2d 173 (Wis. Ct. App. 2016).

147. *Id.* at 176.

148. Roberto Grilli et al., *Practice Guidelines Developed by Specialty Societies: The Need for a Critical Appraisal*, 355 LANCET 103 (2000); Jonathan L. Mezrich & Charles S. Resnik, *Panacea or Sham? Legal Issues of Vertebroplasty*, 13 J. AM. COLL. RADIOLOGY 663 (2016).

149. *Id.*

individuals in routine practice.”¹⁵⁰ This skepticism requires the clinician to go beyond the Levels of Evidence analysis to delve into the details of the methods and results of each research study.

III. APPLICATION

Apart from the ranking of levels of reliability for the various studies, there remains the “fit” of the study to the “facts of the case.” This “fitness” issue is often considered in the context of the “application” level of the analysis. The courts often ask whether the expert’s opinion is an “unfair extrapolation” from the studies being relied upon.

On the issue of factual application, the United States Supreme Court in *General Electric Co. v. Joiner*¹⁵¹ explained:

[N]othing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the *ipse dixit* of the expert. A court may conclude that there is simply too great an analytical gap between the data and the opinion proffered.¹⁵²

However, the *Daubert* analysis “is not intended to supplant the adversarial process.”¹⁵³ Even “shaky” expert testimony may be admitted if the *Daubert* standard is satisfied, subject to cross-examination.¹⁵⁴ More specifically, “[v]igorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence.”¹⁵⁵

150. Jeremy Howick et al., *Explanation of the 2011 Oxford Centre for Evidence-Based Medicine (OCEBM) Levels of Evidence (Background Document)*, OXFORD CTR. FOR EVIDENCE-BASED MED. (May 1, 2016), <https://www.cebm.net/index.aspx?o=5653>.

151. *Gen. Elec. Co. v. Joiner*, 522 U.S. 136 (1997).

152. *Id.* at 146.

We further hold that, because it was within the District Court’s discretion to conclude that the studies upon which the experts relied were not sufficient, whether individually or in combination, to support their conclusions that Joiner’s exposure to PCBs contributed to his cancer, the District Court did not abuse its discretion in excluding their testimony.

Id. at 146-47.

153. *Bielskis v. Louisville Ladder, Inc.*, 663 F.3d 887, 894 (7th Cir. 2011).

154. *Gayton v. McCoy*, 593 F.3d 610, 616 (2010) (“Determination on admissibility should not supplant the adversarial process; ‘shaky’ expert testimony may be admissible, assailable by its opponents through cross-examination.”).

155. *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 596 (1993).

In *Bayer ex rel. Petrucelli v. Dobbins*,¹⁵⁶ the Wisconsin Court of Appeals adopted *Daubert* as the standard of admissibility for Wisconsin state courts.¹⁵⁷ *Bayer*, a medical malpractice case, provides a useful example. In *Bayer*, a mother sued her obstetrician for medical malpractice claiming that his negligence during delivery resulted in her baby having a permanent brachial plexus injury.¹⁵⁸ In defense, the obstetrician cited numerous scientific articles supporting the expert's opinion that it is more likely that maternal forces caused the injury, not the physician's actions.¹⁵⁹ The trial court excluded the defense's use of the "maternal forces" literature claiming, "the problem that I see with everything that is being done on this from the defense standpoint is that these articles are not distinguishing between permanent brachial plexus injuries and temporary brachial plexus injuries."¹⁶⁰ In essence, the court excluded the compendium of research studies because the articles' measured outcome, of both permanent and temporary brachial plexus injuries, differed from the patient's measured outcome of only a permanent brachial plexus injury.¹⁶¹ The Wisconsin Court of Appeals reversed and remanded because "neither the Bayers nor the circuit court have explained why the defense experts should not be permitted to extrapolate from the multiple peer-reviewed articles that, while supporting the maternal forces theory for causation, fail to distinguish between temporary and permanent brachial plexus injuries."¹⁶²

Bayer suggests that if a study follows proper protocol then the courts should permit an expert to extrapolate from the scientific literature to the facts of the case. The level of "unfair extrapolation" will always be a *Daubert* "application" question, but courts are less likely to exclude such evidence if the underlying principles and methodologies of the studies are reliable and the experts can explain the relevancy of any differences in "fitness" of the facts.

