

THE NRC BULLET-LEAD REPORT: SHOULD SCIENCE COMMITTEES MAKE LEGAL FINDINGS?

D.H. Kaye*

ABSTRACT: For decades, analysis of concentrations of elements in bullet lead have been used in the United States to link defendants to crimes in which fragments of bullets have been recovered. In response to mounting criticism of the practice, the FBI commissioned the National Academy of Science to review the Bureau's procedures for making these measurements and drawing inferences from them. The Academy's report confirms the validity of the instrumentation but identifies weaknesses in the statistical methods for declaring matches and for describing the significance of these matches. It also concludes that courts correctly could apply existing legal doctrine to uphold testimony that a match makes it more probable (by an unspecified amount) that the defendant is the source of the fragments. The article questions the committee's legal reasoning with regard to the probative value of the evidence and its conclusion as to the admissibility of the evidence under the scientific-validity standard articulated in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* It also raises a question as to whether and how blue-ribbon

*Regents' Professor and Fellow, Center for the Study of Law, Science, and Technology, Arizona State University. This paper is an expanded version of a presentation at the Joint Statistical Meetings, Toronto, Canada, Aug. 2004, and the Sixth International Conference on Forensic Statistics, Tempe, Arizona, Mar. 2005. It does not consider the opinions and commentary published since these conferences. See *State v. Behn*, 868 A.2d 329, 331–32 (N.J. Super. Ct. App. Div. 2005) (concluding that “the expert testimony was based on erroneous scientific foundations and its admission met the requirements for granting a new trial on the ground of newly discovered evidence”); *Commonwealth v. Fisher*, 870 A.2d 864, 870 (Pa. 2005) (holding that a press release describing the NRC report did not provide a basis for questioning a capital verdict that was unavailable to a petitioner for post-conviction relief at the time of the trial); Michael O. Finkelstein & Bruce Levin, *Compositional Analysis of Bullet Lead as Forensic Evidence*, 13 J.L. & POL'Y 119 (2005). On September 1, 2005, the FBI announced that “after extensive study and consideration, it will no longer conduct the examination of bullet lead.” FBI Laboratory Announces Discontinuation of Bullet Lead Examinations (Sept. 1, 2005), available at http://www.fbi.gov/pressrel/pressrel05/bullet_lead_analysis.htm. The author is grateful to Paul Giannelli, Michael Finkelstein, and Jay Koehler for comments on a draft of this paper.

scientific committees studying forms of scientific evidence should offer explicit legal opinions on admissibility.

CITATION: D.H. Kaye, *The NRC Bullet-Lead Report: Should Science Committees Make Legal Findings?*, 46 *Jurimetrics J.* 91–105 (2005).

For decades, analyses of the composition of bullet lead have been used to identify the source of bullets.¹ In most cases, markings from a bullet's manufacture and its passage through the gun can be used to help associate it with a particular gun or set of guns.² When the bullet is too deformed or a bullet fragment is small, however, these clues are not available. Nevertheless, the lead alloy used in bullets contains a variety of trace elements, and the concentrations of these elements vary across production runs and individual bullets. If unused cartridges can be found in the possession of a suspect, similarities in these concentrations might forge a link between the crime-scene bullet and other bullets owned by the suspect.

Although simple enough in principle, implementing this approach is more complicated. The error in the measurements must be small relative to the variations in the concentrations of the elements across bullets, and such quantities must be known (or at least estimated) to understand the strength of inferences about the origin of the bullets. Despite a glaring lack of data on such matters, expert witnesses have been willing to testify that two bullets come from "the same manufacturer's source (melt) of lead"³ or even "the same box of ammunition."⁴ As the defense bar finally began to recognize the vulnerability of this kind of testimony,⁵ the FBI decided to address such issues. It commissioned a study by the National Research Council (NRC) of its methods for compositional analysis of bullet lead (CABL).⁶

The report of the NRC committee appeared last year.⁷ This complex and thoughtful report affirms the suitability of inductively coupled plasma optical

1. NATIONAL RESEARCH COUNCIL, COMMITTEE ON SCIENTIFIC ASSESSMENT OF BULLET LEAD ELEMENTAL COMPOSITION COMPARISON, *FORENSIC ANALYSIS: WEIGHING BULLET LEAD EVIDENCE 1* (2004) [hereinafter NRC REPORT].

2. *See, e.g.*, 1 J. HOWARD MATTHEWS, *FIREARMS IDENTIFICATION* (1962).

3. NRC REPORT, *supra* note 1, at 93.

4. *Id.* at 92; *see also* *Commonwealth v. Daye*, No. 11238-11246, 2005 WL 1971027 at *1 (Mass. Super. Ct., Aug. 3, 2005) ("Special Agent Riley testified as an expert as to how, thorough testing of the bullets, he was able to determine that one of the bullets found in Patricia Paglia's body and the bullet found in the basement of the Rye, N.H. home were 'from the same box of ammunition, or from another box of ammunition that was produced at the same place on or about the same date.'").

5. *See* Mark Hansen, *Bullet Proof*, A.B.A. J., Sept. 2004, at 58.

6. The FBI requested "an impartial scientific assessment of the soundness of the scientific principles underlying CABL to determine the optimum manner for conducting the examination and to establish scientifically valid conclusions that can be reached using the examination." NRC REPORT, *supra* note 1, at ix.

