EVIDENCE SYNTHESIS AND GENERAL CAUSATION: KEY METHODS AND AN ASSESSMENT OF RELIABILITY

*Douglas L. Weed**

TABLE OF CONTENTS

I. INTRODUCTION

A dozen years ago, the Supreme Court changed the landscape of the intersection of science and the law. It ruled in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*[1] that judges were to act as evidentiary gatekeepers in the courtroom, applying new standards for the reliability of scientific testimony.[2] Prior to *Daubert*, most courts relied upon the less stringent general acceptance standard announced in *Frye v. United States*.[3] After *Daubert*, judges were to examine the reliability of scientific expert testimony on the basis of several factors: (1) the testability of the scientific hypothesis or technique; (2) whether the theory or technique has been

* Director, Cancer Prevention Fellowship Program, National Cancer Institute; B.S., 1974, M.D., 1977, Ohio State University; M.P.H., 1980, Ph.D., 1982, University of North Carolina.

1. Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579 (1993).

2. *See id.* at 592–93 (holding that the Federal Rules of Evidence superseded the previously used general acceptance standard).

3. Frye v. United States, 293 F. 1013 (1923); *see Daubert*, 509 U.S. at 585.

published and peer reviewed; (3) the known or potential rate of error in the theory or technique; (4) the existence and maintenance of standards and controls; and (5) the extent to which the theory or technique has been generally accepted within the scientific community.[4] *Daubert* also changed the role of judges. After *Daubert*, judges were required to be much more proactive in screening and in not admitting scientific evidence that was not, in their assessment, both reliable and relevant, guarding the proceedings against what some had perceived to be an increase in the use of so-called "junk science" in the courtroom.[5]

Judges were now expected to think like scientists, although it has been pointed out that the Court expected judges to "think like scientists as *they* imagined scientists think."[6] This is no small feat: vast differences persist between the cultures of science and law, reflecting different educational backgrounds as well as fundamental differences in the respective institutions.[7] These differences are particularly apparent in the assessment of scientific evidence for general causation in toxic tort litigation.[8] There is a consensus in both the scientific and legal

4. *Daubert*, 509 U.S. at 593–94.

5. Sheila Jasanoff, *Law's Knowledge: Science for Justice in Legal Settings*, 95 AM. J. PUB. HEALTH 49, 57 n.8 (Supp. 1 2005).

6. *See id.*

7. A committee convened by the National Academy of Sciences to discuss the impact of *Daubert* and subsequent Supreme Court decisions and to identify questions for future study concluded that science and the courts "operate with a different set of values and rules." The National Academies of Science, Technology, and Law Panel, *Meeting Summary: Committee on Daubert Standards* (forthcoming 2006) (on file with author).

8. Differences in the use and definitions of key terms represent opportunities for confusion and misunderstanding between the cultures of law and science. The Supreme Court, for example, used the words "theory" and "technique" to represent concepts that most biomedical scientists, including epidemiologists, would likely refer to in their practices as "hypothesis" and "method," respectively. *See* Kumho Tire Co. v. Carmichael, 526 U.S. 137, 149–50 (1999) (citing *Daubert*, 509 U.S. at 592–94). Are these important differences in language, or is it reasonable to substitute one term for the other with minimal confusion?

In the practice of epidemiology and other biomedical sciences, theory and hypothesis have the following relationship: scientific theories, which provide general explanations for observable phenomena, generate hypotheses to be tested in empirical studies. A general theory of carcinogenesis, for example, could generate a variety of hypotheses relating to specific chemicals and specific cancers, which are then examined one by one in population-based (epidemiologic) studies as well as laboratory-based (toxicologic) studies. The hypothesis, "Does chemical *X* cause cancer *Y*?"—the prototypical problem associated with general causation—would not typically be called a theory by biomedical scientists. *See generally* Douglas L. Weed, *Theory and Practice* communities that scientists in their practices and judges functioning as evidentiary gatekeepers evaluate the scientific evidence of disease causation differently.[9]

The problem goes even deeper. Within science, the evaluation and interpretation of evidence in any particular situation can result in considerable variability. When scientists from different disciplines interpret the same evidence, they may come to different, even opposing, conclusions. Precisely that phenomenon surprised and confounded the court in *Soldo v. Sandoz Pharmaceuticals Corp*.[10] Three scientists, working independently, were appointed by the court to offer their conclusions in a case involving a medication, Parlodel, and its potential to cause strokes.[11] Two of the three experts, a neurologist and an epidemiologist, determined that a claim of causation based on the limited amount of evidence provided was not warranted; however, the third expert, a pharmacologist, disagreed.[12] In this particular case, the court agreed with the "majority"

---

in Epidemiology, 954 ANNALS N.Y. ACAD. SCI. 52, 53–54 (2001). Using "theory" for "hypothesis," however, is unlikely to cause much confusion.

