J Forensic Sci, July 2009, Vol. 54, No. 4
doi: 10.1111/j.1556-4029.2009.01081.x
Available online at: www.blackwell-synergy.com

Bruce Budowle,[1] Ph.D.; Maureen C. Bottrell,[1] M.S.; Stephen G. Bunch,[1] Ph.D.; Robert Fram,[1] M.A.; Diana Harrison,[1] B.S.; Stephen Meagher[1]; Cary T. Oien,[1] M.S.; Peter E. Peterson,[1] Ph.D.; Danielle P. Seiger,[1] M.F.S.; Michael B. Smith,[1] B.A.; Melissa A. Smrz,[1] M.S.; Greg L. Soltis,[1] M.S.; and Robert B. Stacey,[1] M.A.

A Perspective on Errors, Bias, and Interpretation in the Forensic Sciences and Direction for Continuing Advancement*

ABSTRACT: The forensic sciences are under review more so than ever before. Such review is necessary and healthy and should be a continuous process. It identifies areas for improvement in quality practices and services. The issues surrounding error, i.e., measurement error, human error, contextual bias, and confirmatory bias, and interpretation are discussed. Infrastructure is already in place to support reliability. However, more definition and clarity of terms and interpretation would facilitate communication and understanding. Material improvement across the disciplines should be sought through national programs in education and training, focused on science, the scientific method, statistics, and ethics. To provide direction for advancing the forensic sciences a list of recommendations ranging from further documentation to new research and validation to education and to accreditation is provided for consideration. The list is a starting point for discussion that could foster further thought and input in developing an overarching strategic plan for enhancing the forensic sciences.

KEYWORDS: forensic science, error, confirmation bias, context bias, interpretation, education, training, ethics

The application of science to characterize forensic evidence has revolutionized the investigation of crimes over the past 100 years. While not always individualizing, the forensic sciences can provide meaningful evidence for excluding or including a group of individuals or items as the source of evidence and thus provide another piece of the puzzle to help determine guilt or innocence. Many disciplines may have begun in an *ad hoc* manner; yet experiential inferences and foundational research have impacted the practices and helped build robust fields. In recent years, however, the forensic sciences have seen increased commentary about their practices and whether the analytical results presented in legal proceedings are reliable. This more intensive review is due to several factors, which include: (i) forensic scientists today are more sophisticated and are professionally questioning their disciplines to improve practices; (ii) scientists outside of the forensic disciplines have increased awareness of and interest in the forensic fields; (iii) the public awareness of the forensic sciences has heightened due to popular entertainment media; (iv) an aggressive adversarial legal system; (v) the Daubert and Kumho admissibility standards; and (vi) documented examples of errors that have been committed (1), especially those that have resulted in wrongful convictions (2).

There always is a need to make a process better by improving a method or developing a new method. Such improvements in themselves do not necessarily call into question the reliability of current or past methods. Critiques and suggestions abound on better methodologies. However, there seems to be little doubt that, for example, DNA can be extracted from evidence so that a profile can be generated and compared with a reference sample, or that a latent print can be detected by use of laser excitation and the features from the evidentiary print can be compared with a reference print, or that the elemental composition of a glass sample can be determined and the profile can be compared with other glass samples. Instead, many criticisms tend to focus on two areas: (i) errors, and (ii) the significance or weight of evidence.

We, therefore, focus the majority of this paper first on addressing the issues of error: measurement error, human error, contextual bias, and confirmatory bias. Error can be caused by measurement inaccuracies and human mistakes, and be due to contextual bias and confirmatory bias. Since it is impossible to completely eliminate the possibility of an error, understanding how errors can arise and employing a sound quality assurance (QA) program, that emphasizes peer-review, can minimize them.

Next we address the need to properly convey evidentiary weight. There has been discussion that all disciplines should consider using a "DNA-based" model of quantification of inculpatory evidence to assess significance of an observation (3). Some critics suggest that if a similar statistical model is not followed the result is meaningless or unreliable (3). However, very few methods can use "DNA-based" statistical approaches and employing such an approach may unintentionally overstate the weight of the evidence. A lack of a specific statistic does not mean a method is unreliable. Nonetheless, in lieu of a quantitative assessment, it is imperative that a qualitative statement is provided that appropriately conveys the significance of the match or association. Forensic methods typically identify relevant features, make comparisons, and exclude or fail to exclude. These actions rarely, from an analytical point, are called into question. The conveyed significance of a result using

[1]FBI Laboratory, 2501 Investigation Parkway, Quantico, VA 22135.

*This is publication number 08-13 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation.

Received 11 April 2008; and in revised form 10 Oct. 2008; accepted 25 Oct. 2008.

qualitative statements can be useful as long as the scientific assumptions and findings are well explained.

We should note that we purposely do not discuss in any detail admissibility standards such as Frye and Daubert. While these do have an impact on the forensic sciences, we believe using a legal standard for determining what is "good science" adds to the concept that "as long as it gets admitted everything is fine." If good science is carried out, then legal admissibility should be met readily under either standard of admissibility. Therefore, in this paper we stress on improving the sciences, not meeting an admissibility standard.

Lastly, we provide a list of recommendations for consideration as the disciplines of the forensic sciences continue to evolve. Hopefully, these recommendations will foster further thought and input in developing an overarching strategic plan for enhancing the forensic sciences.

To provide a starting point to better appreciate some of the practices of the forensic sciences, six disciplines—latent prints, glass comparisons, shoe print comparisons, handwriting analyses, morphological hair comparisons, and firearms and tool mark analyses—and their practices are briefly described in separate articles (4–9). It can be difficult to begin a discussion on the forensic sciences without a basic understanding of the processes. We strongly recommend that anyone interested in constructively critiquing a forensic science discipline become intimately familiar with the foundations and practices of that discipline. Otherwise misinformation or misunderstanding will likely prevail which would be of little benefit for improving the forensic sciences. These brief summaries (4–9) should inform and help guide the reader to additional sources for a better fundamental understanding of current practices.

Ascertainment Bias

The forensic field has a constraint not routinely encountered in other scientific disciplines, namely the law (10). Legal proceedings can play a role in the review of technology and its validity (10,11). Science and the law do not extract information and reach conclusions in the same manner. It is a tenet of science to continuously question one's beliefs and findings. Using the scientific method, a hypothesis is proposed, and experiments are carried out to test the hypothesis. If the data do not refute the hypothesis, the hypothesis gains more support, and through incremental steps, the hypothesis becomes grounded and accepted as reasonable and reliable. Constructive criticism is considered a healthy approach for building a better process.

In an adversarial legal system a defense attorney has a responsibility to attempt to create doubt, and is not always required to provide any data to support such doubt. In so doing, the attorney is performing his/her proper and appropriate role to provide a vigorous defense. Similarly, there are times where the prosecution may challenge the admissibility of a method that the defense seeks to proffer. In the courtroom, one may exploit the standard practice of science "to question" in order to suggest that there is a lack of consensus, even if most of the relevant scientific community agree that the approach is reliable. Moreover, the admissibility of scientific evidence is more likely to be challenged when the evidence does not support one side's position. Thus, one should be cautious of suggesting advocacy in lieu of objectivity. A finding that supports the theory that an individual committed the crime probably will be introduced by the prosecution, but is unlikely to be introduced by the defense. In most situations where forensic science analysis excludes an individual as a source of the evidence, the prosecution does not proceed with trying the individual. These cases usually are dismissed. Typically, inculpatory evidence is entered into evidence by the government in legal proceedings. So it is expected that most scientific evidence entered into a court proceeding by the prosecution supports its position. This circumstance in no way should be misconstrued as forensic scientist bias. It is a bias in ascertainment as to what is entered into court proceedings. Ascertainment bias is systematic distortion in the data collected or observed so that a true measure of the frequency of occurrence can not be made. The legal proceedings do not entertain most forensic science analyses where an exclusion was obtained. The vast majority of forensic scientists want to obtain the correct result whether it is exculpatory or inculpatory. The exculpatory evidence often is filtered out before court proceedings.