7. *Id.*

emission spectroscopy to measure the concentrations of seven elements in lead samples taken from bullets.⁸ It explains the dubious statistical procedures the FBI had been using to declare that a vector of measurements from one sample was essentially identical to that from another sample⁹ and proposes statistically superior approaches.¹⁰ The report also clarifies the statistical and logical inferences about the origin of two bullets that can be made from these data.¹¹ In all these regards, the report is exemplary.

However, the NRC committee also offers opinions on “the legal aspects of CABL evidence,”¹² especially the admissibility of this evidence in federal courts. This article describes the committee’s reasoning as to these legal issues. It finds the reasoning wanting in certain respects and suggests that courts should not accord the resulting conclusions great weight. It then reflects on a broader question—whether NRC committees empaneled primarily for their scientific expertise should opine on the admissibility of the forensic-science evidence that they have studied.

I. ANALYZING LOGICAL RELEVANCE

A. The Subtle CIVL

The legal portion of the report first argues that analytical chemists or technicians can provide relevant evidence by testifying that two bullets originate from the same “compositionally indistinguishable volume of lead,” or “CIVL.”¹³ The committee offers the concept of a CIVL as a more precise term than “source,” which might be a billet (made by remelting large blocks of smelted lead), or a wire (extruded from a billet or mix of billets), or something else.¹⁴ Yet, inasmuch as the boundaries of a CIVL never are known or measured, this concept is hardly free from ambiguity. A CIVL can range from “[a]t the very least, . . . several wires”¹⁵ to an entire “vat of molten lead whose composition is not altered during

8. *Id.* at 12–25.

9. These involved three seven-element measurements on three fragments from the two samples to be compared. *Id.* at 29. Thus, there were nine seven-element measurements per sample. The FBI had three rules for declaring matches. (1) For each element, use a within-bullet standard deviation (SD) of each bullet, form a two-SD window about each bullet’s mean, and declare a match if the two windows overlap (for every element). *Id.* at 30–31. (2) Look for an overlap between the two bullets in the range of the replicate measurements on each bullet. *Id.* at 31. (3) “Chaining”—a rather peculiar method that seems comparable to saying that if A is a friend of B, and if B is a friend of C, then A is a friend of C. *Id.* at 31–35.

10. The committee proposed computing a distance function for differences in the seven-element measurements or computing a series of *t*-statistics and using one or the other of these quantities to define a match. *Id.* at 35–70.

11. *Id.* at 71–85.

12. *Id.* at 85.

13. *Id.* at 85 (defining the term).

14. *Id.* at 82–84 (decrying “[t]he poor definition and understanding of the term ‘source’” and defining these terms).

15. *Id.* at 85.

Kaye

the pouring of bullets.”¹⁶ All we really know about a CIVL is that it is “produced during one production run at one point in time [and] is at least as large as the sample taken for analysis.”¹⁷ And even this characterization is to some extent a matter of faith.¹⁸

Having created the acronym, the committee leans on it in arguing that “matches” are relevant evidence within the meaning of Federal Rule of Evidence 401. It asserts that “[t]here are two aspects of relevance in this context: the likelihood that crime scene bullets came from the same CIVL as the defendant’s bullets and the likelihood that the crime scene bullets came from the defendant.”¹⁹ In reality, only the second proposition determines relevance. If the elemental measurements do not make it more (or less) probable “that the crime scene bullets came from the defendant,” then the laboratory findings are irrelevant—regardless of what one concludes about the “CIVLity” of the bullets being compared. The report, however, largely ignores the pertinent proposition in favor of the one that is more tractable.²⁰ To explain the logical problem here, I begin by outlining the committee’s likelihood ratio, then I return to the fundamental question of what it does and does not establish.

B. The Committee’s Likelihood Ratio

The NRC committee adopts the following definition of probative value:

A description of the probative force of evidence is given by the likelihood ratio for such evidence. The likelihood ratio for bullet lead match data is the probability that two bullets would match if they came from the same CIVL divided by the probability they would match (coincidentally or through error) if they came from different CIVLs.²¹

16. *Id.*

17. *Id.* (emphasis omitted).

18. Suppose that there is substantial inhomogeneity in the material being extruded. The elemental composition of the wire that emerges then could be highly variable. The report acknowledges that “[i]t is not clear from the data available whether the concentration of Sb is segregated in the billet or wire.” *Id.* at 83. It recognizes that “a paucity of data also exists for the spatial dependence of concentration of the other impurities along the length of the wire” *Id.* Nevertheless, it postulates that “their . . . lower concentration should make spatial inhomogeneities less likely.” *Id.* Indeed, the committee envisages a CIVL as a blob of well mixed molten material of appreciable volume, whose local concentrations of impurities do not change much from that point on. The report contains the following “finding”:

The committee’s review of the literature and discussions with manufacturers indicates [sic] that the size of a CIVL ranges from 70 lbs in a billet to 200,000 lbs in a melt. This is equivalent to 12,000 to 35 million 40-grain .22 caliber longrifle bullets from a CIVL compared with a total of 9 billion bullets produced each year.

Id. at 107; *see also id.* at 167.

19. *Id.* at 96.