The distinction between technique and method, however, may help prevent a potential source of confusion regarding one of the key methods for inferring causation from scientific evidence: the so-called causal criteria to be discussed in more detail later in this Article. In a description of these criteria intended for the legal community, Green, Freedman, and Gordis state that: "There is no formula or algorithm that can be used to assess whether a causal inference is appropriate . . . . While the drawing of causal inferences is informed by scientific expertise, it is not a determination that is made by using scientific methodology." Michael D. Green, D. Michal Freedman & Leon Gordis, *Reference Guide on Epidemiology*, *in* FED. JUDICIAL CTR., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 333, 375 (2d ed. 2000). Taken literally, these authors do not consider the causal criteria to be a scientific methodology, therefore revealing a common bias within the scientific community against the use of more qualitative methods, but more importantly, potentially putting the method (or technique) at risk of being considered unscientific. *See id.* (noting it is not a determination that is made by using scientific methodology). The use of the more expansive concept of technique would, in this instance, alleviate any confusion. More importantly, it is reasonable to assert that this methodology (technique) is generally accepted in the scientific community, inasmuch as it is widely cited in epidemiological textbooks and is widely recognized in the epidemiological community. *See id.* at 374–75 (footnotes omitted) (discussing epidemiology causation technique). However, the method is not free of controversy, a topic discussed later in this Article.

9. There is also a distinction to be made between the way scientists evaluate evidence in their practices and how they represent evidence in the courtroom as experts. In this Article, I will confine my remarks to the former.

10. Soldo v. Sandoz Pharms. Corp., 244 F. Supp. 2d 434 (W.D. Pa. 2003).

11. *Id.* at 441–42.

12. *See id.* at 528–29.

and excluded the scientific testimony on the grounds of its unreliability.[13]

One way to help bridge the gap between the cultures of science and law is to educate judges about the scientific methodologies used by epidemiologists and toxicologists (to name two prominent disciplines) in their studies of the possible causal relationships between exposures and human disease outcomes. The Federal Judicial Center, for example, has produced two editions of a reference manual on scientific evidence that describes how science "works," as well as a number of chapters on the techniques and methods of epidemiology, toxicology, and statistics.[14] Workshops on science in the courtroom, entitled "Science for Judges," have been offered to state and federal judges. Evidence features prominently in law school curricula, and a vast body of legal scholarship exists on published cases following *Daubert*. On the other side of the fence, the scientific literature has increasingly examined this intersection of law and science. For example, a recent supplement to the American Journal of Public Health discusses *Daubert* and related decisions.[15]

Despite all these efforts, the post-*Daubert* world has seen very little discussion of an important topic: the methods of evidentiary synthesis used by scientists to assess a body of evidence for the purpose of making claims about disease causation. These are the so-called "weight of evidence" or "synthetic methods" of epidemiology, toxicology, and other biomedical sciences. In the example of *Soldo*, these methods were implicitly used by the court-appointed scientific experts to assess the available evidence.[16]

13. *Id.* at 577. The court held that the scientific testimony offered by the plaintiff did "not satisfy the *Daubert*/Fed. R. Evid. 702 requirements of being scientifically reliable and having a valid scientific connection to the facts of this lawsuit." *Id.*

14. FED. JUDICIAL CTR., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE (2d ed. 2000); FED. JUDICIAL CTR., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE (1st ed. 1994). Visit the Federal Judicial Center's online Publications and Video Catalog at http://www.fjc.gov/library/fjc_catalog.nsf to view excerpts from the reference manual.