To underscore the differences inherent in the scientific and legal approaches, suppose that a microscopic hair comparison method has been unequivocally demonstrated to be reliable. The method is used to characterize some highly probative biological evidence from a crime scene and a reference sample from an accused individual. If the microscopic morphology features from the evidence and an individual are different, the appropriate interpretation is that the evidence sample could not have originated from that individual. Rarely in such a case should that individual be charged with the crime. More importantly, the reliability or admissibility of the evidence will not be challenged by the defense because it is exculpatory. In contrast, if the evidence and the known hair from an individual have the same or similar morphological features, then the data lend support to the prosecution's hypothesis and combined with other evidence, the accused may be charged. If there is a trial, the hair evidence may be used in court. It is in this latter scenario that challenges to the reliability of hair evidence tend to occur. The defense has the purview to vigorously challenge the admissibility of the evidence; yet, the burden of proof is on the proponent of the evidence. Indeed, scientific arguments in the courtroom are classic examples of bias in ascertainment; typically those analyses that fail to exclude the accused are the ones that are challenged.

There have always been challenges to the use of science in legal proceedings. In the adversarial system the evidence is criticized in a negative, nonconstructive manner. As a result, the courtroom can pervert the evaluation of science. The same analytical methods used in other fields are used in many forensic analyses and the basic foundations of the science are the same. However, questions regarding admissibility which would not be considered valid arguments in the relevant scientific community can arise in the courtroom. Thus, validity and reliability are better sought in the scientific arena where a positive constructive approach is used to improve processes. Science is continuously evolving. Testability in the scientific arena has been and is a better venue than the courtroom for determining the validity of forensic science methods.

Two points arise here that warrant further consideration. The first point is that for many years the forensic science community has pointed to successful admissibility of its science findings, and the opportunity to cross examine expert witnesses, as support of a technique's "general acceptance" and "reliability." Again philosophically we do not advocate successful admissibility as demonstrating good science. Instead, carrying out good science should result in successful admissibility. The forensic science community should advocate continued evaluation and testing of the scientific validity and reliability of methods used to characterize forensic evidence. The second point is that we believe that bias is not a serious pervasive concern (although it obviously exists, see below). The forensic science community should consider documenting the numbers of "inclusions," "exclusions," and "inconclusives" (or whatever terms are used for the comparison process) of their laboratory results.

Such data, if it were to be published, would likely support the proposition that forensic scientists are not overly biased and do provide substantial testing that can benefit either accused individuals or the government.

Quality Assurance in Forensic Science

Quality performance is an essential component for obtaining reliable results and for reducing the chance of error. QA provides the infrastructure to promote high performance, address errors that arise, and improve processes. Education (including continuing education) and training are essential components to maintaining a high quality operation. The ultimate goals of a high quality system are to minimize the occurrence of error and to develop and encourage an environment for improving processes and services. A well-developed QA program provides for quality products or services; conversely, not developing or following QA or quality control (QC) practices can result in poor quality or unreliable results. Adherence to using validated and documented protocols, tested reagents, calibrated equipment, appropriate control samples, recognized, detailed, and methodical documentation requirements, and independent review of operations, results, and interpretations are necessary for obtaining reliable results with confidence (see [12 (5.4.5 *Validation of methods*), 13–15]). If these principles and categories of a quality system are enacted, then the desired goals can be met.

Continuous improvement must be a cornerstone of the forensic sciences in their ongoing efforts to make products, services, or processes better. These efforts can seek incremental improvement over time or breakthrough improvement immediately. Among the most widely used tool for continuous improvement is a four-step quality model—the plan-do-check-act (PDCA), also known as the Deming Cycle or the Shewhart Cycle (16). The PDCA is (i) Plan—Identify an opportunity and plan for change; (ii) Do—Implement the change on a small scale; (iii) Check—Use data to analyze the results of the change and determine whether it made a positive difference; and (iv) Act—If the change was successful, implement it on a wider scale and continuously assess your results. If the change did not work, begin the cycle again. There are other approaches that are based on the same principles as that of the Deming cycle, such as six sigma which is synonymous with DMAIC (Define, Measure, Analyze, Improve, and Control) (17), that also could be considered.

High quality does require dedicated resources. However, poor-quality products and services are far more costly in operations, reliability, and credibility. Forensic laboratories must follow a robust system if they wish to maintain high quality performance. While no quality system can completely eliminate the potential for committing an error, adherence to a well-designed quality system will enable a laboratory to strive for constant improvement in their processes. In particular, a good QA system allows for identifying limitations, focusing on minimizing risk of error, and instituting methods of detecting error.

The forensic community has several accepted avenues to design its quality system. There is no one specified set of rules for the development of a forensic quality management system; several are acceptable. For example, the FBI Laboratory selected to use the requirements described in ISO/IEC 17025:2005, *General Requirements for the Competence of Testing and Calibration Laboratories* (12) and the ASCLD/LAB *International* supplemental document (18) for its quality management system. Other accrediting bodies for forensic laboratories use the International Laboratory Accreditation Cooperation (ILAC)-Guide 19:2002, *Guidelines for Forensic Science Laboratories* (19) as a supplemental document. For example, Forensic Quality Services has *Requirements for Accreditation*-1 (FRA-1) which are derived from ILAC Guide 19 (20).

Competent accreditation programs based on the International Standard of ISO/IEC 17025 (*General Requirements for the Competence of Testing and Calibration Laboratories*) (12) supplemented with appropriate forensic requirements (18) already provide forensic laboratories with the requirements to develop a robust and comprehensive quality system. Many forensic science laboratories are taking the necessary steps to regulate themselves by employment of QA programs that are appropriate to their operations and stringent. The laboratories also can demonstrate that the requirements are being carried out by participating in a rigorous external accreditation process. The external accreditation program is designed to reveal areas for remediation and improvement from an outside, unbiased viewpoint. To date, forensic laboratory accreditation has been voluntary, except in the states of New York, Texas, and Oklahoma (21–23). While many federal, state, and local forensic laboratories have developed quality systems and are actively participating in a forensic accreditation program, there are still too many forensic science facilities that function in various law enforcement agencies and private facilities that are providing services and are not accredited. We strongly advocate mandatory accreditation for any entity which provides forensic services.

The requirements of ISO/IEC 17025 (12) are divided into a number of categories. Some of the requirements (derived and modified from ISO 17025) are listed in Table 1. This list is an overview of some of the major requirements of ISO/IEC 17025. When all the sub-clauses are examined there are over 400 separate requirements to which a laboratory must conform in order to be accredited by one of the appropriate accrediting bodies. One of the current accreditation programs has developed additional requirements that are specific to forensic science laboratories (18). In total, the described elements and criteria serve as a road map for a laboratory to establish a comprehensive and robust quality system. By addressing these criteria, a laboratory system can put in place a high quality infrastructure.

As mentioned previously, the adversarial system is not a constructive process for addressing error, although it can be effective in raising awareness of error. If an error does occur (after all, humans are involved), the adversarial system tends to focus on the error itself. Of course that error is germane to a particular case, but one should not concentrate solely on the error itself, but also what caused the error, that it was corrected, and how it was corrected. It is not so much that an error occurred that is important, but what was done about the error that is the primary concern. In QA, this involves a root cause analysis, i.e., a thorough analysis of all the potential causes of the problem (see [12, 4.10.2 *Cause analysis*]). The opportunity to improve the individual and the infrastructure, when necessary, should be embraced, so such an error and any other mitigating risks (that were identified by a thorough review) are ameliorated. When errors are ignored, there should be great concern raised. QA programs are essential to define the process and documentation for addressing an error. Nationally mandated accreditation programs assure that all practicing laboratories are meeting acceptable practices for monitoring and improvement of people and processes. National programs also foster communication so that the discovery of an error in one laboratory can bring a benefit to other laboratories within the accredited network.