20. The strategy is a common one—if you cannot answer the question, pick a different question that you can answer.

21. NRC REPORT, *supra* note 1, at 96.

It will be helpful to express this idea algebraically. To keep things simple, we can boil the situation down to making a single measurement (instead of three measurements) of the concentration of a single element (instead of seven elements) in each of the two bullets, B_1 and B_2 . Let x_1 denote the measurement on B_1 , and let x_2 be the measurement on B_2 . Let C_1 denote the CIVL from which B_1 emerged, and let C_2 be the CIVL from which B_2 originated. A match is declared if the difference between the two measurements, $d = |x_1 - x_2|$, falls below some more-or-less arbitrary threshold D . In this terminology, the committee's likelihood ratio is computed as

$$LR_{NRC} = \frac{P(d < D \mid C_1 = C_2)}{P(d < D \mid C_1 \neq C_2)}. \quad (1)^{22}$$

The report then labors to establish that this likelihood ratio exceeds 1 and hence supports the hypothesis that the bullets came from the same CIVL compared to the hypothesis that they came from two different CIVLs. Indeed, “for the sake of discussion,”²³ it floats a figure of 450 for this likelihood ratio or Bayes factor.²⁴

But LR_{NRC} is a poor measure of the probative value of the evidence for two reasons. First, even with respect to the hypothesis of CIVLity, it does not capture the richness of the data. Second, and more fundamentally, a high value for this likelihood ratio does not mean that the evidence is probative of the legally relevant proposition—the possible association between bullets from the crime scene and bullets in the defendant's possession.

C. The Likelihood *Ratio* and the “Probative Force” of Continuous Data

The first point, which is somewhat technical, arises when the committee abruptly switches from the frequentist analysis in the chapter on “Statistical Analysis” to the Bayesian perspective in the chapter on “Interpretation.” The legal chapter treats the data as if the chemists are measuring a categorical variable—either the bullets match or they do not. The underlying measurements, however, pertain to continuous elemental concentrations. The better the “match” (the closer that d is to zero), the stronger the support for the hypothesis $C_1 = C_2$. Values of d that are just above this critical value are practically as probative as values that just

22. One reviewer of this paper suggested that this equation reflects only the probability of a match as a result of coincidence and ignores the reference to “error” in the committee's verbal description. However, d is a measured quantity that is affected by error.

23. NRC REPORT, *supra* note 1, at 96.

24. The Bayes factor is the probability of a hypothesis in light of the data divided by its probability prior to the data. In the simple situation portrayed in this part of the report, the likelihood ratio is also the Bayes factor, and the committee depicts it as both. *Id.* at 96–97.

squeeze into the acceptance region. A likelihood function that reflected this fact could provide a better measure of probative value.²⁵

Probably, this does not matter. I am not enough of a statistician to do the Bayesian analysis of probative value correctly,²⁶ and hence to determine whether the committee's inefficient or inappropriate measure of probative value provides some kind of average probative value across all "matches." And, I do not know what to make of the committee's concession that "[h]ere, the likelihood ration [sic] is not defined strictly as statisticians would use the term, but in a way that has been acceptable in court."²⁷ If this concession is meant to imply that LR_{NRC} is not the correct measure of probative value of CABL evidence in a given case, then the criticism leveled above is simply that there is a gap in the committee's reasoning about likelihood and probative value. If LR_{NRC} is not the likelihood ratio as

25. This point was well-ventilated in connection with the presentation of similarities and differences in the molecular weight measurements obtained from gel electrophoresis of DNA fragments in forensic settings. See Donald A. Berry, Comment, 9 STAT. SCI. 252 (1994); D.A. Berry et al., *Statistical Inference in Crime Investigations Using Deoxyribonucleic Acid Profiling*, 41 APPLIED STAT. 499 (1992); J. Buckleton et al., *A Continuous Model for Interpreting the Positions of Bands in DNA Locus-Specific Work*, 31 J. FORENSIC SCI. SOC'Y 353 (1991); A. Collins & N.E. Morton, *Likelihood Ratios for DNA Identification*, 91 PROC. NAT'L ACAD. SCI. 6007 (1994); B. Devlin, *Forensic Inference from Genetic Markers*, 2 STAT. METHODS IN MED. RES. 241 (1993); B. Devlin et al., *Forensic Inference from DNA Fingerprints*, 87 J. AM. STAT. ASS'N 337 (1992); Ian W. Evett et al., *An Illustration of the Advantages of Efficient Statistical Methods for RFLP Analysis in Forensic Science*, 52 AM J. HUM. GENETICS 498 (1993); David Jarjoura et al., *Likelihood Ratios for Deoxyribonucleic Acid (DNA) Typing in Criminal Cases*, 39 J. FORENSIC SCI. 64 (1994); D.H. Kaye, *The Relevance of "Matching" DNA: Is the Window Half Open or Half Shut?*, 85 J. CRIM. L. & CRIMINOLOGY 676 (1995).

26. Assume, for concreteness, that the measured values of the single-element concentration are observations of *normal* random variables with means μ_1 and μ_2 , and variances σ_1^2 and σ_2^2 , respectively. The means reflect the elemental concentration in the CIVL for each bullet, and the variances come from small inhomogeneities within a CIVL and from instrumental measurement error. The observed difference $d = |x_1 - x_2|$ is then normally distributed with mean $\delta = |\mu_1 - \mu_2|$ and variance $\sigma^2 = \sigma_1^2 + \sigma_2^2$. The probability density of d is thus $f(d; |\mu_1 - \mu_2|, \sigma_1^2 + \sigma_2^2)$, where f is the normal probability density function. That is,

$$f(d; \delta, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2}\left(\frac{d-\delta}{\sigma}\right)^2\right].$$

Given an estimate of σ , one can use Bayesian or frequentist procedures to estimate the value of δ , but this does not reveal whether the bullets came from the same CIVL—only whether they came from CIVLs for which $\mu_1 = \mu_2$.