156. *Bayer ex rel. Petrucelli v. Dobbins*, 885 N.W.2d 173 (Wis. Ct. App. 2016).

157. *Id.* at 180.

158. *Id.* at 175.

159. *Id.* at 180 ("Dobbins cited over twenty peer-reviewed publications supporting his claim that maternal forces of labor caused Unity's injury.").

160. *Id.* at 178.

161. *See id.* at 182-84.

162. *Bayer*, 885 N.W.2d at 182.

A. *Evidence-Based Subjects for Cross-Examination of the Expert Beyond the Pyramid Ranking of the Studies*

1. Bias and Statistical Error in the Medical Literature

Unfortunately, it is well established that bias and statistical error result in a high percentage of false results in medical literature.¹⁶³ This question of potential research errors should be queried whenever any peer-reviewed study is relied upon by an expert when giving expert testimony. A few corollaries exist to predict the truthfulness of an article.¹⁶⁴ In general, the greater the financial interests, other interests, and prejudices in a scientific field, the less likely the research findings are to be true. For example, studies performed by drug manufacturers on their own drugs are less likely to be true due to financial bias. But bias need not just be financial. Some scientists bias their results to obtain public acclaim or to acquire tenure. In “scientific bias,” researchers alter the results to reinforce the prevailing scientific theory. These biases need not be either conscious or intentional to be a source of unreliable evidence.¹⁶⁵

Even assuming the absence of any bias, research studies can be false simply because of statistical mishaps. These statistical mishaps occur more frequently when the studies are smaller, when the studied effect size is smaller, or when a study reveals an unexpected result.¹⁶⁶ Because of the susceptibility to error, the scientific literature needs to be critiqued before being used by a physician or before it is relied upon in expert testimony.

2. Critique Based upon a Study Using Journal Quality and Study-Type: Ranking Scientific Journals

Not all peer-reviewed research is equally reliable. Consequently, physicians and expert witnesses need to have techniques to sift out weak studies. One way to start this literature critique is to start with the medical journal quality. Journal quality can be measured using impact factors. An impact factor is a number that reflects the average number of times a particular journal has been cited by other journals. A higher impact factor means more citations and, presumably means, a greater likelihood of

163. John P.A. Ioannidis, *Why Most Published Research Findings are False*, 2 PLOS MED. e124 (2005); M. Carrington Reid et al., *Use of Methodological Standards in Diagnostic Test Research. Getting Better but Still Not Good.*, 274 J. AM. MED. ASS'N 645 (1995); Anne W.S. Rutjes et al., *Evidence of Bias and Variation in Diagnostic Accuracy Studies*, 174 CANADIAN MED. ASS'N J. 469 (2006).

164. Ioannidis, *supra* note 163, at e124.

165. *Id.*

166. *Id.* (The lower the pretest odds, the less likely it is to be true).

scientific respect. One limitation of the impact factors is that a citation in a low-ranking journal is given equal weight to a citation from a high-ranking journal. The Eigenfactor value overcomes this limitation by giving weight to citations that come from highly rated journals.¹⁶⁷ A journal may be ranked under the Eigenfactor by researching EIGENFACTOR.org. The online SCImago journal rank¹⁶⁸ is an example of rating journals by Eigenfactor. If a witness relies upon a study, counsel for either the proponent or opponent should review the Eigenfactor to credit or discredit any reliance given to the study in the expert's opinion. While the quality of a scientific journal is relevant in the assessment of a given research article, this ranking is not determinative. There are countless examples of incorrect research results in highly esteemed journals, and there are many great articles published in lesser known journals.