15. *See* Jasanoff, *supra* note 5, at 57 (arguing in *Daubert* the Supreme Court misconceived the nature of scientific practice and as a result "turn[ed] judges into naïve epistemologists, a role they are poorly suited to assume by training, skills, or knowledge" who unthinkingly apply the *Daubert* criteria as a checklist). This is a very interesting argument considering the ten-year effort to educate judges and others in the legal profession about the nature of science. *See* Joe S. Cecil, *Ten Years of Judicial Gatekeeping Under* Daubert, 95 AM. J. PUB. HEALTH 74, 74 (Supp. 1 2005) (arguing judges and others "inquire more deeply [than before] into the reasoning and methodology that supports the expert opinions").

16. These methods were likely set forth in the experts' reports. *See* Soldo v. Sandoz Pharms. Corp., 244 F. Supp. 2d 434, 442 (W.D. Pa. 2003) (noting that before the

Methods of synthesis can be distinguished from the general scientific methods discussed by philosophers of science. They can also be distinguished from the analytical methods scientists use to design and analyze data from individual studies. Put another way, there are three ways to categorize scientific methods. First, there are the general philosophical approaches to scientific method aligned with such notables as Karl Popper or Carl Hempel, two philosophers whose ideas were considered by the Supreme Court in *Daubert*.[17] These methods are often described in terms of their logical structure, for example, "inductive" or "hypothetico-deductive." Another more recent example is Lipton's "inference to the best explanation."[18]

A second type of scientific method is analytic. These more quantitative methods can represent specific study designs (e.g., the "case-control" study, the "cohort" study, or the randomized clinical trial); numerical (often statistical) methods, such as regression analysis or linkage analysis; or any number of techniques used in laboratory studies. DNA assays are an excellent example.

The third type of scientific method to be examined in some detail in this Article is more synthetic than analytic. These methods are typically applied to a body of evidence that may include studies from different scientific disciplines.[19] Synthetic methods, as an example, could examine a series of case control studies from epidemiology in combination with laboratory studies from toxicology. Synthetic methods can also be used to combine information (i.e., results) from several different yet similarly

Daubert hearing, the three experts submitted reports stating their opinions of causation surrounding Parlodel and the plaintiff's stroke).

17. Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 593 (1993) (invoking the works of Popper and Hempel when discussing whether a theory can be tested (citations omitted)).

18. PETER LIPTON, INFERENCE TO THE BEST EXPLANATION (2d ed. 2004).

19. The importance of these synthetic methods to the current legal framework ten years after *Daubert* cannot be overstated.

> Recent cases consider more difficult questions arising from the differing methodologies used in various areas of science. The current legal framework that assesses admissibility in terms of professional practice outside the courtroom is poorly suited to cases that require expertise across a wide range of specialties and force judges to choose from among different scientific methodologies.

Cecil, *supra* note 15, at 74. Synthetic methods do not force such choices but rather provide more or less reliable techniques for combining evidence from different relevant scientific disciplines.

designed studies within one discipline, such as a series of randomized clinical trials or observational (cohort) studies.[20] Important examples of synthetic methods include the systematic narrative review and the causal criteria of epidemiology. These, along with such methods as meta-analysis, are often referred to as "weight of evidence" (WOE) methods.[21]

As one of the first attempts to introduce the legal community to these important and less well-understood techniques of research synthesis, this Article offers a brief overview of a few key methods. In addition, I provide an assessment of their reliability, as a practicing scientist and public health professional, in the spirit of what is expected of trial court judges post-*Daubert*.[22] Because so many of these methods are captured under the umbrella of WOE, this Article begins with that expression.

II. WEIGHT OF EVIDENCE (WOE)

WOE is a widely used phrase in scientific and policy-making literature, often in the context of health risk assessment.[23] WOE, however, is also widely used in assessments of the efficacy of medical treatments, environmental hazards, forensic analyses, and in the social sciences applied to medical and public health issues.[24]

WOE has several distinct uses in contemporary scientific practice. First, it most often appears in a metaphorical sense, pointing to a body of scientific evidence without reference to any specific methodology.[25] Here, WOE could be replaced by "summary interpretation" or "synthesis" of the

---

20. *See* Douglas L. Weed, WEIGHT OF EVIDENCE: A REVIEW OF CONCEPT AND METHODS, 25 RISK ANALYSIS 1545, 1552 (Supp. 1 2005) (noting that weight of evidence "may . . . point to a number of longstanding interpretive methodologies or their combinations, or it may refer to innovative methods qualitatively or quantitatively combing several types of evidence").