This sticking point is worth belaboring. If B_1 and B_2 come from the same CIVL (if $C_1 = C_2$), then $\delta = 0$, but the converse does not hold. A CIVL from a melt last month, for instance, could have the same elemental concentrations as a CIVL from another melt this month. It is easy enough to write the likelihood function when $\delta = 0$, but the alternative hypothesis $C_1 \neq C_2$ is composite and does not correspond to any specific function. Perhaps the best one can do is to reframe the hypothesis in terms of bullets derived from CIVLs having the same means. This would make the problem tractable, but it would not indicate the probative value of the data with regard to the hypotheses $C_1 = C_2$ versus $C_1 \neq C_2$.

27. NRC REPORT, *supra* note 1, at 96 n.110.

understood by statisticians, how do we know it functions as a properly computed likelihood ratio would? Perhaps this gap can be filled, and it can be established that LR_{NRC} suffices in lieu of a more appropriate likelihood function for thinking about relevance *vel non*.

If, however, the statement is supposed to mean that because legal cases have considered likelihood ratios for two hypotheses rather than likelihood functions for the parameters of a model, it is desirable to do the same thing here, then the rationale is without merit. One ought not to force large square pegs into small round holes. If one is going to use a Bayes' factor to describe "probative force," the law surely permits (and probably requires) statisticians to use the one that applies to the problem at hand. There is no court opinion that requires statisticians to use methods designed for categorical data with continuous data, and the NRC report cites no cases that find "acceptable" any sort of likelihood ratio, let alone one that applies only to categorical testimony about a match when it is clear that the degree to which items match is important.²⁸ One can choose to simplify and condense, of course, but one of the more compelling reasons for using likelihood methods is to cope with variables, such as the refractive index of glass, that do not lend themselves to simple yes-no, match-nonmatch reports.²⁹

D. Same Elemental Composition, Different CIVLs

As indicated above, the fact that two bullets appear to have the same elemental composition does not mean that they are from the CIVL. CIVLs from different times or places could happen to have the same elemental concentrations. But how likely is *this*? The report merely asserts that if two bullets that have the approximately same elemental concentrations, they are more likely to be from the same CIVL than from different CIVLs.³⁰ At first glance, this claim seems too obvious to dispute. By definition, a CIVL is so homogeneous that differences within it are undetectable. But how much overlap is there among different CIVLs? The report states that "[t]he frequency of coincidentally identical CIVLs is unknown."³¹ Apparently, the production process could be characterized by a few large, "coincidentally identical CIVLs" along with thousands of smaller, highly variable ones. In the absence of a more definitive study of the issue, the basis for the committee's conclusion that its likelihood ratio justifies an inference of CIVLity is unclear.³²

28. For descriptions of cases in which testimony on likelihoods has been introduced, see D.H. KAYE ET AL., *THE NEW WIGMORE, A TREATISE ON EVIDENCE: EXPERT EVIDENCE* § 12.4.2(b) (2004).

29. See *id.* § 12.4.2(a); Geva Maimon et al., *Assessing Spatial Heterogeneity in the Refractive Index of Float Glass*, 46 *JURIMETRICS J.* ___ (2005).

30. NRC REPORT, *supra* note 1, at 97–99.

31. *Id.* at 99.

32. The report describes an FBI study as indicating that "the frequency of coincidental false positives is quite low," but it then criticizes the study for being based on a statistically biased sample, for not being conducive to "quantifiable error rates," and for not having been published. *Id.* A revised version of the study recently was published as Robert D. Koons & JoAnn Buscaglia, *Forensic*

E. Same CIVL, Different Endpoints

Suppose it were clearer that small differences in the elemental concentrations of two bullets implied a common CIVL origin—that this much was, in the committee’s words, a “scientifically supportable conclusion.”³³ Would a small value for d then warrant an inference to the conclusion that B_1 and B_2 once were in the defendant’s possession? Is it safe to move from “same CIVL” to “same box” or even “same or nearby boxes”?

To be clear about the pertinent propositions, symbols are again helpful. Let the difference in the measured concentrations again be d ; let the hypothesis be that B_1 and B_2 were nearby one another at the endpoint of the distribution system for bullets be N ; and let the hypothesis that they were not near (they were far) be F . Then the probative value of d (with regard to *these* hypotheses) is given by the following likelihood ratio:

$$LR = \frac{P(d|N)}{P(d|F)}. \quad (2)$$

How LR might be realistically estimated is not obvious, but it depends in part on how bullets are packed and shipped—on how they move from the production line, to the wholesaler, to the retailer, to the customer. Figure 1 illustrates the problem. It shows how the mixing of bullets from different CIVLs can lead to bullets from many CIVLs in many boxes, frustrating any assumption that two