Error and Error Rates

Even with safeguards in place, errors can occur in any endeavor involving humans. The errors we address here are not those due to

TABLE 1—*Some of the QA requirements derived and modified from ISO 17025.*

Organization—a laboratory must be organized in a manner that promotes quality work and that meets the needs of the customer
Management system—management must create policies, practices, and an environment that promote quality work
Document control—the laboratory must establish procedures for the control of all the documents that make up its quality system. These would include such documents as the Quality Manual, Operations Manuals, Procedure Manuals, etc.
Review of requests, tenders, and contracts—the laboratory must have procedures to review requests for work so as to ensure that both the laboratory and the ''customer'' understand the requirements of the work and that the laboratory has the capability to perform the work
Service to the customer—the laboratory must seek customer feedback, both positive and negative, that will help the laboratory improve the quality of its work and better meet the needs of the customer
Complaints—the laboratory must have a policy and a procedure for the resolution of complaints received from its customers or from other parties
Control of nonconforming testing—the laboratory must have a policy and a procedure to deal with testing that does not conform to its own procedures. When appropriate, corrective action will be implemented
Improvement—the laboratory shall work to continually improve the effectiveness of its quality system through a variety of quality assurance activities, such as the use of audit results, corrective and preventive actions, management reviews, etc.
Corrective action—a laboratory must have a policy and a procedure and designate appropriate authorities for implementing corrective action when problems are identified. The procedure will include a cause analysis, the selection and implementation of the corrective actions, the monitoring of the corrective actions, and the need for additional audits
Preventive action—procedures must be in place to direct preventive action. When improvement opportunities are identified by the laboratory, action plans will be developed and implemented to take advantage of the opportunities for improvement. Preventative action should be a pro-active process
Control of records—the laboratory must develop procedures concerning records. Among other requirements, they must be held secure and in confidence. Back-up and access procedures must be in place to protect electronic records
Internal audits—the laboratory shall have a schedule and procedure to conduct audits. Audits should cover all the activities of the laboratory to ensure that its activities comply with the requirements of the management (quality) system
Management reviews—the laboratory's top management is required to review the laboratory's management (quality) system and testing activities to ensure suitability and effectiveness. ISO 17025 directs a number of topics that must be covered during the management review
Personnel—the management of the laboratory is required to ensure the competence of the individuals who perform the examinations to include proficiency testing. Among other requirements, the laboratory is required to have a training program that is relevant to the work that will be performed
Accommodation and environmental conditions—a laboratory must provide an environment that will facilitate the correct performance of the testing. Included in the requirement is the need for effective separation where incompatible activities occur so as to prevent cross-contamination
Test and calibration methods and method validation—a laboratory must use appropriate methods and procedures so as to ensure correct results in its testing. Methods that have been appropriately validated and published are recommended for use. The requirement also provides guidance for the appropriate validation of new methods
Equipment—a laboratory must have the proper equipment so as to accomplish the correct performance of the examinations. The requirement also outlines who will operate the equipment and how it shall be maintained
Measurement traceability—a laboratory performing measurement must be able to demonstrate traceability to an appropriate standard. Reference materials, where possible, must be traceable to a certified reference material
Sampling—when a laboratory uses a part of a substance or material for testing as a representative sample of the whole, then a sampling plan must be developed and used for the selection of the testing sample
Handling of test and calibration items—procedures must be established for the transportation, receipt, handling, protection, storage, retention, and disposal of test items. These requirements represent appropriate evidence handling that should be in place in all forensic laboratories
Assuring the quality of test and calibration results—a laboratory must have procedures for checking the validity of tests. These include, when appropriate, the regular use of certified referenced materials, proficiency testing programs, and repeat testing
Reporting the results—a laboratory must report the results of all tests. It is not permissible to allow results to be unreported. Additionally, the requirement provides that information must be included in the report that makes clear what was tested, the results, who the test was performed for, and who conducted the test. Also, opinions and interpretation must be clearly marked in the report

methodology; these can be calculated and defined for most analytical processes. However, with the more subjective component comparison analyses, human limitations can result in errors and/or a lack of consistency among practitioners. These vulnerabilities need to be addressed prophylactically (preferably) or when a concern arises. It is important to identify potential error and focus on those areas where error is most likely to occur. The most critical error would be a false association (but one could argue that equally critical would be a false exclusion of evidence that would have failed to associate a victim with a crucial piece of evidence; or from the victim's standpoint a false exclusion of a suspect would also be critical). Obviously, it is important to know if an error has occurred in a case analysis that results in a false match or inclusion, a wrongful exclusion, or that causes the conclusion to be overstated (Note: understating the weight of the evidence is typically not criticized, because with a degree of uncertainty for assessing the weight of the evidence, it is desirable by the forensic science community to be conservative).

Admittedly human errors do occur (1). But, presenting the possibility of error as a practitioner error rate is not advocated. An error rate in the context of a scientific discussion is defined as a continuous, repeatable, consistent action that yields a predictable level of false positive or false negative results in casework. Providing error rates, along with or in combination with the association, has been proffered as a meaningful way to convey the strength of the evidence (24–27). However, suggesting that a specific error rate must be presented adds little value to the discussion on reliability. A community-wide error rate is not meaningful, because it falsely reduces the rate of error for those who might commit the most errors and wrongly increases the rate for those who are the most proficient. Moreover, when an error of consequence occurs, for instance a false inclusion caused by human error, QA demands that corrective action be taken which includes review of cases analyzed by the examiner prior to and after discovery of the error. Because the corrective action taken must be such that the individual will no longer commit that error, or at least make it less likely to do so, that past error can no longer impact negatively to the same degree on the individual's future performance. In fact, the practitioner likely is better educated and thus less prone to err in the same manner. Thus, the calculation of the individual's current error rate should not include that past error. Cumulative error over time is not a meaningful mechanism for assessing current error rate. We do not suggest that the error(s) should be ignored in the laboratory or in court. Instead it is more meaningful to convey error qualitatively; i.e., what was the error, what caused the error, what is the consequence of the error, and what was or is being done about the error. One also should not presume that the absence of prior errors precludes an individual from making an error in a particular case. Attorneys, if they believe it is useful, should make use of such information during examination and/or cross-examination of an expert witness.

In legal proceedings, some may want to focus on diminishing the weight of evidence based on a hypothetical error rate that does not necessarily apply to the case at hand. As an example, Saks and Koehler (3) declared that "the practical value of any particular technology is limited by the extent to which *potentially* important errors arise" (italics added). This statement means that the fact that an error is possible necessarily lessens the value of the evidence, even if an error has not occurred in the specific case. A known error rate (wrongly calculated from a proficiency test mistake [3,24]) is at best some indirect measure of the verity of the proposed results in any given case, but can never be a direct measure of the reliability of the specific result(s) in question (28). Many forensic disciplines have nonconsumptive forms of examination. The most direct way to measure the reliability of the purported results is to have another qualified expert conduct his/her own review, as is advocated by the National Research Council for DNA analyses (25); DNA analysis does not present an error rate (which should not be confused herein with the probability of a coincidental match). Reanalysis by a qualified examiner would be more meaningful and less costly than entertaining experts espousing hypothetical errors and error rates (28).

One can point to proficiency tests to indicate that errors do occur (as an example see [29]), and these should be evaluated for improving the sciences and improving on weak points in a system (25,30). We purposely do not address such proficiency test results herein because of a number of potential issues that need to be considered, which are not available to us for this paper. These include: what was the purpose of proficiency testing? what was the format and construction of the tests? were the tests properly designed? were all examiners who took these tests "qualified" examiners with sufficient knowledge skills and ability? was the purpose of the tests well understood? were the interpretation guidelines of the laboratory followed and was variation in the degree of conservatism considered a difference? did the examiner ignore in-house guidelines and attempted to provide a result? was what constitutes an error well defined? were the differences due to classifying an "inconclusive" versus a "conclusive" interpretation? were the differences "exclusion" versus "inclusion"? etc. Others may want to weigh in on these issues in a future paper(s).