3. Critiquing the Details of a Study's Methodology

a. Flow diagrams

After critiquing a study based on general features, such as its ranking within Oxford's Level of Evidence, potential bias, and the ranking of the journal of origin, the expert needs to review and potentially critique the detailed methods of the study. Reading the methods section of papers can be confusing, even for experts. To simplify this process, readers are encouraged to create flow diagrams. These diagrams create a visual aid that breaks down the methods of a study into many different steps.¹⁶⁹ These diagrams can be a useful aid for physicians, expert witnesses, judges, and juries as they all seek to understand the detailed methods of a research study.

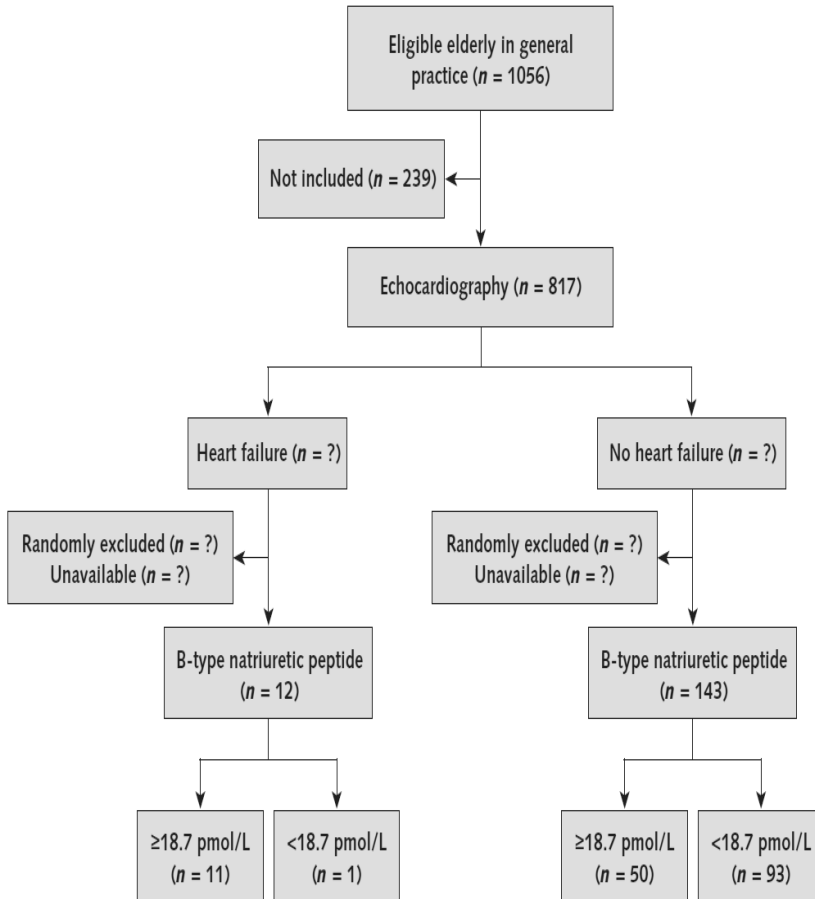
These flow diagrams are best explained by example. Here is an example of a flow diagram, created by the QUADAS group, of a cohort study that sought to determine whether B-type natriuretic peptide levels could be utilized to diagnose heart failure:¹⁷⁰