Significance of Bullet Lead Compositions, 50 J. FORENSIC SCI. 341 (2005). Although the researchers caution that their data set “does not, nor is it meant to, represent any production volumes, or geographic or temporal distributions of bullets, such as would be required for calculations of frequency of occurrence in a general bullet population,” they conclude that “[f]or a representative sampling of all bullets involved in FBI casework over the 1989 to 2002 period in the U.S., the probability of two unrelated bullets having indistinguishable compositions is on the order of 10^{-3} to 10^{-5} .” *Id.* at 343, 350. They offer the following, more general remarks about efforts to estimate coincidental match probabilities with trace evidence:

We believe that it is neither possible nor appropriate to calculate reliable probabilities of chance occurrence of indistinguishable items of non-biological trace evidence. This is particularly true for manufactured items and for comparison of highly discriminating variables such as elemental concentrations. Reasons for this include the lack of databases that are both of sufficient size and representative of the distributions of the measured variables in the evidentiary material. For items whose measured characteristics in the population are either geographically or temporally variable, such databases are impossible to obtain. This fact has proven to be one of the greatest drawbacks preventing the adoption of probabilistic methods of evaluating evidentiary significance. However, the inability to acquire databases appropriate for precise frequency of occurrence calculations does not mean that the significance of two samples having indistinguishable properties is low, only that it cannot be calculated precisely. The inability to calculate precise and accurate significance statistics applies generally to all comparative characteristics of trace evidence, except possibly those whose distributions are controlled by time-invariant genetic rules (e.g., biomarkers), or whose characteristics are deemed to be unique (e.g., fingerprints, toolmarks). Despite the lack of precise significance statistics, comparative trace evidence examinations continue to provide highly significant conclusions and meaningful assistance to all participants in the legal process.

Id. at 350.

33. NRC REPORT, *supra* note 1, at 96.

bullets are more likely to end up in the same box if $C_1 = C_2$ than if $C_1 \neq C_2$. Yet, if $C_1 = C_2$ does not indicate an enhanced probability that B_1 and B_2 resided in the same box—in the box owned or accessed by the defendant—then there is *no* probative value in the evidence said to prove that $C_1 = C_2$. In other words, even if a small value of d is highly probative of CIVLity, it may be entirely nonprobative as to the proposition N that is of interest to the jury—whether the two bullets were near neighbors at the end of the production and distribution chain.

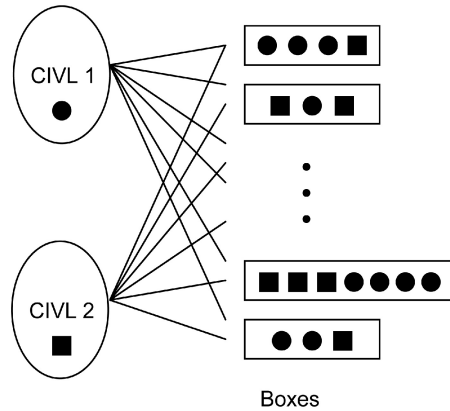


Figure 1. How Bullets from Two CIVLs Can Be Arranged in Many Boxes

The committee confessed ignorance as to how likely it is that bullets from one CIVL segregate in the same boxes, and it is probably the case that no one knows.³⁴ In the absence of such knowledge, the relevance of the testimony endorsed by the committee—that $LR_{NRC} > 1$ and that this favors the view that the bullets came from the very same CIVL³⁵—is unclear. Suppose, for instance, that the crime-scene bullet is the type depicted as a square in Figure 1. In other words, it is definitely from CIVL 2.³⁶ Because a great many boxes (presumably possessed by a great many gun owners) contain such bullets, it is not clear that the conclusion suggested by the NRC’s version of the expert testimony makes it even slightly more probable that the defendant’s box is the source of the crime-scene bullet. Of course, the details of Figure 1 are unreal, but the general picture it depicts is consistent with the report’s observations that “[i]t is routinely found that a single box contains multiple distinct compositional groupings—as many as 14,”³⁷ and

34. *See id.* at 82.

35. The report approves of testimony “that having CABL evidence that two bullets are analytically indistinguishable increases the probability that two bullets came from the same CIVL, versus no evidence of match status.” *Id.* at 107.

36. This assumes that, as in Figure 1, every other CIVL is distinguishable from CIVL 2.

37. *Id.* at 84.

that “distribution can lead to varied scenarios regarding retail dispersion of bullets from a distinct compositional group.”³⁸ As such, the report clearly recognizes that “[a] conclusion that two bullets came from the same melt does not justify an expert in further testifying that this fact increases the odds that the crime bullet came from the defendant.”³⁹ But if this is so, it would appear that CIVLity is logically irrelevant.

The report stops short of this necessary conclusion. Instead, what it taketh with one hand, it giveth with the other: “The large number of bullets made from a single melt and the absence of information on the geographic distribution of such bullets precludes such testimony *as a matter of expertise*. Such an inference is a matter for the jury.”⁴⁰ Left unexplained is how a judge or juror can conclude that a small *d* shifts the probability that the defendant had access to the crime-scene bullet when the committee failed to spell out the necessary connection between LR_{NRC} and *LR*. As a result, the treatment of the logical relevance and probative value of the evidence seems incomplete and unpersuasive.