Some descriptions of error rates are misunderstood or misleading at best. For example, Saks and Koehler (3) stated that the false positive error rate for microscopic hair comparison is 12% based on a study of morphological hair comparisons and mitochondrial DNA (mtDNA) analysis in a study by Houck and Budowle (31). Saks and Koehler espoused that 12% of hair matches are in error because the *higher resolving* mtDNA test excluded some hair comparisons that were categorized as "a failure to exclude." The Houck and Budowle study contains no data on false positive errors. Instead, it is a comparative study of the different resolving capacities of the methods. Consider a case where a blood grouping analysis was performed and assume that the test is valid and reliable. Also consider that the evidence has been run independently in several laboratories, and all results are consistent. The blood group type for the evidence sample is type A and the reference sample from the suspect is also type A. Based on population genetic statistics, about 60% of the population can be excluded as a source of the evidence sample. The result is reliable, consistent, reproducible, and valid; it just is not particularly informative. The same evidence subsequently is analyzed with today's DNA technology, and the two samples are found to have different genetic types. Assuming no sample mix-up, the suspect could not be the source of the evidence. If the philosophy of Saks and Koehler were followed, the blood group result of type A and the failure to exclude would be called an error in typing. However, it should be obvious that this is not the case—the methods of blood grouping and DNA typing simply differ in their resolving capacity. This blood group and DNA typing example is no different than a microscopic hair examination and mtDNA sequencing. In some scenarios mtDNA can be more resolving than hair microscopy such as hairs from unrelated individuals, and in others microscopy can be more resolving than mtDNA, such as with hairs from maternal relatives. This difference in resolving power should not be misconstrued for establishing an error rate, if a meaningful one could ever be calculated.

However, the difference in resolving power between microscopic and mtDNA analyses should not be ignored. Other than for maternal relatives, where microscopic analysis is more resolving than mtDNA sequencing, studies such as that by Houck and Budowle (31) can indicate the resolving power/limitations of microscopic analyses when hair from unrelated individuals is compared. The examiners in any forensic discipline that have information such as these should be prepared to convey such limitations so that a nonscientist will be less likely to give more weight to the evidence than should be given. Given the nature of the forensic sciences, it is critical that the forensic scientist is vigilant and always sensitive to the possibility that laboratory results may be misinterpreted by those involved in the investigation or legal process.

Subjectivity and Bias

A criticism, not unique to the forensic sciences, that is raised consistently is that subjectivity can affect reliability (32). Interpretations of scientific analyses have some degree of subjectivity (33). Some might suggest that subjectivity will impact the reliability of a result. However, subjectivity does not necessarily equate to unreliability (11,33). For example, a laboratory could subjectively place a higher threshold than needed for effecting an "association" to reduce the chance of false matches and consequently increase the number of inconclusive interpretations. While subjective, the practice of placing a conservative, higher threshold on an interpretation is not detrimental to an accused individual.

The suggestion is that a scientist's subjectivity, and thus biases, promote him or her to intentionally or unintentionally "fudge" interpretations towards "matches" (28). Yet, in our experience and although anecdotal, the outcome of interpretations (i.e., exclusions) would seem to refute this as a general practice. The forensic science community should make a concerted effort to document the percentage of exclusions to better demonstrate that subjectivity does not necessarily equate to "fudging."

As stated above, there is no *a priori* basis to conclude that subjectivity automatically correlates with unreliability. In fact, some protocols may allow interpretation in specific circumstances (usually those with limited quality or quantity) if the results support an exclusion of a suspect (or item associated with a suspect), but would not allow any inferences for inclusion (34,35). Such a practice is somewhat subjective but is intentionally designed to avoid a false inclusion.

Yet, biases exist in all of us (36). These biases can impact negatively on our judgments if they are not recognized and mechanisms are not in place to minimize wrongful interpretations (i.e., that can result in errors). The two types of biases that are of most concern for any scientist are confirmation bias and contextual bias.

Confirmation bias is a proclivity to search for or interpret additional information to confirm beliefs and to steer clear of information that may disagree with those prior beliefs (32). Contextual bias (some have used the term context effect) is using existing information or consistency to reinforce a position (37–39). In other words, contextual bias is where the forensic scientist uses other evidence to believe that the specific evidence being analyzed is related to a particular reference sample(s). These biases result from the natural tendency of putting things in a perspective to foster communication and to organize thinking. They are a part of the human psyche which enables us to pursue objectives, to categorize and classify things in context, and to communicate. Information is needed to effect a meaningful and reliable interpretation. We cannot deny that these biases exist; they are necessary for human beings to function (32). Unfortunately, they also may cause a loss of objectivity. Therefore, personal biases might override sound judgment, may affect interpretations in certain circumstances (37), and need to be minimized.

Some critics have suggested blind analysis is a possible way to alleviate the effects of contextual and confirmation biases (26,27,39). However, no one would want a doctor to make a differential diagnosis or advise a patient on a course of treatment without taking a medical history and full knowledge of the symptoms; that would be a formula for disaster. A hair or shoeprint examination carried out without knowing the estimated time between the crime and collection of reference samples could lead to erroneous interpretations. Ignoring elimination samples, when interpreting analytical results from evidence in a rape case, can provide false leads and reduce the power of the analysis. Complete ignorance to case specific information exhibits poor judgment and should not be considered. The difficulty is in determining what relevant information to request and what is superfluous. Some could argue that an examiner should not have knowledge that an eyewitness placed the suspect at the scene, or that the suspect confessed, or that a completely different item of physical evidence associated the suspect with the scene or victim. Such information could influence an examiner's interpretations and conclusions, but in some cases may be useful in making decisions on what samples are the most meaningful to analyze (particularly in cases with an inordinately large number of evidentiary samples). Some blinding of this ancillary information may have merit and should be considered by the forensic community. On the other end of the spectrum, recently, a letter describing sequential unmasking approach has been proffered for DNA interpretation (40). This letter has some points that are difficult to reconcile, such as a case manager solely deciding what to test, how to test, and to supervise testing. This suggestion would strip the laboratory of a wealth of experience in carrying out an analysis and would rely on only one individual to effect case analyses.

Because of the myriad case scenarios, there is no absolute or obvious guide available. There even may be situations where ancillary information could be invaluable. However, caution should be exercised when using additional information in casework, so that bias is not introduced that may result in error in interpretation (41). The best way to overcome and prevent potential biases in judgment is through peer review. Blind verification is a form of internal peer review that can reduce the chance of error and is complementary to the external review that is inherent in the adversarial legal system. Blind verification is defined as an independent second examination of an item(s) of evidence by another qualified examiner, who does not know the conclusion of the original examiner. Withholding the interpretation of the first examiner from a second independent examiner can decrease the effects of bias. The protocol should ensure that the blind verification process includes both associations and nonassociations. Thus, the second reviewer is not aware in any way even of the general outcome derived from the first examiner.

Because of the QA systems in place, biases are not routinely problematic. However, the forensic field should continue to undertake mechanisms to mitigate the effects of bias whether unintended or intended. Education and training should stress that subjectivity and bias exist in all human beings and these can impact judgment (33). The forensic scientist needs to recognize that this is not necessarily a negative thing; it gives human beings direction, focus, and motivation. However, if bias overrides sound judgment and the consideration of alternative explanations, then it could be detrimental.

Since confirmation and contextual biases are inherent in the psyche of human beings, science advocates independent confirmation and peer review to overcome these potential weaknesses. We suggest that the same mechanisms are in place in forensic science and do effectively reduce confirmation and context bias. Investing in the requirements of QA and following the tenets of science of independent review and peer review is the best way to reduce the effects of bias. Blind verification, independent review of an analysis, and/or retesting (particularly for nonconsumptive analytical techniques) are the best and most cost effective methods to reduce error due to bias (25,28).

Qualitative Versus Quantitative Assessments of Results from Forensic Examinations

There has been discussion about applying the DNA quantitative statistical model to other disciplines such as handwriting, latent print examinations, shoe print comparisons, trace evidence analysis, and tool mark comparisons to assess the weight of the evidence (3). Following the DNA quantitative statistical model may be foolhardy for these other disciplines. Statistical models need to be analysis-specific in order to be meaningful and contextually useful. While statistical models may not be in place for some analyses (for a number of reasons), statistical techniques have other value as well in that they may help determine what questions should be asked and what answers can be obtained given the available data, as well as can be used to determine what characteristics are dependent and independent with regard to one another. While quantitative statistical methods are used in many analyses, research should be carried out to determine if reliable approaches can be developed for assessing the significance of variation for the particular assay and what other value statistical assessments can provide. However, we strongly recommend that following the "DNA business" model could be very effective for other disciplines. The term "DNA business" model refers to a forensic science discipline system that has a strong QA program, which is nationally mandated, and standards of operation to ensure high performance.