21. *See* Sheldon Krimsky, THE WEIGHT OF SCIENTIFIC EVIDENCE IN POLICY AND LAW, 95 AM. J. PUB. HEALTH 129, 129 (Supp. 1 2005); Weed, *supra* note 24 at 1546, 1550.

22. If judges are expected to think as they imagine scientists think, then it may be helpful for them to see how a scientist (in this case, an epidemiologist) would assess the reliability of the synthetic methods of evaluating evidence employing the *Daubert* checklist. *See Daubert*, 509 U.S. at 593–94 (identifying factors to be considered in determining the admissibility of expert testimony pursuant to Rule 702 of the Federal Rules of Evidence).

23. *See, e.g.*, Krimsky, *supra* note 21 at 129.

24. *See, e.g.*, *id*. at 129–30 (footnotes omitted).

25. Weed, *supra* note 20 at 1546–47.

evidence without any change in meaning.[26] Second, in some situations, a WOE approach specifically refers to a technique in which "*all* available evidence" should be examined and interpreted.[27] This approach is usually contrasted with a "strength of evidence" approach that includes only selected studies, the selection of which typically is made on the basis of criteria such as quality, statistical significance, positive results, or some combination of these.[28] Third, often a WOE method refers directly to some other synthetic method, such as the systematic narrative review, meta-analysis, or the so-called "causal criteria" associated most often with the public health discipline of epidemiology.[29] Fourth, a WOE method may point to an institutional approach to synthesis, such as the risk assessment guidelines of the Environmental Protection Agency or the analogous guidelines for the assessment of chemical carcinogenicity used by the World Health Organization's International Agency for Research on Cancer, both of which combine human (epidemiological) evidence with laboratory-based (toxicological) evidence.[30] Finally, in relatively rare instances of health-risk assessment, a WOE approach involves a method that assigns numerical weights to individual scientific studies and creates summary numeric assessments using mathematical algorithms.[31]

Multiple definitions and uses of WOE are important problems for those charged with determining the extent to which methods employed to synthesize scientific evidence are generally accepted within the scientific community. There is no single, well-established WOE method for synthesizing scientific evidence.[32] Indeed, it is fair to say that, taken alone, a WOE method, without further specification, may have no direct relationship to either explicit weighting schemes or to other, better defined (although not necessarily widely-accepted) synthetic methods. The obvious solution to these problems is to require that anyone claiming to have produced a WOE assessment of general causation provide the details of his or her approach. Unfortunately, it is possible, indeed it is likely, that unspecified WOE methods have been described within the peer-reviewed scientific literature without adequate detail. Put another way, the fact that

26. *Id.* at 1546.
27. *Id.* at 1547.
28. *Id.* at 1547–48 (footnote omitted).
29. *Id.* at 1548.
30. *See id.* at 1550–51.
31. These five WOE uses emerged from a systematic review of the scientific literature as represented in PubMed, a database supported by the National Library of Medicine. *See id.* at 1545–46.
32. *See* Krimsky, *supra* note 21, at 129.

a peer-reviewed scientific publication describes the synthetic methodology as a WOE approach is no guarantee that the method actually used has been described, much less represents a well-accepted methodology. As noted in many peer-reviewed scientific publications, it is likely the expression is used solely in its metaphorical form.

III. SYSTEMATIC NARRATIVE REVIEW

A systematic narrative review is a type of summary literature review that puts special emphasis on the following: (1) a clear statement of its purpose; (2) the literature search methods used to identify the studies that may be included in the final synthesis of evidence; (3) inclusion/exclusion criteria for individual studies (typically based on the quality of those studies); and (4) any other synthetic methods used in the assessment of the evidence (e.g., meta-analysis or causal criteria).[33] Literature reviews, systematic or not, are a longstanding form of synthetic method and are ubiquitous in science, appearing in peer-reviewed journals, in textbook chapters, and in background material for research funding applications. The purposes of systematic reviews can include: claims regarding general causation, efficacy of therapeutic or preventive interventions, recommendations for such interventions, and recommendations for future research.[34]

Concerns about the quality of published literature reviews first surfaced in the late 1980s in medical practice journals; an analysis of published reviews in epidemiology revealed prominent lapses in quality ten years later.[35] Guidelines for writing systematic reviews of scientific literature likely have made improvements in the quality of systematic reviews of medical therapy, adverse effects of treatments, and economic evaluations of health care, although these reviews tend to focus on the published randomized clinical trials. The quality of reviews in epidemiology has not been examined in a decade, and it is unclear if the quality of published literature reviews in toxicology has ever been examined.