To rescue the committee’s conclusion, one would have to argue that, in general, when bullets come from the same CIVL, it is more likely that they will wind up in the same box. If this is usually true, then CIVLity increases the probability that the defendant, who had a similar bullet on his person or property, is associated with the crime-scene bullet. As noted above, the report does not defend this assumption, and the committee regards it as outside the scope of expert knowledge. If, however, it is something that is reasonable for anyone to assume, then the committee’s willingness to let a jury rely on it is more plausible.⁴¹

II. AVOIDING *DAUBERT* AND *KUMHO TIRE*

In addition to opining that CABL evidence is logically relevant, the NRC committee concludes that, this evidence, with certain statistical refinements, satisfies the special standard articulated in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*⁴² for the admission of scientific evidence. The report nicely describes the criteria for judging scientific validity articulated in *Daubert*, which concerned an alleged causal link between the anti-nausea drug, Bendectin, and congenital limb defects. These criteria were reiterated in *Kumho Tire Co. Ltd. v. Carmichael*,⁴³ a case involving engineering testimony on the cause of a tire blowout.

38. *Id.*

39. *Id.* at 102.

40. *Id.* (footnotes omitted)

41. If true, the assumption enables a CABL match to satisfy the minimal requirement of logical relevance. Further analysis would be required to establish the admissibility of the “match” under the common-law balancing test codified in Federal Rule of Evidence 403.

42. 509 U.S. 579 (1993).

43. 526 U.S. 137 (1999).

Having outlined these and other pertinent legal rules, the report then evinces considerable skepticism as to whether *any* of the *Daubert* factors warrants admitting CABL evidence. The committee writes:

- “Additional testing would be needed to fully satisfy the . . . testing requirement.”
- “[T]he subject has not received . . . broad [peer] review [M]ore such work would be needed to provide a strong basis for this admissibility factor.”
- “[T]he . . . factor requiring a theory to have a known error rate is only partially satisfied.”
- “The [factor requiring the] existence and maintenance of standards controlling the technique’s operation . . . in significant part is not satisfied.”
- “The analytical technique . . . has general acceptance of the scientific community However, . . . the FBI is the only laboratory performing this type of lead analysis for forensic use, so any inquiry into ‘general acceptance’ will not provide the broad consensus that this factor assumes.”⁴⁴

These are the kinds of statements one would expect to find in legal briefs or judicial opinions. Some are more defensible than others (it is, for instance, far from clear that there can be no “general acceptance” as that term is used in the case law, just because the number of laboratories using a technique for forensic purposes is small), but one also would expect such a string of consistently negative assessments to be followed by the conclusion that the technique fails to satisfy the *Daubert* standard. The committee, however, draws back from this precipice. It states that “[t]he fact that the specifically mentioned *Daubert* factors are not fully satisfied does not mean that CABL evidence should not be admitted”⁴⁵ It reads *Kumho Tire* as giving a court license to admit scientific evidence that flunks—in whole or “in significant part”—each and every one of the criteria for admissibility set forth in *Daubert*.

This is not a plausible reading of *Kumho Tire*. *Kumho* merely holds that a trial judge evaluating the admissibility of nonscientific expert testimony “*may consider*” the specific factors enumerated in *Daubert* for passing on scientific evidence.⁴⁶ In this context, *Kumho* states that “the test of reliability is ‘flexible,’ and *Daubert*’s list of specific factors neither necessarily nor exclusively applies to all experts or in every case.”⁴⁷ A farmer who testifies as an expert on which crops grow best in local fields need not publish his theories and findings in professional journals. But the testimony of an analytical chemist or forensic scientist as to the results of Optical Emission Spectroscopy of bullet lead and Hotelling’s T^2 test (or any other hypothesis test) lies at the very core of *Daubert*. If the *Daubert* criteria can be ignored in such a case, then the factors that the *Daubert* Court took the trouble to describe have no power. If they do count but are outweighed by other considerations, then those countervailing factors must be

44. NRC REPORT, *supra* note 1, at 100–01.

45. *Id.* at 101.

46. D.H. KAYE ET AL., *supra* note 28, at § 9.1.

47. NRC REPORT, *supra* note 1, at 101 (quoting *Kumho Tire*).

identified and weighed against the poor showing of CABL regarding the enumerated factors. The *Daubert* standard is flexible but not infinitely elastic. The committee's approach stretches it beyond the breaking point.⁴⁸

III. OF SCIENCE AND LEGAL ANALYSIS: A CALL FOR MODESTY

The legal analyses in the NRC report raise a wider question than the admissibility of CABL evidence. The National Academies have convened four committees to report on forensic science techniques. The first committee produced a report in 1979 on voice spectrographic identification.⁴⁹ Although the report described legal standards for admitting scientific evidence and the factual issues that needed to be resolved in light of these standards,⁵⁰ and it urged that "forensic applications be approached with great caution,"⁵¹ it purportedly took "no position on admissibility"⁵² and took pains to differentiate between "value judgment[s] . . . properly made by the judicial or legislative body" and scientific

48. The report does not actually *advocate* admitting even the watered-down version of CABL evidence that it deems scientifically sustainable. Its legal conclusion is a double negative: The failure to meet any of the *Daubert* criteria does *not* mean that a court should *not* admit the evidence. This is a remarkably lenient standard. As three concurring Justices remarked in *Kumho*:

[T]he opinion of the Court . . . makes clear that the discretion it endorses—trial-court discretion in choosing the manner of testing expert reliability—is not discretion to abandon the gatekeeping function. [I]t is not discretion to perform the function inadequately. Rather, it is discretion to choose among *reasonable* means of excluding expertise that is *fausse* and science that is junky. Though, as the Court makes clear today, the *Daubert* factors are not holy writ, in a particular case the failure to apply one or another of them may be unreasonable, and hence an abuse of discretion.