There have been efforts to suggest that the DNA statistics model should be applied to all forensic science disciplines for assessing the significance of an evidentiary result where there has been a failure to exclude the evidence and a reference sample(s) as possibly originating from the same source. If one were to apply the population genetics statistical approaches used for evaluating DNA evidence to other types of evidence, unintended error could result. Population dynamics are very different for other types of evidence, particularly such items as shoeprints and trace materials, since their populations change upon manufacturing requirements and public demand. Many features that define items are not randomly distributed and do not even approximate random distribution. Therefore, it is often difficult to quantify evidence by employing the same principles used for DNA analyses, such as the random match probability, at least with the current state-of-the-art and understanding. Yet, we do support that further research be considered for assessing whether statistical methods can be developed for quantifying the significance or weight of a failure to exclude an evidence and reference sample as possibly being from the same source (or batch or lot).

Some have espoused specifically that microscopic hair comparisons alone (i.e., where no mtDNA sequencing results were obtained) should be deemed inconclusive because an association is not accompanied with a statistical assessment of the weight of the observation (42). Such a recommendation may have unintended consequences for those that see forensic methodology as solely a prosecution tool when in reality evidentiary results can serve to include or exclude a person. Consider two suspects and hair evidence: Suspect 1 is excluded but the hair comparison fails to exclude suspect 2. No mtDNA result was obtained. Suspect 1 may want to put forth a strategy to convey that suspect 2 could be the source. This strategy for Suspect 1's defense might be hampered if the result of a microscopic hair comparison was excluded because it is not statistically quantifiable. Suspect 2 will want to question the reliability of the microscopic hair comparison methodology. As long as done properly and given appropriate significance qualitatively, comparison evidence also can be quite valuable for excluding potential sources.

Moreover, a quantitative assessment of association (43) may not be necessary in a number of scenarios. In certain cases, the frequency of occurrence of a hair in a population may not be necessary for a case. For example, consider a scenario where hairs are recovered from the inside of a windshield from a car involved in an accident. Neither of the two people in the car admits to being the driver. The entire "population" in question in this case is the two people in the car. It only is necessary to resolve these two people as the source of the hair.

Of course, the real issue is not that the evidence is helping one side or the other; we indicate this only to make a point that the tool is not necessarily solely a prosecution tool. The most important issue is the reliability of the criteria used to make determinations. The conclusions excluding a suspect should be based on the same reliable criteria as that for inclusions. Thus, full consideration of discrimination power of relevant characteristics could be relevant. But the hair evidence in the automobile crash, which is a closed population analysis, may not require as much use or specificity of the discrimination power of microscopic analysis to infer the weight of the evidence as would be needed in an open population analysis.

Regardless, it would be comforting to some to be able to quantify evidence in a manner that would emulate that of DNA typing, but such is not possible. So what is to be done? First and foremost, a lack of a quantitative model does not mean that the method is unreliable and of no value. The identification and comparison of features and the resulting interpretation of exclusion, failure to exclude, or inconclusive can be made without quantification. Yet a lay person or even another scientist not experienced in the specific discipline may not appreciate the significance of the observation. Therefore, in lieu of a quantitative approach, it is imperative that the weight of the evidence be explained qualitatively so that fact finders or other scientists can appreciate the limitations of the analysis and comparison. We strongly recommend that each discipline document and provide to the legal and greater scientific communities the limitations associated with qualitative interpretations, the features used to effect an interpretation, and relative rarity or commonality of those features. The information will be useful for communication and would assist the prosecution and defense in mounting support or criticism as necessary. This activity should be a primary mission for all current Scientific Working Groups (SWGs).

How Should the Community Proceed?

Need for Specific Definitions Regarding Qualitative Associations in Case Reports

An immediate and compelling need is to provide more definition or supporting data in case reports for those disciplines that offer qualitative statements regarding an interpretation of a result or comparison. Clearly, a number of comparison-based disciplines do not lend themselves readily to a quantitative approach such as that applied in forensic DNA analyses. This does not mean that we do not support efforts to develop appropriate statistical models to quantify the evidence results; such endeavors should be pursued. But, in lieu of a quantitative approach, there is additional information that can help convey the significance of an association, identification, or failure to exclude (for a summary see [4–9]) and references within these citations). Such additional information might be provided during courtroom presentation to assist the trier of fact. But, most cases never reach the courtroom because they are plea-bargained or the additional information may not be elicited during testimony. Merely giving a qualitative statement in a report that states that the interpretation is an association is insufficient to help investigators, the prosecution or the defense, appreciate the significance of a comparison. Therefore, we recommend that the discipline specific SWGs with all due speed define what additional information should be placed in reports so that the significance of qualitative statements can be better conveyed to all parties. In addition to information regarding the strength of the interpretation (for an example see [44]), appropriate supporting information in the report should encompass the analyses, comparisons, associations, conclusions and other interpretations drawn from the data generated or other information gathered during a forensic evidence examination (12,45).

Although semantics are a difficult topic to address, when terms such as "association" or "match" are used in a qualitative statement they may convey to some people stronger significance than other terms such as "failure to exclude." The same could hold true for terms like "dissociation." Others may find the current terminology reasonable and acceptable. Because there may be an unintended contribution to bias (i.e., conveying more strength than intended), existing terminology should be reviewed for best practices in report writing. Regardless, as suggested above, when such terms are used they should be fully described in the case report so that the meaning of terms, such as "association" or "match," are understood in context. An alternate approach is to use instead the term "failure to exclude," which may seem to some more acceptable. Though even with "failure to exclude," a description of the meaning of the findings of the comparison is warranted. Even the term "inconclusive" merits definition. The former suggestion is preferable because no matter what term is used, it is likely that someone will prefer an alternate term. We strongly urge the various SWGs and laboratory systems to better define these qualitative terms when an evidence sample and reference sample cannot be excluded as originating from the same source.

Need for National Forensic Science Education and Training Requirements

Currently, forensic science education and training are carried out through several mechanisms, such as academia, web-based training, short courses, and in-house crime laboratory training. There are education and training requirements for individual law enforcement laboratories; but with the exception of DNA there are no established mandatory acceptable minimal standards across the U.S. for the respective disciplines. While some programs may be exemplary, there is no guarantee that all forensic scientists are receiving the same basic fundamentals. Thus, performance quality may vary substantially. National standards need to be instituted; these are not just infrastructure related but also should address specific topics and details requisite for education and training.

We strongly advocate that all forensic scientists, i.e., those practicing particularly a criminalistics forensic science discipline, have a minimum of a bachelor's degree in a natural science or a similar degree that might be more appropriate for a particular discipline from an accredited college or university (46,47). A forensic scientist must have a basic education steeped in science to carry out effectively the analytical processes of a forensic discipline. Primarily, forensic scientists need to understand and practice scientific principles such as hypothesis building, problem solving, considering alternate explanations, documenting work and activities in a methodical manner, and laboratory work (32,46). Moreover, all forensic scientists should have coursework in statistics. Statistics allows one to frame experiments in a cogent fashion, to better evaluate and interpret data, and consider alternate explanations when appropriate. A basic science education that includes statistical training can be achieved through the current well-developed academic infrastructure, although the specifics of a statistic course(s) need to be defined. It is important that a forensic scientist be educated in the fundamentals of statistics, but specific approaches or concepts should be developed that are forensically relevant to include modeling, likelihood ratios, match probabilities, variation of analytical parameters, individuality, subjectivity, objectivity, uncertainty, and frequentist and Bayesian approaches (for examples see [48–50]). A forensic science statistics course should be developed that incorporates basics and forensic applications and that could be introduced into required curricula. A basic science education that includes statistical training might be achieved through the current well-developed academic infrastructure.