167. Carl T. Bergstrom et al., *The Eigenfactor™ Metrics*, 28 J. NEUROSCIENCE 11433 (2008).

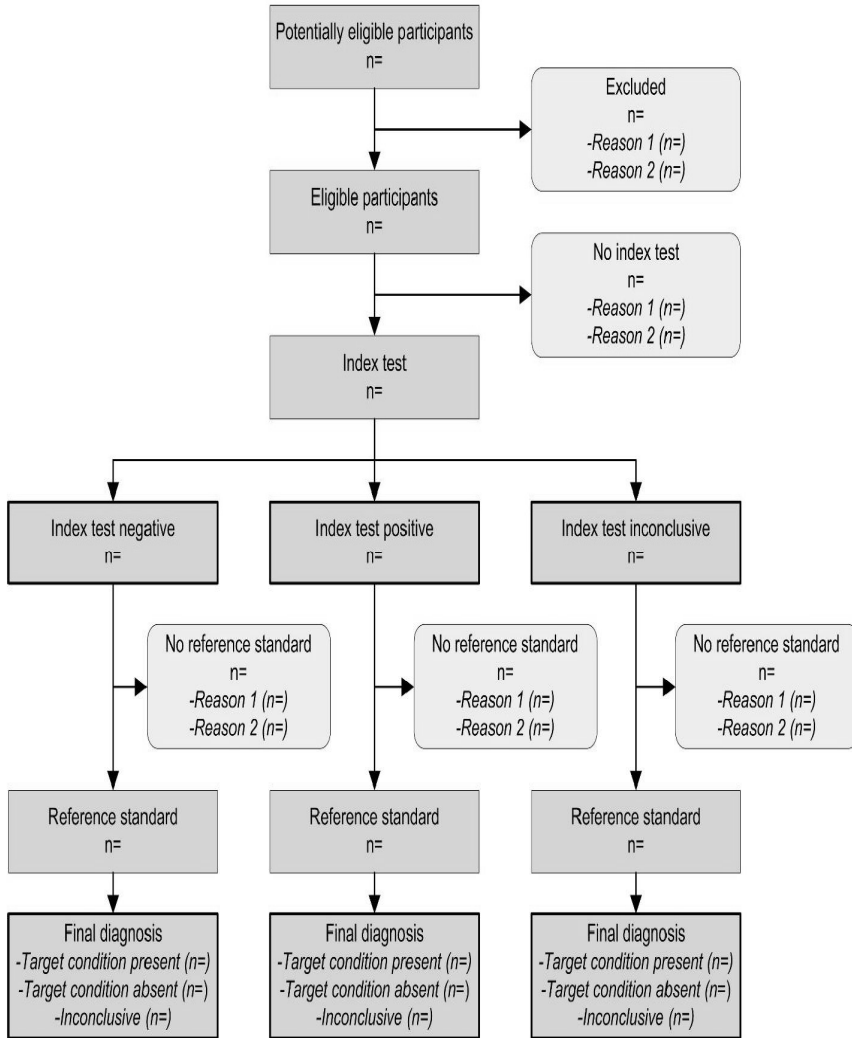
168. *SCImago Journal Rank*, SCIMAGO, <http://www.scimagojr.com> (last visited Nov. 28, 2018).

169. Penny F. Whiting et al., *QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies*, 155 ANNALS INTERNAL MED. 529 (2011).

170. *Id.*; Helen E. Smith et al., *Biochemical Diagnosis of Ventricular Dysfunction in Elderly Patients in General Practice: Observational Study*, 320 BRIT. MED. J. 906 (2000).