Kumho Tire Co., Ltd., v. Carmichael, 526 U.S. 137, 158–59 (1999) (Scalia, J., concurring).

49. NATIONAL RESEARCH COUNCIL, COMMITTEE ON EVALUATION OF SOUND SPECTROGRAMS, ON THE THEORY AND PRACTICE OF VOICE IDENTIFICATION (1979).

50. *Id.* at 38–57, 62–63.

51. *Id.* at 2. The committee pointedly wrote that:

The practice of voice identification rests on the assumption that intraspeaker variability is less than or different from interspeaker variability. However, at present the assumption is not adequately supported by scientific theory and data. Viewpoints about probable errors in identification decisions at present result mainly from various professional judgments and fragmentary experimental results rather than from objective data representative of results in forensic applications. The Committee concludes that the technical uncertainties concerning the present practice of voice identification are so great as to require that forensic applications be approached with great caution.

Id. The committee also opined as follows:

Estimates of error rates now available pertain to only a few of the many combinations of conditions encountered in real-life situations. These estimates do not constitute a generally adequate basis for a judicial or legislative body to use in making judgments concerning the reliability and acceptability of aural-visual voice identification in forensic applications.

Id. at 60.

52. *Id.* at 68. However, the report does recommend that voice spectrographic analysis not be admitted as evidence without an explanation of "the inherent limitations in the method and in the performance of examiners . . . in order to protect against overvaluation of such evidence." *Id.* The committee further states that "this [procedure] can best be achieved through the testimony of opposing experts, or perhaps through the testimony of an expert appointed by the court to explain the limitations of voicegram evidence." *Id.* at 69.

and statistical knowledge.⁵³ A 1992 report on DNA identification technology went to the other extreme, advising courts that they could take judicial notice of certain facts⁵⁴ and (in some instances) not clearly distinguishing between questions of legal policy and scientific practice.⁵⁵ A follow-up DNA report,⁵⁶ which appeared four years later, eschewed overtly legal conclusions⁵⁷ but included some tendentious suggestions.⁵⁸ The latest effort on bullet lead delivered twenty-three “major findings and recommendations,”⁵⁹ of which about a quarter are judgments of law.⁶⁰

Should NRC committees try to confine themselves to assessments of science and to avoid offering overt legal opinions (unless the latter is actually part of their charge⁶¹)? Or is the effort to avoid explicit judgments as to what courts can or should decide about the admissibility of scientific testimony a naive and foolish

53. *Id.* at 63.

54. NATIONAL RESEARCH COUNCIL, COMMITTEE ON DNA TECHNOLOGY IN FORENSIC SCIENCE, DNA TECHNOLOGY IN FORENSIC SCIENCE 23 (1992) (recommending that “[c]ourts should take judicial notice of three underpinnings of DNA typing . . .”), 133 (describing what should be judicially noticed “even in the context of a *Frye* hearing”). Judicial notice is a doctrine that allows a court to dispense with case-specific proof of a fact that is not subject to reasonable debate. *See* 2 MCCORMICK ON EVIDENCE §§ 328–30 (John W. Strong ed., 5th ed. 1999).

55. *See* D.H. Kaye, *The Forensic Debut of the National Research Council’s DNA Report: Population Structure, Ceiling Frequencies and the Need for Numbers*, 34 JURIMETRICS J. 369 (1994) (criticizing the report’s assertion that “[t]o say that two patterns match, without providing any scientifically valid estimate (or, at least, an upper bound) of the frequency with which such matches might occur by chance, is meaningless,” COMMITTEE ON DNA TECHNOLOGY IN FORENSIC SCIENCE, *supra* note 54, at 74). This assertion was taken by many courts to mean that scientific practice dictates the legal conclusion that quantitative estimates of random match probabilities must be part of the testimony about a DNA match. *Id.*

56. NATIONAL RESEARCH COUNCIL, COMMITTEE ON FORENSIC DNA SCIENCE: AN UPDATE, THE EVALUATION OF FORENSIC DNA SCIENCE (1996).

57. The committee’s eleven recommendations are collected *id.* at 4–8. None involve legal conclusions.

58. *E.g., id.* at 199 (observing that “[t]he argument that jurors will make better use of a single figure for the probability that an innocent suspect would be reported to match never has been tested adequately. The argument for a single figure is weak . . .”) (footnote omitted).

59. NRC REPORT, *supra* note 1, at 109–13.

60. For example, based on its understanding of Federal Rule of Evidence 403 and the common law principles that this rule codifies, the committee offers formal recommendations as to the testimony that should and should not be allowed or likewise required on direct examination. *Id.* at 112–13.

61. The NRC report states that “the National Research Council . . . was called on to evaluate CABL scientifically, statistically, and legally.” *Id.* at 10. Yet, the only reference to a legal evaluation in the actual “questions in the statement of task accepted by the NRC” occurs in the following question: “Can significance statements be modified to include effects of such factors as the analytical technique, manufacturing process, comparison criteria, specimen history, and legal requirements?” *Id.*; *cf. supra* note 6 (quoting more of the charge to the committee).