The National Institute of Justice's Technical Working Group on Education (TWGED) and the American Academy of Forensic Sciences' Forensic Education Programs Accreditation Commission (FEPAC) have developed minimum curricula expectations guidelines and accreditation criteria, respectively, for forensic science curricula (46,51,52). These entities have taken a major step in establishing formalized coursework requirements for forensic science programs at the undergraduate and graduate levels. They define and support the basic infrastructure for university-level forensic science programs and coursework. However, details regarding the minimum required information to be conveyed in the coursework, statistics, the resources required to educate a student, and the required expertise to instruct have yet to be addressed adequately. The specific requirements of coursework need to be defined to ensure that all students receive the same fundamental education so they can be instructed in accordance with nationally defined standards. The result would be better prepared students and a more qualified talent pool from which crime laboratories could select new hires.

It can be expected that basic science education has sufficient development at the accredited college and university level. However, due to limited resources and limited operational experience at the academic level, discipline specific training typically is carried out within the crime laboratory (although a very few forensic science academic programs do support well the training function). Training is a formalized, structured process intended to provide a level of scientific knowledge and expertise required to conduct forensic analyses germane to a discipline (46). Most training to develop the competency and proficiency of new scientists is done within the operational laboratory. Albeit time and resource demanding, the forensic laboratory should have incentive, expertise, and the equipment in place to carry out the training function. We believe that there are some very good training programs within crime laboratories, but expertise and resources vary substantially across the U.S. This variation is exacerbated in some organizations/agencies where the typical 6 months to 3 years of residency in a structured training program is not required. Some may consider a 1–2 week course sufficient; that practice should not be considered acceptable for any forensic discipline. Even when guidelines for training exist, they may not be widely followed. Standards and/or standardization of training curricula are not in place on a national level, and there is no authority to ensure that all agencies adhere to standard training curricula.

Training also may be a task assigned to a qualified examiner who may be carrying a full workload, may or may not have the requisite in-depth knowledge to train in the theories and fundamentals required to become proficient, and/or may or may not have the ability to teach. Because of this, there is no consistency in essential training requirements; there is no guarantee that all new scientists (or for that matter current scientists) are trained to the same minimum level nationwide. Given that the experiential based disciplines are taught by example, there is no guarantee that all scientists (or the vast majority) agree or recognize what is a sufficient amount of information to determine that the examinations are complete and appropriate and to render an association or identification. This disparity needs to be addressed. Training curricula and defined coursework need to be formalized and applied nationwide. Documented training programs should be developed to train an individual to have the requisite knowledge of the theories, procedures, and analytical techniques necessary to produce reliable results and conclusions, and adequately present evidence.

There should be minimum requirements for practical laboratory exercises and demonstrated competency through national testing programs and practical examinations (53). These requirements would apply to the more prevalent discipline practices. Some subspecialties, where there are only a few individuals carrying out the analyses, should at least be required to demonstrate competency. A standard training curriculum and required competency testing would be beneficial for establishing qualified scientists while maintaining consistency throughout a discipline. The training should be structured, measurable, and documented (46). National training programs would be a good way to ensure consistency and proficiency and should be considered. However, such programs may not be practical for access to all or may be too resource demanding currently. This could be achieved, in part, by the development and use of a web-based training database for the initial training of scientists who are selected to be trainers. The forensic training in the crime laboratories should be given by practitioners who are trained to teach. Laboratory trainers in each discipline should be educated in training requirements, how to evaluate training performance and teaching methods.

Continuing Education Requirements

Science is an ever-evolving process. New technologies and techniques are constantly being developed. Novel and better ways to interpret data will arise. Improvements in providing services continue. Continuing education in discipline-specific areas must be provided in order for forensic scientists to remain current in their areas of specialization and/or gain a higher level of expertise (46). Topics could include best practices and guidelines concerning the collection, analysis, and interpretation of forensic evidence to help ensure quality and consistency in the use of forensic technologies and techniques to analyze evidence. Also, continuing education should capitalize on disseminating information on the use of current and novel technologies and other research in the forensic sciences.

Many disciplines and their respective SWGs either require or recommend a prescribed number of hours of continuing education per year and professional assessment (15,53). Each scientist should participate in or attend scientific meetings, take courses, or participate in research. Obtaining continuing education is a costly but necessary burden for the crime laboratory and nontraditional learning opportunities should also be considered, such as seminars and web-based workshops or meetings. One area of continuing education that has been underemphasized and is invaluable is review of current, relevant scientific literature. There are many mechanisms to pursue continuing education and forensic laboratory management must support such efforts and provide sufficient time and resources to ensure forensic scientists maintain their expertise and keep abreast of the state of their disciplines. Forensic scientists need to be aware of and well versed in new and potential developments that could improve their current practices so the field can be sufficiently dynamic for meeting the challenges of the 21st century.

Training in Professional Conduct and Ethics

Professional conduct cannot be over-emphasized. Public service is a public trust, requiring employees to place loyalty to the Constitution (at least in the U.S. Federal government), the laws, and ethical principles above private gain. The public places its trust and confidence in law enforcement and in all of us involved in forensic science examinations. Therefore, the forensic scientist should carry out the highest level of ethics and professional standards. A Code of Ethics likely exists in many of the professional forensic organizations and various agencies (54). Some are agency specific and some are science specific. All crime laboratories, both public and private, should have a code of ethics/conduct. These codes need to stress the best interests of society through government service and justice, government responsibility for professionalism and cost effectiveness, integrity (both professionally and regarding evidence), objectivity, staying within the bounds or limits of what the science can provide, maintaining confidentiality, complying to legal demands such as disclosure, and being truthful (11,54,55).

There are a number of forensic science organizations that have codes of ethics/conduct; these should be evaluated and modified accordingly for instituting professional codes that are relevant. Another part of ethics and responsibilities are those relevant to management and its responsibilities to its employees. For example, management has an obligation to provide an environment with resources and training for its forensic scientists to acquire and maintain competency. One possible set of ethics guidelines as a starting point for management is "ASCLD Guidelines for Forensic Laboratory Management Practices" (56).

Embedded within these basic standards of ethics are testimony, QA, scientific rigor, science culture, hypothesis building, questioning and improving practices, considering alternate explanations, and understanding the causes of bias and error (32). Currently, coursework and training in professional conduct are not mandatory (54,55). Because professionalism is at the root of the forensic sciences, we highly recommend that a course or seminar covering the basic tenets of ethics and professional conduct be developed and required for all forensic scientists (both those in the public and private sectors).

Review of Practices

It is easy to accept forensic science discipline-specific practices as valid and reliable when there is no challenge in the courtroom or when there are unsuccessful challenges to the admissibility of scientific results. Some of the traditional disciplines are steeped in experience that support their reliability but may not have the same validation studies as more novel disciplines. This lack of formalized validation does not necessarily equate to unreliable science. Indeed, in our opinion there is high confidence in the results obtained from forensic science analyses. Yet, we should all review and question our practices to improve and strengthen the foundations of the science. If weaknesses are found (57,58), they should be rectified willingly and with advocacy. In fact, forensic scientists should question their practices routinely, i.e., as stated earlier carry out good science principles.

We cannot, in the space of this paper, provide direction on review of all practices. Therefore, only one example is given here, and that is on sufficiency. Sufficiency is a concept that all forensic scientists encounter. It can be the limit of detection or limit of quantitation of an assay, or it can be related to the degree of attribution (34,35,59–62). One example of sufficiency applies to friction ridge analysis. In this regard, sufficiency relates to the amount of information necessary to render an interpretation of "identification of a single source to the exclusion of all others." While, historically, sufficiency was based on a specified minimum number of ridge characteristics, or "points," this numerical standard was abandoned because it relied only on a subset of the features in a latent print pattern. Instead a holistic approach was advocated, and examiners are trained empirically to recognize what constitutes sufficiency (63,64).

