References

1. Keto RO. Analysis and comparison of bullet leads by inductively-coupled plasma mass spectrometry. J Forensic Sci 1999;44(5):1020–6.
2. Box GEP, Tiao GC. Bayesian inference in statistical analysis. Wiley, 1992.
3. Schmitt SA. Measuring uncertainty: An elementary introduction to Bayesian statistics. Addison-Wesley, 1969.

Donald I. Promish, M.S.
68 Richardson St.
Burlington VT 05401-5002

Author's Response

Sir:

Mr. Promish appears to have missed one of the major points put forth in my article. Many of the bullets within a box do not resemble others from the same box any more than they resemble bullets from other manufacturers. This was illustrated graphically in Fig. 2 on page 1024. Calculating an "average" elemental composition for a box, and then showing that individual bullets from that box more closely resemble that average than they do averages of other boxes is meaningless. What matters, in a real case scenario, is whether known and questioned bullets match.

The purpose of bullet lead analysis, in most situations, is not to brand identify unknown bullets, as suggested in the final paragraph of Mr. Promish's letter. This is because the trace elemental composition of bullets from a single manufacturer can vary almost continuously over time, depending on the source for the lead. The analyst's opinion is usually limited to stating that the suspect bullet is consistent with (or could have come from) the same source as the known bullet. This is the same standard that applies to other forms of trace evidence as well, such as paint chips, glass fragments, and fibers, in the absence of a physical match. The "scalar products," or correlation values, were calculated in an effort to quantitate the quality of the match between two bullets, and arrive at a means of unambiguously distinguishing between known matching and non-matching bullet leads.

While statistical approaches to data interpretation may be useful, they are, as stated in the final paragraph of my paper, "not a substitute for direct comparison of the raw elemental data when formulating an opinion as to the similarity of two bullets". A 95% probability that two bullets match means little when a look at the raw data shows that they could not be from the same melt.

I appreciate Mr. Promish's enthusiasm, and applaud his volunteering his probabilistic approach. More of this type of thinking needs to be applied to the forensic sciences. However, it must be applied with caution. I question the advisability of using the probability of a bullet match as a decision threshold for arrest, prosecution, or finding of guilt. Such decisions can only be based on a much broader scope of evidence, which could include bullet comparison. The bullet analyst cannot be expected to give a qualified opinion as to the certainty of a match; his findings must be either positive or negative to be of use.

Raymond O. Keto, M.F.S.
National Laboratory Center
Bureau of Alcohol, Tobacco, and Firearms
Rockville, MD 20850

Verbal Conventions for Handwriting Opinions

Sir:

A paper of mine in Science and Justice (1) on reporting conventions was recently the subject of a spirited debate on Docexam_L, the forensic document examiners e-mail forum (membership enquiries to andersonc@docexam.com.au). The rough and tumble of an e-mail discussion is all very well but it seems to me that if the core issues are to be resolved then the appropriate forum lies within the pages of a peer-reviewed journal. Nothing of what I say here is original and the subject has been covered in greater detail elsewhere but it seems to me to be appropriate that I should state my case in the journal that carried the letter that announced the reporting convention with which I take issue.

The letter from McAlexander, Beck and Dick (2) is to be applauded for its motivation. It promotes the idea that there is a need for standardization of terminology among experts when they express opinions. It also argues convincingly for the need to consider handwriting evidence probabilistically. Rightly, the authors pointed to the weaknesses of phrases which appear, regrettably, to be in widespread use in the forensic science world: I refer, in particular to the use of "could have" and "consistent with." I agree with McAlexander et al. that these phrases should have no place in any convention for expressing the weight of an item of scientific evidence.

The letter described a reporting convention which became the subject of ASTM standard E 1658–96 (3). My copy is headed "Standard Terminology for Expressing Conclusions of Forensic Document" (sic).

Laudable though the attempt at standardization undoubtedly is, I must point out that there is a serious problem with the manner in which the convention uses the notion of probability. In my opinion, the reporting convention is incompatible with a logical approach to evidence interpretation. Whether or not one agrees with me that this is a problem depends on whether or not one wishes to view handwriting comparison as having a sicentific rationale. If is scientific, it has to be logical; it follows that if probability is to be invoked, then the laws of probability cannot be violated.