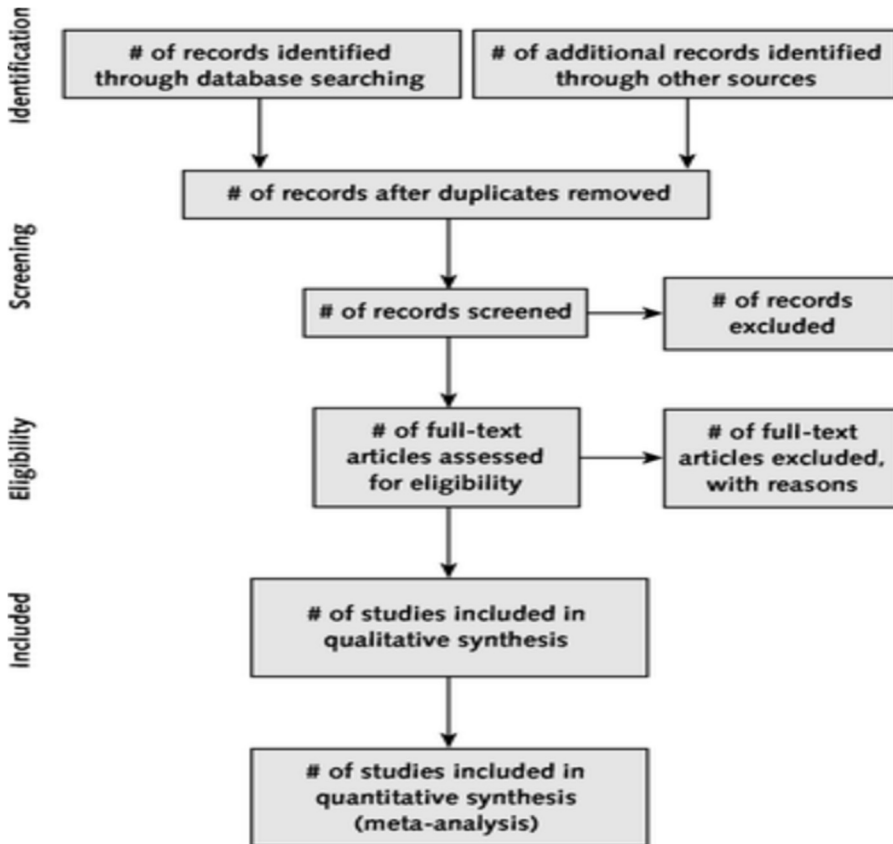


Here is an example flow diagram created by the STARD group for diagnostic studies:¹⁷¹



171. Patrick M. Bossuyt et al., *The STARD Statement for Reporting Studies of Diagnostic Accuracy: Explanation and Elaboration*, 138 ANNALS INTERNAL MED. W1 (2003). Patrick M. Bossuyt et al. for the STARD Group, *STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies*, BRITISH MED. J. (Oct. 28, 2015), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4623764/>.

Flow diagrams can also be created for systematic reviews and meta-analyses:¹⁷²



b. Challenging the methods (CASP)

After creating a flow-diagram of a study, the physician or witness is now ready to critically challenge the details of the study's methodology. Standardized techniques have been developed to analyze the methods of a study.¹⁷³ One of these methods is the Critical Appraisal Skills Program

172. David Moher et al., *Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement*, 151 ANNALS INTERNAL MED. 264 (2009).

173. Lisa P. Lavelle et al., *Evidence-Based Practice of Radiology*, 35 RADIOGRAPHICS 1802 (2015); Penny F. Whiting et al., *QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies*, 155 ANNALS INTERNAL MED. 529 (2011).

(“CASP”).¹⁷⁴ CASP directs the reader to ask pointed questions of the study in assessing for bias and applicability. CASP has different questionnaires for each of the types of studies discussed previously. CASP also has a few questionnaires for study types that we have not discussed, such as qualitative studies. A physician, expert witness, or lawyer can use the CASP questionnaires to better form their opinion about the biases of the study and the potential applicability of the study in their specific clinical situation.

- c. Understanding and critiquing research statistics: comparing two groups to determine any statistical difference

- (1) Null hypothesis

A “null hypothesis” states that two measured populations have no statistically significant difference.¹⁷⁵ To use smoking and lung cancer as an example, a null hypothesis would state that the incidence of lung cancer is the same in those who smoke compared to those who do not smoke. If the lung cancer study reveals a statistically significant difference between the groups, then the null hypothesis would be rejected and the conclusion would be that there is a correlation between smoking and lung cancer.

- (2) P-value

A p-value quantifies the likelihood that there is a statistically significant difference between the two groups compared in the null hypothesis. If a p-value is .01 then there is a 99% probability that the null hypothesis is false. Conversely, if a p-value has a 5% chance that the null hypothesis is false, then there is a 95% probability that chance alone could account for the measured differences between the groups. Scientists in general prefer to have a p-value of less than .05 before saying that a study reveals a statistically significant likelihood of the null hypothesis being false. If the p-value is greater than .05 then the null hypothesis cannot be rejected.

- (3) Type 1 error

A “type 1” error occurs when a study rejects a true null hypothesis.¹⁷⁶ Even if all bias has been removed, a type 1 error can occur strictly out of chance alone. For example, if a study rejects a null hypothesis based on a p-value of

174. *CASP Checklists*, CRITICAL APPRAISAL SKILLS PROGRAMME, <https://casp-uk.net/casp-tools-checklists> (last visited Nov. 28, 2018).