Problems continue to plague the assessment of the quality of systematic narrative literature reviews. Peer review is no guarantee that the review is of high quality, much less systematic. Methods for assessing

33. Weed, *supra* note 20, at 1548.

34. *Id*. at 1548–49.

35. *See* Rosalind A. Breslow, Sharon A. Ross & Douglas L. Weed, *Quality of Reviews in Epidemiology*, 88 AM. J. PUB. HEALTH 475, 475 (1998).

the quality of literature reviews are rudimentary at best and confuse a systematic review with a quantitative meta-analysis of studies. Finally, it is important to point out the fact that some scientists perceive any review of the literature to be a publication of dubious distinction, despite the critical importance of this form of research to the question of general causation.

IV. CAUSAL CRITERIA (HILL'S CRITERIA)

A list of considerations is typically used by epidemiologists to examine scientific evidence for the purpose of determining if an observed association between an exposure and a disease is causal.[36] Historically, evidentiary criteria relevant to infectious disease causation appeared in the scientific literature as early as the mid-nineteenth century. In the late 1950s and early 1960s, the epidemiologic community turned its attention to the so-called chronic (typically non-infectious) diseases—heart disease, cancer, and stroke, for example—examining how the earlier criteria could be used. In 1964, five criteria for causation were used in the U.S. Surgeon General's *Report on Smoking and Cancer*.[37] A year later, a British medical statistician, Sir Austin Bradford Hill, published a list of nine considerations that expanded upon the 1964 list.[38] These so-called "causal criteria," sometimes called "Hill's Criteria," remain an important synthetic method for epidemiology forty years after their original appearance.[39] The list of nine criteria includes: strength of association, consistency, specificity, temporality, dose-response, biologic plausibility, coherence, experimentation, and analogy.[40] It should be noted that at least one of these criteria, biologic plausibility, points directly to scientific evidence from other laboratory-based scientific disciplines such as toxicology or molecular biology. In other words, these criteria both synthesize evidence from different types of epidemiologic studies as well as synthesize evidence from different scientific disciplines. Typically, these criteria are used to assess causation after an association between the exposure and the disease has been established; that is, there are a sufficient number of published studies with statistically significant positive results to warrant a judgment

36. Douglas L. Weed, *Causation: An Epidemiologic Perspective (In Five Parts)*, 12 J.L. & POL'Y 43, 46 (2003) (footnote omitted) (enumerating the following considerations: consistency, strength of association, dose response, biological plausibility, coherence, temporality, specificity, and analogy).

37. *Id.* at 45 (footnote omitted).

38. Austin Bradford Hill, *The Environment and Disease: Association or Causation?*, PROC. ROYAL SOC'Y MED. 295, 295–99 (1965).

39. *See* Weed, *supra* note 36, at 46 n.8.

40. Hill, *supra* note 38, at 295–99.

that the association exists.[41] On the other hand, these criteria have traditionally been applied when little or no laboratory-based evidence exists, reflecting Hill's argument that an observed association may make biologic "sense" even in the absence of supporting laboratory-based evidence to that effect.[42] Recent descriptions of this criteria-based practice of causal inference have emphasized the significant variability that emerges due to three longstanding problems: (1) selection and prioritization of the criteria; (2) assigning rules of evidentiary decision-making to the criteria; and (3) the role of subjectivity and values in applying the method.[43]

V. ASSESSING THE RELIABILITY OF EVIDENCE SYNTHESIS METHODS

Daubert provides five factors for assessing the reliability of scientific testimony: testability (objective challenge), peer review and publication, rate of error, operating standards, and general acceptance in the scientific community.[44] Assuming no particular order or priority, I begin with the related issues of peer review and operating standards. In addition, I will focus only on the methods of systematic review and causal criteria, putting aside the WOE concept and its associated methods because there are so many meanings in current practice. Moreover, the systematic review and the causal criteria can be considered WOE methods as described.

