

Error Terror in Forensic Science: When Spectroscopy Meets the Courts

A big question in forensic science today is, “How do we best report uncertainty?” The answer to which approach is “best” turns out to be surprisingly complex, for many reasons.

Glen P. Jackson

In the 1930s, a researcher at the National Bureau of Standards (the predecessor to the National Institute for Standards and Technology [NIST]) was so confident in the precision of his reported analytical result that he was apparently prepared to “eat the apparatus and drink the ammonia” if he was wrong (1). Many of us who rely on sophisticated spectroscopic instruments to make chemical measurements—such as to identify seized drugs—can probably relate to this sense of confidence, even if we aren’t quite willing to gamble with the same stakes. A big question in forensic science today is, “How do we best report uncertainty in casework?” The answer to which approach is “best” turns out to be surprisingly complex, and for many reasons. In some situations, a prosecuting attorney in a certain jurisdiction will require the crime lab to report a result in a certain way. In some disciplines, error rates are simply not known. In other cases, a particular technique might have quite well known error rates, but the community might be divided about how best to handle or discuss uncertainty.

The field of DNA analysis probably leads the way in error reporting, both because the error rates are well known and because there is a strong consensus on how to communicate the results: that is, likelihood ratios. At the other extreme, comparison techniques such as fingerprints, hair, and bite marks have experienced very bad press because of exaggerated claims by

analysts regarding the confidence of “matches” (2). All three of these comparison techniques have been involved in wrongful convictions and exonerations, but the resulting impact has been very different for each discipline. For example, the National Institute of Justice (NIJ) has funded a significant number of grants to better assess and report error rates in fingerprint comparisons, and the actual error rates are becoming better understood (3). Whereas fingerprint comparisons are now regaining credibility as a result of changes since the National Academy of Sciences (NAS) report, the same cannot be said for bite-mark comparisons and hair microscopy—the former of which is no longer admissible in certain jurisdictions (4). Perhaps additional research in these areas will enable a more scientific or rigorous defense of the claimed inclusion and exclusion rates, but only time will tell.

To complicate matters, recent research has shown that reporting difficult concepts like likelihood ratios can be totally misinterpreted by triers of fact. In one example, subjects acting as jurors were asked to rank the weight of evidence in cases involving likelihood ratios that weakly favored the prosecution’s argument (5). The subjects completely misinterpreted the weight of evidence and inferred that the evidence actually favored the defense’s argument. Studies like these reinforce a two-part problem in reporting error and uncertainty in forensic science results. The first

problem is knowing the error and the second problem is effectively communicating the error.

Although the problem of uncertainty in chemical measurements has long been a topic of intense scrutiny, most of the attention has focused on the uncertainty of quantitative measurements, not qualitative. As a result, most governing bodies have well-developed standards for measuring and reporting quantitative results—such as using the concept of expanded uncertainty (6)—but fewer guidelines exist for measuring or reporting the uncertainty of qualitative analytical results. For example, we have very good guidelines for how to derive a result such as “ephedrine concentration was found to be 20 ± 3 ppm (95% confidence interval),” but we have fewer guidelines as to how to assess or communicate the probability that the substance being quantified is in fact ephedrine and not another substance like pseudoephedrine. Of course, in reality we typically perform method validation studies before using a particular method, and such method validation typically establishes some level of true and false positives and negatives or simply the limits of detection in different complex matrixes. However, reassuring oneself and stakeholders through the acquisition of true negative and true positive results is quite mathematically different from a meaningful numerical measure of uncertainty. The bottom line is that unless we analyze a large number of

known negatives and positives—with true results—using a given spectroscopic method, it is difficult to mathematically justify having a high degree of confidence in a result.

In the last decade, the Innocence Project (7) and the 2009 NAS report (8) have both served to raise awareness about the current practices in certain forensic science disciplines. In response to the recommendations in the 2009 NAS report, congress funded the establishment of the Organization for Scientific Area Committees (OSACs) under the management of NIST—a well-respected and independent standards-setting organization. OSAC committees and subcommittees have now been formed in a variety of areas to help promote standards for how each forensic science discipline should operate and to help provide uniformity between disciplines, such as with nomenclature, education, and training. When the subject matter requires it, these subcommittees contain practitioners and academicians who specialize in the use of spectroscopic and other analytical techniques, and the various subcommittees are always in need of additional help from researchers who can help develop national standards in these areas. Consider this an open invitation to nominate yourself to serve on a subcommittee or task group in your area of expertise.