175. Michael D. Green et al., *Reference Guide on Epidemiology*, in FED. JUDICIAL CTR. ET AL., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 122 (2d ed. 2000); BRIAN EVERITT, THE CAMBRIDGE DICTIONARY OF STATISTICS (1998).

176. ROBERT H. RIFFENBURGH, STATISTICS IN MEDICINE (3d ed. 2012).

.05, then 5% of the time an error was made and the null hypothesis was actually true. If the study has bias, and most studies do have some bias, then the chance of a type 1 error can be greater.

(4) Type 2 error

A “type 2” error occurs when a study finds no difference between the measured groups even though a real difference exists.¹⁷⁷ Studies that do not enroll enough patients are at risk for type 2 errors because the results may not be strong enough to show a statistically significant result. Such studies are said to be “underpowered.”¹⁷⁸ Increasing the number of enrolled patients decreases the risk of a type 2 error but also increases the cost of the study.

(5) Power calculation

A trial should be big enough to have a high chance of detecting, as statistically significant, an effect if it exists. In other words, a trial should be “powered” to reduce the possibility of a “type 2” error. This power calculation can be performed before the study begins to measure the appropriate size of the study. The power calculation depends on the expected level of difference between the two groups that would constitute a clinically significant effect and the mean/standard deviation of the principal outcome variable.

d. Using statistics to show association between an agent and outcome

(1) Absolute risks and relative risks

The “absolute risk” is the probability of an event occurring in a given patient population.¹⁷⁹ The relative risk is calculated as the quotient of the “absolute risk of the test group” divided by the “absolute risk of the control group.” A relative risk of 1.00 means that the measured outcome is equally prevalent in the test group as in the control group. A relative risk of 2.00 means that the measured outcome is twice as common in the test group as in the control group.

Some courts use a relative risk of 2.00 as evidence that a particular causative agent is more likely than not to have been the cause of the negative

177. Patrick M. Bossuyt et al., *The STARD Statement for Reporting Studies of Diagnostic Accuracy: Explanation and Elaboration*, 138 ANNALS INTERNAL MED. W1 (2003). Patrick M. Bossuyt et al. for the STARD Group, *STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies*, BRIT. MED. J. (Oct. 28, 2015), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4623764/>.

178. *Id.*

179. RIFFENBURGH, *supra* note 176.

outcome.¹⁸⁰ A relative risk of less than 2.00 can still be attributed to the causative agent, but it is not more likely than not. However, using the 2.00 relative risk as a determinant factor does confuse the notion of general causality and specific causality. A study showing a toxin causes cancer with a relative risk of 1.8 reveals the general causality of that toxin. However, a given patient may have a greater or lesser risk depending on that patient's other risk factors. An expert witness will be needed to extrapolate the specific relative risk to the patient from the research study general causality relative risk.

A few other terms are often used:

Relative Risk Reduction = $1 - \text{Relative Risk}$

Absolute Risk Reduction = Absolute Risk of Therapy – Absolute Risk of Control

Number Needed to Treat = $1 / \text{Absolute Risk Reduction}$

(2) Odds ratio

An odds ratio (“OR”) is another measure of association between an exposure and an outcome. The OR represents the odds that an outcome will occur given a particular exposure compared to the odds of the outcome occurring in the absence of that exposure. The odds ratio can be calculated from a 2 x 2 frequency table.

		Outcome Status	
		+	-
Exposure Status	+	A	B
	-	C	D

$$OR = (A/C) / (B/D) = (A*D) / (B*C)$$

If the OR equals one, then there is no association between the exposure and the outcome. If the OR is greater than one, then the exposure is positively associated, or correlates, with the outcome. If the OR is less than one, then the exposure is negatively associated with the outcome.

Case-control studies use odds ratios instead of relative risk because relative risk cannot be calculated in these studies. In case-control studies, the prevalence of disease is unknown so absolute risks cannot be calculated.