Systematic review and causal criteria typically appear in peer-reviewed publications. Whether peer review provides a satisfactory level of assurance that either method meets standards of quality presumes such standards exist and that peer reviewers require authors adhere to those standards. Standards for systematic reviews have emerged in recent years,[45] and it is fair to say that the quality of these publications also has improved. These standards focus upon transparency—the purpose of the review, the literature search techniques used, the evidence that was included and excluded, reasons for exclusion, the extent to which bias and other systematic errors were considered, and, when appropriate, which

41. *See generally* Weed, *supra* note 36, at 49–50 (using the example of smoking and cancer).

42. Hill, *supra* note 38, at 295.

43. *See* Weed, supra note 36, at 46–48, 51–52.

44. Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 592–94 (1993) (citations omitted).

45. *See* Douglas L. Weed, *Methodologic Guidelines for Review Papers*, 89 J. NAT'L CANCER INST. 6, 7 (1997) (challenging scientists to disclose methods used to summarize evidence and prepare reviews).

additional synthetic methods were incorporated into the review.[46] One example of a synthetic method is the use of causal criteria or meta-analysis and specific parameters associated with those methods, such as the definitions and rules of evidence associated with each causal criterion.[47]

Standards for these same parameters (i.e., standard definitions and rules of evidence for the causal criteria), however, are not so clear-cut. Although the method is subject to peer review, there is a tendency for practitioners to exhibit considerable flexibility, especially in the rules of evidence assigned to the criteria, that smacks of subjectivity. Nevertheless, general practice patterns exist. Some criteria are regularly used, such as consistency, strength, dose-response, biologic plausibility, and temporality, to name those most commonly applied in cancer epidemiology.[48] There is, however, no standard subset. Standards for how each criterion is defined and what rules of evidence are assigned are also of interest; the definitions are reasonably standardized in practice, while the accompanying rules are less so.

Turning to testability and rates of error, the overall reliability of any scientific method can be assessed by comparing its results to those of a better method known as a "gold standard."[49] When such a standard methodology exists, then rates of error can, in principle, be calculated. No such gold standard exists for either of the synthetic methods described here. A discussion of how they might be objectively challenged can, nevertheless, be undertaken in more philosophical terms. General (philosophical) schema for critically assessing these methods have been discussed in epidemiological literature by examining the extent to which a method solves the problem it was intended to solve, identifying errors in its solutions, and eliminating those errors.[50] What could emerge from such

---

46. *See id.* at 6 ("The reader of any review should have a clear idea of the search techniques used, what evidence was assessed, and what evidence was excluded.").

47. *See id.* at 7.

48. Douglas L. Weed & Lester S. Gorelic, *The Practice of Causal Inference in Cancer Epidemiology*, 5 CANCER EPIDEMIOLOGY, BIOMARKERS & PREVENTION 303, 308 (1996).

49. *See* David L. Sackett et al., *Evidence Based Medicine: What It Is and What It Isn't*, 312 BRIT. MED. J. 71–72 (1996), *available at* http://bmj.bmjjournals.com/cgi/content/full/312/7023/71 (discussing the use of randomized trials as "gold standards" of evidence).

50. *See* Douglas L. Weed, *Methods in Epidemiology and Public Health: Does Practice Match Theory?*, 55 J. EPIDEMIOL COMMUNITY HEALTH 104, 108 (2001) (discussing the "[m]ismatches between methods and practice").

objective challenges are improved methods, potentially reducing some of the subjectivity that plagues their application.[51]

VI. CONCLUSION

Two methods of evidence synthesis used to assess general disease causation, the systematic narrative review and the causal criteria, are generally accepted in the scientific community, are subject to peer review, have explicit (sometimes implicit) standards of practice, and have no known rates of error, but can be objectively challenged at least in theory. It is tempting, therefore, to declare them reliable in a general sense, with much room for improvement. As a cautionary note, however, it is important to remember that considerable variability remains in how they are applied in practice. In a particular application of these methods, an assessment of reliability should be undertaken.

51. *See id.* (explaining the current situation and outlining what possible solutions will bring).