NATIONAL COMMISSION ON FORENSIC SCIENCE

National Institute of Standards and Technology
U.S. Department of Commerce

Optimizing Human Performance in Crime Laboratories through Testing and Feedback

Type of Work Product: Adjudication of Public Comments on Draft Document

Itemized Issues and Adjudication Summary:

DOJ-LA-2016-0011-0036 (Kimberly Murga)

(1) "Blind testing is a great idea that needs extensive work to develop a successful model before it can be widely utilized in crime labs."

The Subcommittee agrees with this comment. The revised document includes the following statement acknowledging that a period of trial and error may be needed to learn how to run such programs most effectively:
In order to facilitate development of research programs of the type described here the Commission recommends that the Department of Justice and other funding agencies provide funding to laboratories willing to undertake such studies. Pilot projects in which laboratories establish research programs while monitoring how well these programs work, will be particularly valuable in charting a path forward on this issue. The practical experience of laboratories that pioneer the development of such programs should be recorded and disseminated for the benefit of the entire forensic science community. A period of trial and error will undoubtedly be necessary to learn how best to set up and run effective research programs within forensic laboratories.

(2) Error rates are difficult to calculate for tests like GC/MS and extensive validation has been done.

The Subcommittee agrees with this statement and believes that it would nevertheless be worthwhile to conduct research with blind samples to verify such analytic systems are working as expected in practice and to explore any limitations or uncertainties that may arise in casework.

DOJ-LA-2016-0011-0023 (Cecelia Crouse)

(1) Should this be a recommendation rather than a views document, as it includes some recommendations?

The primary goal of this document is to express the views of the Commission on these issues, rather than direct the Attorney General to take specific actions. Accordingly, the subcommittee has chosen to frame this as a views document. We note that previous Commission views

documents have included some policy recommendations.

(2) and (5) How can the rate of error that might occur in a forensic laboratory be estimated from data on the occurrence of error in research studies?

These comments raise an interesting technical question best addressed by statisticians and which is beyond the scope of this document.

(3) (a) (1) The document should be limited to addressing performance of practitioners, not of laboratories

Because laboratory performance is directly linked to the performance of individual practitioners, the subcommittee sees no need to distinguish lab performance from practitioner performance in this document.

(3)(a)(ii) The term "overall quality" needs to be defined and distinguished from "high reliability."

The subcommittee appreciates this suggestion. The terms accuracy and reliability are now explained in footnote 1. The term "high reliability organization" is explained in Footnote 2. The term "overall quality" as used in the first sentence of the Background section has a common sense meaning. In the opinion of the Subcommittee it requires no further explanation.
(3)(b)(c)-"Forensic science evidence"

We have replaced this phrase with the phrase "a forensic scientist's conclusions" in order to distinguish the conclusion from the underlying evidence (e.g., traces, samples) on which it is based.

(3)(d)—"Forensic science service providers"

For clarity we changed the term to forensic science practitioner, as suggested.

(3)(e)(f)—Claim that proficiency testing and peer review provide feedback to examiners that is equivalent to knowing ground truth through the use of known-source research samples.
The subcommittee respectfully disagrees with these comments. The text of the views document explains clearly the basis for the statements that the comment challenges.

(4) (a)(b)(c) Comments on terminology.

After careful consideration, the subcommittee found these comments ultimately unpersuasive and opted to maintain the existing language of the document rather than adopt the proposed alternative terminology.

(6) (b) It is not the purpose of proficiency testing to hone or improve analyst's skills. Limitations of the reliability and accuracy of analytic methods should be established through method validation research, not proficiency testing.

The subcommittee agrees with the feedback regarding the term "hone" and it has been excised from the document. The comment about the limitations of proficiency testing for improving skills of experienced analysts is entirely consistent with statements in the current version of the views document on this issue. The subcommittee respectfully disagrees with the claim that research on the performance of examiners when testing blind samples in the course of routine casework is unnecessary in light of method validation research. Methods validation research has not been performed in every forensic science discipline. Even in those disciplines where methods validation research has been conducted, however, important questions still exist about the accuracy of the underlying methods in practice.

(7) The challenges in forensic laboratories are much greater than those of medical laboratories- this should be noted.

The subcommittee agrees with this comment and we have added a footnote (note 18), making this point and arguing that the greater challenges in forensic laboratories make research on laboratory performance even more important for forensic science laboratories than for medical laboratories.