(3) Confidence intervals

Confidence intervals describe a range of values in which the true value lies within a certain degree of probability. In general, confidence intervals use a

180. King v. Burlington N. Santa Fe Ry. Co., 762 N.W.2d 24, 36 (Neb. 2009).

95% probability. This means that the chance that the true value lies within the range specified by the confidence intervals is 95%. Odds ratios and relative risks are often stated with confidence intervals. If the stated confidence interval of an odds-ratio or a relative risk overlap with 1.00, then there is no statistically significant correlation between the causative agent and the outcome.

e. Using statistics to appraise diagnostic tests

Many different statistical terms are used to quantify the quality of a diagnostic test. These terms are derived from a standard 2 x 2 table showing the result of the index test on the left and the result of the reference test on the top, as shown in the table below.¹⁸¹

		Result of Reference Test	
		Disease Positive A+C	Disease Negative B+D
Result of Index Test	Test Positive A+B	True Positive A	False Positive B
	Test Negative C+D	False Negative C	True Negative D

From the above table, multiple important definitions emerge:

(1) True Positive, false positive, true negative, and false negative

True positive, false positive, true negative, and false negative are descriptive terms for a given test result.¹⁸² A test result is said to be a “true positive” if the index test was positive and the disease was present. A test result is a “false negative” if the index test result is negative but the disease is present.

(2) Sensitivity and specificity

Sensitivity and specificity are numeric values describing the quality of a test. Sensitivity is calculated as the quotient of true positive cases divided by all patients with the disease: $A/(A+C)$. A highly sensitive test is a good screening test because it has a low false negative rate.

181. RIFFENBURGH, *supra* note 176.

182. *Id.*

Specificity is calculated as the quotient of the true negative cases divided by all patients without disease. A highly specific test is good at ruling out disease because there are few false positive results.

Here are some related terms:

Positive predictive value: $A/(A+B)$

Negative predictive value: $D/(C+D)$

Accuracy: $(A+D)/(A+B+C+D)$

(3) Likelihood ratios

Likelihood ratios can be used on diagnostic tests to determine the probability of disease. According to Bayes' theorem, the odds that a patient has a disease equals the pretest odds of disease multiplied by the test's likelihood ratio.¹⁸³ Each test has two likelihood ratio values, one to be used in the setting of a positive test result and the other to be used with a negative test result. These likelihood ratio values can be calculated from a test's sensitivity and specificity.¹⁸⁴

Likelihood ratio of a positive test: $sensitivity / (1-specificity)$

Likelihood ratio of a negative test: $(1-sensitivity) / specificity$

Let us say for example that the positive likelihood ratio of a test is 8.0 and the pretest odds for disease is 1:3 (25% probability). If we perform the test and the test is positive, then the post-test odds of disease is 8:3 (73% probability). This calculation helps reveal the importance of knowing the pretest odds of disease in interpreting the results of any diagnostic test.¹⁸⁵

IV. CONCLUSION

In the early 1990's both the courts and the medical profession began to be concerned about investigating the reliability of expert opinions relied upon in their different professions. The courts have addressed the problem by adopting the *Daubert* standard of admissibility, but the courts have given little guidance evaluating different levels of reliability of health-related testimony under the broad *Daubert* standards. At the same time, the medical profession has adopted the principles and methodology of "evidence-based medicine" for teaching and practice guidelines for physicians and health-related professions.¹⁸⁶ This paper recommends that legal practitioners

183. *Id.*

184. *Id.*

185. Lisa P. Lavelle et al., *Evidence-Based Practice of Radiology*, 35 RADIOGRAPHICS 1802 (2015).

186. Gordon Guyatt et al., *Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine*, 268 J. AM. MED. ASS'N 2420 (1992); see also Benjamin Djulbegovic &

become more familiar with the techniques employed by evidence-based medicine in the direct and cross-examination of experts in both health-related and analogous fields where both expert testimony and learned treatises are relied upon in support of expert testimony.