The first standard to be adopted and posted on the NIST OSAC Registry of Approved Standards (9) was ASTM E2329-14; “Standard Practice for the Identification of Seized Drugs”. This document specifies the minimum standards required of an analytical scheme before one can identify a scheduled drug. An example of an acceptable scheme would be the use of Fourier transform infrared (FT-IR) spectroscopy, gas chromatography–mass spectrometry (GC–MS), and color spot tests, all on the same sample. Immediately after posting the standard, however, NIST administrators issued a statement on the registry to raise concerns about some of the language in the approved standard. Specifically, NIST leadership were

concerned with the wording “an appropriate analytical scheme effectively results in no uncertainty in reported identifications,” because the wording could be construed to mean that there is actually no error, which of course is not the intent of the language. The standard is now under revision at ASTM, and although the community is likely to agree on some vaguely different wording in the revised standard, two issues will probably remain: First, the revised wording will not change the fact that we will not actually know the false positive and false negative rates any more than we do now; and second, the new language is not likely to affect the problem of poor communication with the jurors. These two problems could largely be overcome through federally funded research projects and assistance from social scientists, but that will take several years to achieve.

Regarding the actual error rate of seized drug analysis, analytical chemists are rarely taught to use sophisticated approaches like Bayesian networks during their deliberations. Given the growing trend in applying Bayes’ theorem and likelihood ratios to forensic interpretations (10), it may be helpful to consider how Bayes’ theorem can be applied to spectroscopic or drug chemistry casework. For example, what are the (posterior) odds that a baggie of white powder seized from a suspect actually contains cocaine versus another innocuous substance, given that the powder was found to contain cocaine by a seized drug analyst? Some useful background information in the case might be as follows: First, the baggie was seized during the process of a tape-recorded drug deal for cocaine, and second, the powder gave unambiguously positive results for cocaine using two color spot tests, FT-IR, and GC–MS, which meets the standards set forth in E2329-14. Determining the posterior odds requires taking the product of the prior odds and the likelihood ratio (11). The prior odds includes values such as the probability that a suspect caught in a drug deal for cocaine would happen to have a baggie of white powder con-

taining cocaine. It’s hard to conceive of a probable reason why a non-drug-dealing person would carry a baggie of white powder, or why a suspected drug dealer would broker a deal without having the contraband in his possession, which is to say that even in the absence of any chemical tests, the prior odds alone would indicate that the white baggie has a very high probability of containing cocaine.

The likelihood ratio of the testing result is at the heart of this discussion, and it’s value is determined by the analytical scheme and the frequency of incorrect determinations. The likelihood ratio takes into account the frequency of true negatives, true positives, false negatives, and false positives for the analytical scheme. Unfortunately, we do not have accurate nationwide statistics from which to derive a meaningful likelihood ratio for a given identification, and such a number would not be universally applicable, given the number of scheduled drugs and the diversity of analytical schemes that are permissible under ASTM E2329-14. That said, anyone who has experience performing FT-IR or GC–MS for seized drugs can appreciate the selectivity and sensitivity that complementary techniques would offer to an analytical scheme, and would probably be willing to gamble quite high stakes that a positive result is in fact positive. (Note: Maybe not offering to drink ammonia if he or she is wrong, but this same sense of confidence.) The general point here is that the product of two very large numbers (the prior odds and the likelihood ratio) is an extremely large posterior odds or probability that the white powder contains cocaine. No one should claim that the error rate of such a drug identification is actually zero: it’s obviously not. But the error rate is likely to be so small as to be not meaningfully different from zero, in terms of its likelihood of occurrence. Whereas the likelihood ratio approach has the advantage of avoiding the topic or error altogether, likelihood ratios on their own don’t really describe absolute probability of an event; they only describe the rela-

tive probability of two propositions.

Practitioners are faced with the unenviable task of educating triers of fact, and given the difficulty of this task, practitioners make a compelling case for using language like “practically zero error” or “effectively zero error” to communicate the uncertainty of an analytical scheme to a jury. However, the recent President’s Council of Advisors on Science and Technology (PCAST) report recommends avoiding such language and instead sticking with empirical evidence (12). That said, “unforeseen error” such as mislabeling or human error in a laboratory management system (LMS), is far more likely to lead to an erroneous result than the analytical scheme itself, and the wording in ASTM E2329-14 refers to the error in the scheme, not the error outside of the scheme. Furthermore, when one propagates or combines error in an analytical scheme, it’s quite normal to ignore the smallest sources of error when one or two sources of error are dominant. Mathematically, this is equivalent to saying that the sources of error with the smallest error are “effectively zero.”