Conveying sufficiency may not be a difficult process by empirical training. It is likely that a degree of uniformity is obtained regarding a sufficiency threshold within a laboratory system, if the trainers are the same individuals and the training program is formalized. However, assessing the degree of uniformity for sufficiency (and the training of the concept and practice of sufficiency) among laboratories has not been done, and the concept or practice is likely to vary. This does not mean that we question the reliability of latent print examinations. The majority of friction ridge impressions either contain a substantial amount of information and are of good quality (or contain portions of good quality) that they pose little problem for analysis or contain so little information and are of such poor quality that they are readily determined insufficient (65). But those impressions that are near the sufficiency threshold may be complex and require additional quality control measures or other considerations. Additionally, we have advocated that there be more research on the minimum qualitative and quantitative criteria required for a conclusion of individuality (66). If a functional and effective quantitative threshold could be developed and instituted, it might reduce the degree of variation that is practiced among forensic scientists especially among different laboratories for asserting identification on those prints that may be borderline in sufficiency.

Absolute uniformity of minimum interpretation criteria is unlikely among forensic scientists without a quantitative threshold. Sufficiency is one example in which the forensic sciences might focus efforts to review and improve current practices. At a minimum, we recommend that all SWGs institute a review of their practices and foundations and then make recommendations for improvement and research direction.

TABLE 2—*Recommendations of actions to be undertaken.*

Require accreditation for any entity providing forensic services
Explore national certification of individual practitioners (both public and private)
Support Scientific Working Groups (SWGs) or similar groups as primary vehicles to carry out research initiatives and collaborations and to address quality issues. The SWGs are essential peer consensus entities that address many of the practices within the disciplines
Develop strategic plan(s) and initiatives for discipline-specific research needs
Develop a list of criteria for development, validation, and implementation of methodologies for each discipline
Develop research needs list (for example see [65] on friction ridge analysis as example)
Develop sourcebooks on the scientific foundations and accepted practices and include reference material, where possible
Publish research needs, validation criteria, and sourcebook in publicly accessible vehicles
Develop national databases on reference materials for better assessment of the weight of evidence
Recognize that current culture of the adversarial system is not constructive for scientific evaluation and develop a positive constructive approach for improvement
Document "inclusions," "exclusions," and "inconclusives" (or whatever terms are used for the comparison process) for all casework analyses
Address the possibility of error in a case by encouraging retesting and/or review (particularly relevant for nonconsumptive analyses)
Institute blind verification as part of any review process
Acknowledge that the DNA statistics model (such as population genetics approaches) is inappropriate for most other disciplines
Develop proper language for conveying the meaning of an association when using qualitative statements to assign the weight of a result
Institute QA standards (which is not the same thing as standardization) with a nationally mandated process, as has been done with DNA typing
Develop standard national training programs for uniform training and assessment of scientists
Train scientists routinely on ethics and bias
Document for each discipline its requirements for obtaining reliable and valid results and the proper limits of interpretation from results obtained
Foster a positive, constructive environment so that there is more incentive to address quality and scientific issues professionally and responsibly

Recommendations to Proceed

Consensus guidelines and standards for quality, education, training, etc. need to be developed for all disciplines; many may be practiced but should be documented. Assumptions, inferences, and significance of all interpretations need to be documented. Additionally, research needs and direction should be documented so all are adequately informed about these practices. We strongly recommend a number of actions be undertaken to ensure that all forensic scientists and their institutions are meeting an acceptable performance level (Table 2).

For any of these recommendations to be effective we strongly advocate adhering to a rigorous QA program, such as previously described. Acceptable QA is best achieved through accreditation. The cost of accreditation is small compared with the potential loss of quality and credibility without it.

As can be seen in the associated summary papers (4–9), disciplines have processes in place that formalize current practices. That is not to say that more research should not be carried out. Research and development in methodologies and technologies from collection to interpretation of the analyses should be advocated and aggressively pursued. Greater outreach to the scientific community should be supported. New methods and techniques require validation. Validation is defined as the process to assess the ability of defined procedures to reliably obtain results, to define conditions that are required to obtain the result, to determine the limitations of the analytical procedure, and to identify aspects that must be monitored and controlled (13,14). Validated methods are essential to the forensic sciences, are inherent in providing quality, and provide stability to continuously evolving scientific fields.

For some analytical results, only qualitative assessments of the significance are conveyed. As discussed earlier, some might suggest that the lack of a definable, statistically derived approach for conveying the significance of an association renders the association meaningless. We do not support such a position. Blind validity testing, or black box testing, in lieu of a specifically defined statistical approach, may be an avenue to evaluate current identification and elimination standards and can help identify possible limitations in an application or discipline (66). Utilizing blind validity testing, examiner interpretations can be tested with various inputs of a range of defined categories of specified items of evidence. This approach would demonstrate whether or not it is possible to obtain a degree of accuracy.

To foster communication, coordinate efforts, leverage resources, and facilitate future direction decisions, we recommend that an International Summit on Forensic Research be convened as soon as possible (and continued periodically). The attendees should include expert practitioners, researchers, professional organization representatives, private industry representatives, academicians, and other interested persons that will help develop a comprehensive strategic plan for forensic science research.

Lastly, we strongly recommend investing in the SWGs and placing them under the successful DNA business model to develop guidelines that can be promulgated to standards nationally. Some SWGs need to expand their membership beyond practitioners to include academic participants, thus achieving a better balance for proceeding forward. These peer consensus SWGs have been and continue to be the best avenues for developing and modifying practices and performing some validation and standards studies.

Conclusion

With a good QA program, errors can be minimized and consequently be used to improve laboratory and personnel processes. Hypothetical error rates add little value to the evaluation of the strength of the evidence in a specific case. Errors of consequence due to mistakes or bias, which are a serious concern for all, are identified and addressed best through peer review by retesting, reanalysis, and/or blind verification.

Forensic science is evaluating itself and is improving its practices on a continuous basis. Improvement does not mean that the current practices are inadequate. Enhancing the forensic disciplines should continue and must be advocated. We have outlined a number of areas that should be given immediate attention. With proper support, to include management, practitioners, and adequate funding, the processes can be effected positively and more expeditiously.