Even if a sponsoring agency starts out seeking an opinion on the admissibility of evidence or other interpretations of law, one might wonder whether the National Academies should accept the task. These organizations are excellent sources of scientific expertise. They are less naturally suited to undertaking legal analyses, especially with committees composed predominantly of scientists who are not familiar with legal methodology.

effort to separate values from facts? Certainly, it has been argued that the line between facts and values is blurry and that technology assessment is as much politics or policy as it is objective analysis.⁶² It would be foolish to deny that social and personal values influence “pure” scientific judgments. The evolution of scientific opinion is a social process,⁶³ influenced but not always fully determined by logic and evidence—particularly in the short-term and especially in regard to “trans-scientific” questions.⁶⁴

That said, it surely is possible for a scientific advisory committee to lay out the facts as it sees them, leaving it to law enforcement agencies, lawyers, judges, and others in the legal system to make appropriate use of these assessments.⁶⁵ Although the refusal to make legal judgments can be more a matter of form than substance, overt conclusions about the application of legal standards are likely to be controversial. The National Academies rarely will be asked to study forensic-science methods unless or until these methods are successfully challenged in some court cases and a prominent division of opinion emerges in the scientific community. In this milieu, criticisms about the legal reasoning in a predominantly scientific report can divert attention from or diminish the impact of the report’s scientific assessments and contributions.

In the end, my point is a narrow one. I am not arguing that lawyers and the law have no role in reports on CABL, voice identification, DNA profiling, or

62. Such arguments were raised in opposition to proposals for a “Science Court” consisting of scientifically trained judges who would be charged with making factual findings about science-laden trial issues such as the health effects of silicone-gel breast implants or mobile cell phones. The idea of a science court was advanced in the late 1960s and 1970s by Arthur Kantrowitz. See Arthur Kantrowitz, *Controlling Technology Democratically*, 63 AM. SCIENTIST 505 (1975); Arthur Kantrowitz, *Proposal for an Institution for Scientific Judgment*, 153 SCIENCE 763 (1967); Arthur Kantrowitz, *Science Court Experiment*, 194 SCIENCE 6 (1976); Arthur Kantrowitz, *The Science Court Experiment: Criticisms and Responses*, 33 BULL. ATOM. SCIENTISTS 44 (1978).

63. Naturally, different authors have different views on which norms and processes are dominant. See generally, e.g., PHILIP KITCHER, *THE ADVANCEMENT OF SCIENCE: SCIENCE WITHOUT LEGEND, OBJECTIVITY WITHOUT ILLUSIONS* (1993); JOHN ZIMAN, *REAL SCIENCE: WHAT IT IS, AND WHAT IT MEANS* (2000); JOHN M. ZIMAN, *AN INTRODUCTION TO SCIENCE STUDIES: THE PHILOSOPHICAL AND SOCIAL ASPECTS OF SCIENCE AND TECHNOLOGY* (1984); Robert K. Merton, *The Matthew Effect in Science*, 159 SCIENCE 56 (1968).

64. See Alvin Weinberg, *Science and Trans-Science*, 10 MINERVA 209 (1972); see also Letter from B. John Garrick, Chairman, Advisory Committee on Nuclear Waste, to Shirley Ann Jackson, Chairman, U.S. Nuclear Regulatory Commission (June 4, 1999), available at <http://www.nrc.gov/reading-rm/doc-collections/acnw/letters/1999/1090144.html> (“[A] question is trans-scientific if it can be posed properly and rigorously within the established scientific framework but cannot be answered by science, given current knowledge and techniques and given reasonable resources.”). The validity of linearly extrapolating from high-dose effects of ionizing radiation to very small doses has been presented as just such a question because “effects of low doses are so small that sample sizes required to discern an effect given natural variability are impossibly large,” making impossible to resolve the “intense and polemical” “disagreements both inside and outside the scientific community over the evidence that either supports or contradicts the LNTH [Linear No Threshold Hypothesis].” *Id.*

65. See James A. Martin, *The Proposed “Science Court,”* 75 MICH. L. REV. 1058 (1977); Jeffrey N. Martin, Note, *Procedures for Decisionmaking Under Conditions of Scientific Uncertainty: The Science Court Proposal*, 16 HARV. J. ON LEGIS. 443 (1979).

other forms of scientific evidence. To the contrary, the science and technology in these areas must develop with the needs and constraints of the legal system in mind.⁶⁶ An appreciation of the legal context and the manner in which experts have been testifying can influence the recommendations for presenting scientific evidence accurately and appropriately. In addition, explicating which findings about the scientific technique feed into which components of the legal rules determining the admissibility of evidence can make the findings as to the science more useful to courts. However, taking the further step of drawing debatable legal conclusions invites trouble without adding much value. Perhaps one lesson to be drawn from the latest NRC report is that it is perilous or unnecessary for a blue-ribbon scientific committee—even one that includes a sprinkling of outstanding lawyers or law professors—to proffer overt opinions on whether legal criteria for the admissibility of scientific evidence are satisfied.

⁶⁶ For example, the Neymann-Pearson hypothesis-testing approach pursued in the bullet-lead report has proved quite useful in some contexts, but arguably it is less well suited to courtroom use. *See generally, e.g.*, BERNARD ROBERTSON & G.A. VIGNAUX, *INTERPRETING EVIDENCE: EVALUATING FORENSIC SCIENCE IN THE COURTROOM* (1995); D.H. Kaye, *Is Proof of Statistical Significance Relevant?*, 61 WASH. L. REV. 1333 (1986).