One final thought regarding the NIST OSACs and it’s mission is this: the legal community is in dire need of help to understand which disciplines in forensic science are junk science and which disciplines are good science. The NIST OSACs are currently developing documents (standards) to inform practitioners within their respective disciplines how to perform certain techniques. What the legal community seems to need are documents describing how a particular technique meets admissibility criteria such as Daubert, Frye, or Federal Rules of Evidence. An excellent example of such a foundational document is “Stable Isotopes and the Courts” by Ehleringer and Matheson (13). Developing standards for how a crime lab should practice a technique is fundamentally very different from developing guidance or review articles to help lawyers understand the scientific rigor and appropriate use of a technique in a given case. Unless the

NIST OSACs work on developing informative review articles like the one described above, it’s hard to see how the standards currently being adopted will help the legal community at all. With this problem in mind, if you are familiar with review articles that help address admissibility standards for spectroscopic techniques, please bring these documents to the attention of the NIST OSAC subcommittee members in your domain.

The fact that forensic science disciplines are openly discussing their limitations, and are actively seeking improvements, makes now an exciting time to be involved in finding solutions, whether through basic research or through the drafting of standards. In addition to needing experts in forensic applications of spectroscopic and chromatographic methods of analysis, the NIST OSAC subcommittees also welcome statisticians and chemometricians to volunteer to serve. We have a golden opportunity to help make a significant and lasting impact on forensic science. More importantly, if we—as a discipline—can develop the tools to better measure, understand, and report the uncertainty in qualitative determinations, we have a unique opportunity to help inform and elevate every other domain of analytical chemistry. I’m confident analytical chemists and spectroscopists can make a difference in error reporting in forensic science, and frankly the practitioners can’t afford to wait around for help. It’s time to address the terror of error in qualitative determinations. If we don’t, we may as well eat our instruments now.

Disclaimer

As a disclaimer, I serve on the NIST OSAC Seized Drug Subcommittee that adopted the standard discussed in this article.

References

- (1) “Round-Table Discussion on Statement of Data and Errors,” *Nucl. Instrum. Methods* **112**, 391–395 (1973).
- (2) S.S. Hsu, “FBI Admits Flaws in Hair Analysis Over Decades,” *Washington Post*, April 18, 2015.

- (3) M.K. Taylor et al., “Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach” (NIST Internal Report 7842, 2012).
- (4) B. Grissam, “Texas Science Commission Is First in the U.S. to Recommend Moratorium on Bite Mark Evidence,” *The Dallas Morning News*, Feb 12, 2016.
- (5) K.A. Martire, R.I. Kemp, M. Sayle, and B.R. Newell, *Forens. Sci. Int.* **240**, 61–68 (2014).
- (6) ISO/IEC Guide 98-1:2009 “Guide to the Expression of Uncertainty in Measurement,” 2009.
- (7) <http://www.innocenceproject.org>.
- (8) Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council, “Strengthening Forensic Science in the United States: A Path Forward” (The National Academies Press, 2009).
- (9) <http://www.nist.gov/forensics/osac/osac-approved-standards.cfm>.
- (10) F. Taroni, A. Biedermann, S. Bozza, P. Garbolino, and C. Aitken, *Bayesian Networks for Probabilistic Inference and Decision Analysis in Forensic Science*, 2nd Ed., (Wiley, New York, New York, 2014).
- (11) T. Bayes and R. Price, *Phil. Trans. Royal Soc. London* **53**, 370–418 (1763).
- (12) PCAST Report on Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. Sept. 2016.
- (13) J.R. Ehleringer and S.M. Matheson, “Stable Isotopes and the Courts,” *Utah Law Rev.*, 385–442 (2010).



Glen P. Jackson

is the Ming Hsieh Distinguished Professor of Forensic and Investigative Science at West Virginia University.

Direct correspondence to:

Glen.Jackson@mail.wvu.edu ■

For more information on this topic, please visit our homepage at: www.spectroscopyonline.com