References

1. Stacey RB. A report on the erroneous fingerprint individualization in the Madrid train bombing case. J Forensic Identif 2004;54(6):706–18.
2. Sheck B, Neufeld P, Dwyer J. Actual innocence. New York, NY: New American Library, Penguin Putnam, Inc., 2000.
3. Saks MJ, Koehler JJ. The coming paradigm shift in forensic identification science. Science 2005;309:892–5.
4. Bottrell MC. Forensic glass comparison: background information used in data interpretation. Forensic Sci Commun 2009;11(2), http://www.fbi.gov/hq/lab/fsc.
5. Bunch SG, Smith ED, Giroux BN, Murphy PD. Is a match really a match? A primer on the procedures and validity of firearm and toolmark identification. Forensic Sci Commun 2009;11(3), http://www.fbi.gov/hq/lab/fsc.
6. Harrison D, Burkes TM, Seiger DP. Handwriting examination: meeting the challenges of science and the law. Forensic Sci Commun 2009;11(4), http://www.fbi.gov/hq/lab/fsc.
7. Oien CT. Forensic hair comparison: background information for interpretation. Forensic Sci Commun 2009;11(2), http://www.fbi.gov/hq/lab/fsc.
8. Peterson PE, Dreyfus CB, Gische MR, Hollars M, Roberts MA, Ruth RM, et al. Latent prints: the state of the science. Forensic Sci Commun 2009;11(4), http://www.fbi.gov/hq/lab/fsc.
9. Smith MB. The forensic analysis of footwear impression evidence. Forensic Sci Commun 2009;11(4), http://www.fbi.gov/hq/lab/fsc.
10. Budowle B, Allard MW, Wilson MR, Chakraborty R. Forensics and mitochondrial DNA: applications, debates, and foundations. Annu Rev Genomics Hum Genet 2003;4:119–41.
11. Bird SJ. Scientific certainty: research versus forensic perspectives. J Forensic Sci 2001;46(4):978–81.
12. ISO/IEC. ISO/IEC 17025:2005 General requirements for the competence of testing and calibration laboratories, 2nd ed. Geneva, Switzerland: ISO Copyright Office, Reference Number ISO/IEC 17025:2005(E), 2005.
13. Budowle B, SWGMGF Members. Quality assurance guidelines for laboratories performing microbial forensic work. Forensic Sci Commun 2003; 5(4): at http://www.fbi.gov/hq/lab/fsc/backissu/oct2003/index.htm.
14. Budowle B, Smith JA, Moretti T, DiZinno J. DNA typing protocols: molecular biology and forensic analysis. Natick, MA: BioTechniques Books, BioForensic Sciences Series, Eaton Publishing, 2000.
15. The International Association for Identification. Scientific Working Group Documents. http://www.theiai.org (accessed October 3, 2008).
16. American Society of Quality (ASQ), Plan-do-check-act cycle. http://www.asq.org (accessed October 3, 2008).
17. Six sigma, http://www.isixsigma.com/sixsigma/six_sigma.asp (accessed October 3, 2008).
18. ASCLD/LAB. Supplemental requirements for the accreditation of forensic science testing and calibration laboratories. ASCLD/LAB-International supplemental requirements, 2006, AL-PD-3003-Ver 2.1. Garner, NC: ASCLD/LAB, 2006.
19. ILAC. Guidelines for forensic science laboratories, ILAC-G19, 2002, http://www.ilac.org/documents/g19_2002.pdf.
20. FQS-I 17025 Accreditation, http://www.forquality.org (accessed October 3, 2008).
21. Texas Code Crim. P. Ann. Art. 38.35, 2005.
22. Oklahoma Stat Ann. Tit. Forensic Laboratory Accreditation Act §74-150.37, 2002.
23. N.Y. Crim. Proc. §995-b.
24. Koehler JJ, Chia A, Lindsey S. The random match probability (RMP) in DNA evidence: irrelevant and prejudicial? Jurimetrics J 1995;35:201–19.
25. National Research Council II Report. The evaluation of forensic evidence. Washington, DC: National Academy Press, 1996.
26. Thompson WC. Subjective interpretation, laboratory error and the value of forensic DNA evidence: three case studies. Genetica 1995;96:153–68.
27. Thompson WC, Taroni F, Aitken CG. How the probability of a false positive affects the value of DNA evidence. J Forensic Sci 2003; 48:47–54.
28. Harmon R, Budowle B. Questions about forensic science. Science 2006;311:607.
29. Grieve DL. Possession of truth. J Forensic Identif 1996;46:521–8.
30. Peterson JL, Lin G, Ho M, Chen Y, Gaensslen RE. The feasibility of external blind DNA proficiency testing. I. Background and findings. J Forensic Sci 2003;48(1):21–31.
31. Houck MM, Budowle B. Correlation of microscopic and mitochondrial DNA hair comparisons. J Forensic Sci 2002;47:964–7.
32. Byrd JS. Confirmation bias, ethics, and mistakes in forensics. J Forensic Identif 2006;56:511–25.
33. Saviano J. The pursuit of objectivity in the examination of forensic evidence. J Forensic Identif 2006;56(6):877–84.
34. Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Budowle B. Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. J Forensic Sci 2001;46:647–60.
35. Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Budowle B. Validation of STR typing by capillary electrophoresis. J Forensic Sci 2001;46:661–76.
36. Nisbett RE, Ross L. Human inference: strategies and shortcomings of social judgment. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1980.
37. Dror IE, Charlton D, Péron A. Contextual information renders experts vulnerable to making erroneous identifications. Forensic Sci Int 2006;156:74–8.
38. Dror IE, Charlton D. Why experts make errors. J Forensic Identif 2006;56:600–16.
39. Saks MJ, Risinger DM, Rosenthal R, Thompson WC. Context effects in forensic science: a review and application of the science of science to crime laboratory practice in the United States. Sci Justice 2003;43:77–90.
40. Krane DE, Ford S, Gilder JR, Inman K, Jamieson A, Koppl R, et al. Sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation. J Forensic Sci 2008;53:1006–7.
41. Office of the Inspector General. The FBI Laboratory: an investigation into laboratory practices and alleged misconduct in explosives-related and other cases. Washington, DC: US Department of Justice, 1997.
42. Stafford Smith CA, Goodman PD. Forensic hair comparison analysis: nineteenth century science or twentieth century snake oil? Columbia Human Rights Law Rev 1996;27:227–91.
43. Robertson J. Forensic and microscopic examination of human hair. In: Robertson J, editor. Forensic examination of hair. London: Taylor and Francis, 1999;79–154.
44. Evett IW, Weir BS. Interpreting DNA evidence. Sunderland, MA: Sinauer Associates Inc., 1998.
45. SWGMAT. Expert reporting guideline. Forensic Sci Commun 2009;11(1), http://www.fbi.gov/hq/lab/fsc/current/standards/2009_01_standards02.htm.
46. U.S. Department of Justice. Education and training in forensic science: a guide for forensic science laboratories, educational institutions, and students. Washington, DC: U.S. Department of Justice, Office of Justice Programs, 2004.
47. Siegel JA. The appropriate educational background for entry level forensic scientists: a survey of practitioners. J Forensic Sci 1988;33:1065–8.
48. Aitken CGG, Taroni F. Statistics and the evaluation of evidence for forensic scientists, 2nd ed. West Sussex: John Wiley & Sons, Inc., 2004.
49. Lucy D. Introduction to statistics for forensic scientists. West Sussex: John Wiley & Sons, Inc., 2005.
50. Taroni F, Aitken CGG, Garbolino P, Biedermann A. Bayesian networks and probabilistic inference in forensic science. West Sussex: John Wiley & Sons, Inc., 2006.
51. FEPAC. http://www.aafs.org/default.asp?section_id=aafs&page_id=committee&subpage_id=fepac (accessed July 31, 2008).
52. Almirall JR, Furton KG. Trends in forensic science education: expansion and increased accountability. Anal Bioanal Chem 2003;376:1156–9.
53. Scientific Working Group for Materials Analysis (SWGMAT). Forensic fiber examiner training program. Forensic Sci Commun 2005;7(2), http://lab.fbinet.fbi/library/fsc/backissu/april2005/standards/SWGMAT_fiber_training_program.pdf.
54. Fleckstein MP, Martone M, Pitluck HM. Law, ethical codes, and the report of the CSSP survey on ethics policies. J Forensic Sci 2001; 46:982–5.
55. Peterson JL. Teaching ethics in a forensic science curriculum. J Forensic Sci 1988;33:1081–5.
56. http://www.waduicenter.com/toolbox/documents/ASCLD-labmgtguide.pdf (accessed July 31, 2008).
57. National Research Council. Forensic analysis: weighing bullet lead evidence. Washington, DC: National Academy Press, 2004.
58. National Research Council. Committee to assess the feasibility, accuracy and technical capability of a national ballistics database. Washington, DC: National Academy Press, 2008.

59. Bodziak WJ. Footwear impression evidence: detection, recovery, and examination, 2nd ed. Boca Raton: CRC Press, 2000;347.
60. Clinical Laboratory Standards Institute. User verification of performance for precision and trueness; approved guideline, 2nd ed. Wayne, PA: Clinical and Laboratory Standards Institute (CLSI Document EP15-A2), 2005.
61. Clinical and Laboratory Standards Institute. Genotyping for infectious diseases: identification and characterization; approved guideline. Wayne, PA: Clinical and Laboratory Standards Institute (CLSI Document MM10-A), 2006.
62. Budowle B, Chakraborty R, Carmody G, Monson KL. Source attribution of a forensic DNA profile. Forensic Sci Commun 2000;2, http://www.fbi.gov/programs/lab/fsc/current/budowle.htm.
63. Evett IW, Williams RL. A review of the sixteen points fingerprint standard in England and Wales. J Forensic Identif 1996;46:49–73.
64. International Association for Identification. Standardization committee report. FBI Law Enforcement Bull 1973;42:7–8.
65. Vanderkolk JR. Forensic individualization of images using quality and quantity of information. J Forensic Identif 1999;49(3):246–56.
66. Budowle B, Buscaglia J, Schwartz Perlman R. Review of the scientific basis for friction ridge comparisons as a means of identification: committee findings and recommendations. Forensic Sci Commun 2006;8, http://www.fbi.gov/hq/lab/fsc.

Additional information and reprint requests:
Bruce Budowle, Ph.D.
FBI Laboratory
2501 Investigation Parkway
Quantico, VA 22135
E-mail: bbudowle@ic.fbi.gov