Probabilistic thinking in relation to forensic science has, until comparatively recently, been seen to be something that evolved in the 1970's, when the paper by Finkelstein and Fairley (4) was an important milestone—though that, in turn, had evolved to some extent from lines of reasoning followed by Mosteller and Wallace (5) in considering the authorship of The Federalist papers. However, recent research at the University of Lausanne (6) has pinpointed the work of Poincaré, Darboux, and Appell as, apparently, the first example of what we now call the Bayesian view of forensic evidence. It is particularly germane that the reasoning of Poincaré and his colleagues was concerned with a critical review of Bertillon's evidence in a notorious handwriting case: the trial of Dreyfus for treason. In modern parlance, we would say that Bertillon committed what Thompson and Schumann (7) called the "prosecutor's fallacy". Poincaré and his colleagues pointed out the error.

It is not necessary for me to explain the Bayesian view here because of the extensive body of literature that now exists in the forensic sphere. Useful introductions to the ideas are provided by Robertson and Vignaux (8) and Aitken and Stoney (9). The key principles that emerge from this view include, first, the notion that the forensic scientist should always consider (at least) two propositions that, in the adversary system of justice, will represent the defence and prosecution positions. Next, the fundamental principle is that the scientist must address questions of the kind "what is the probability of the evidence given the proposition?". Questions of the kind "what is the probability of the proposition given the evidence?" are the province of the jurors, who will not only take into

account the scientific evidence in their deliberations, but also all of the other evidence that is put before them.

In the case where there are no more than two competing propositions then the weight of evidence in favour of one or other of them is a function of the *likelihood ratio*. This is the ratio of the answers to two questions:

What is the probability of the evidence if the prosecution proposition is true?

What is the probability of the evidence if the defense proposition is true?

A likelihood ratio greater than one means that the prosecution proposition is supported whereas the defence proposition is supported when the likelihood ratio is less than one. This inspires the notion of a reporting convention based on the use of the word "supports" together with appropriate qualifiers that has been explained elsewhere—for example, Evett and Weir (10).

The consequence of this logical view of inference in the legal framework is that is is not appropriate for the scientist to frame his/her opinion in the form of a probability for the truth of a proposition. Interestingly, a similar view was reached through different methods and in a broader context by Popper (11) who said (p 394) "I regard the doctrine that the degree of corroboration or acceptability cannot be a probability as one of the most interesting findings of the philisophy of knowledge".

Yet this logically impermissible kind of probability statement underpins the ASTM standard, which embodies such expression as:

"There is *strong probability* that the John Doe of the known material wrote the questioned material, or it is my opinion...that the John Doe of the known material *very probably* wrote the questioned material."

The convention also sanctions the use of the word "unlikely" (as an equivalent to "improbable") within the context:

"It is unlikely that the John Doe of the known material wrote the questioned material."

This suffers the same logical fault as the former. These are probability statements about the truth of propositions.

It is my view, therefore, that if the ASQDE has the policy of promoting the view that forensic handwriting comparison is a science then it must change its reporting convention, because it cannot be logically justified.

References

1. Evett IW. Towards a uniform framework for reporting opinions in forensic science casework. Science and Justice 1998;38:198–202.
2. McAlexander TV, Beck J, Dick RM. The standardization of handwriting opinion terminology. J Forensic Sci 1991;36:311–9.
3. Anon. Standard terminology for expressing conclusions of forensic document. American Society for Testing and Materials, Designation E 1658–96. Reprinted from Annual Book of ASTM Standards Vol. 14.02.
4. Finkelstein MO, Fairley WB. A Bayesian approach to identification evidence. Harvard Law Review 1970;83:489–517.
5. Mosteller F, Wallace DL. Applied Bayesian and classical inference: the case of The Federalist papers. New York: Springer-Verlag, 1964.
6. Taroni F, Champod C, Margot PA. Forerunners of Bayesianism in early forensic science. Jurimetrics 1988;38:183–200.
7. Thompson WC, Schumann EL. Interpretation of statistical evidence in criminal trials—the prosecutor's fallacy and the defense attorney's fallacy. Law and Human Behavior 1987;11:167–87.
8. Robertson B, Vignaux GA. Interpreting evidence: evaluating forensic science in the courtroom. Chichester: Wiley, 1995.
9. Aitken CGG, Stoney DA. The use of statistics in forensic science. Chichester: Ellis Horwood, 1991.
10. Evett IW, Weir BS. Interpreting DNA evidence: statistical genetics for forensic scientists. Sunderland, MA: Sinauer, 1998.
11. Popper KR. The logic of scientific discovery. London: Hutchinson, 1959.

Dr. Ian W. Evett
Forensic Science Service
109 Lambeth Road
London SE1 7LP, England (UK)