(8) Would it be possible to determine ground truth if samples were prepared by another laboratory?

The focus of this views document is on methods used for source determination. For that task, ground truth is easy to establish. Two biological samples either come from the same person or not; two fingerprints either come from the same finger or not. Ground truth is established by creating the samples in a documented manner. The subcommittee believes this point is implicit in the document and that no additional clarification is needed.

(9)(a) "….it is not clear if challenging performance testing is a one-time thing or ongoing."

The document has been re-written to emphasize that what is being proposed is research on laboratory performance. Research of this type would be performed at the discretion of laboratory management and would not be mandatory.

(9)(b) The misspelling noted in this comment has been corrected

(9)(c) "Why would performance tests that are spiked for errors be necessary? Programming an individual for failure just does not seem appropriate."

The revised document now includes extensive discussion of the reasons that researchers may find it useful to test examiners' performance when processing challenging samples. Among the materials that address this question are the following:

Studies that involve highly challenging samples will be particularly valuable for helping examiners improve their skills. For example, latent print examiners sometimes need to make critical decisions about whether a low quality latent print (e.g., a print containing limited detail or distortions) can accurately be identified, or whether the comparison should be deemed inconclusive. Research on this issue will not only address general concerns about the validity

of judgments in such cases, it will also provide feedback that will help examiners improve their decision making in such cases.

And,
Engineers often test products and systems by subjecting them to conditions (e.g., stress, strain, pressure) in excess of normal service parameters. Known as "accelerated life testing," this process is useful for uncovering faults and potential modes of failure in a relatively short time. Nelson, W. (1980). "Accelerated Life Testing - Step-Stress Models and Data Analyses". IEEE Transactions on Reliability (2): 103.doi:10.1109/TR.1980.5220742; Donahoe, D.; Zhao, K.; Murray, S.; Ray, R. M. (2008). "Accelerated Life Testing". Encyclopedia of Quantitative Risk Analysis and Assessment.doi:10.1002/9780470061596.risk0452.ISBN 9780470035498. Research on how well forensic laboratories perform when processing highly challenging samples would have similar benefits.

(9)(d) Should attorneys not be concerned about a link between passing a challenging test and incompetence?

The revised document addresses this concern. It now explains carefully that the occurrence of error in research studies designed to determine the circumstances under which errors occur are not necessarily an indication that an examiner is incompetent. It also urges courts to be circumspect in allowing evidence of errors in research studies to be used to challenge examiner competence.

(10)(a) Database matches are already verified.

As the document points out, research with known source samples would be useful for addressing a variety of questions about the effectiveness of database searches.

(10)(b) "…as of now, as you are aware, we legally can't tell someone how many candidate matches were returned as part of the search."

The document recognizes that it may be necessary to change some of the rules that currently govern access to databases in order to facilitate research of the type proposed. The problem cited illustrates why there is a need (as noted in the document) for the Department of Justice and the FBI to examine current rules and to revise any that would interfere with research of the type proposed.

(10)(c)—Individuals who agree to submit samples for research purposes will go into databases. They will then have to petition to have their profiles removed.

The document discusses this potential problem and recommends ways to deal with it by preventing research samples from being permanently entered into databases.

(11) Errors that examiners make in research studies will inevitably be used against them in court.
The subcommittee recognizes this concern and the revised document responds to it. The

document now explains carefully that the occurrence of error in research studies designed to determine the circumstances under which errors occur are not necessarily an indication that an examiner is incompetent. It also urges courts to be circumspect in allowing evidence of errors in research studies to be used to challenge examiner competence.

(12)(a) Privacy issues will need to be evaluated…

See responses as to 10(b) and 10(c). Steps would of course need to be taken to protect the rights and interests of individuals who submit samples to be used in research. Those issues will need to be addressed for each study by the researchers who conduct the study. In the view of the subcommittee, the fact that such difficulties might arise for some kinds of research is not a reason to forgo moving forward with research in general.

(12)(b) "…cannot truly have a 'fake' database if this is truly a blind proficiency test."

This comment is a bit unclear to the subcommittee. The views document does not propose a "fake" database. It proposes facilitating certain types of research on the performance of bench-level forensic examiners. Among the research studies that might be possible using the proposed approach are studies that could entail searching known-source specimens (e.g., biological samples; latent prints) against existing databases. The value of performing such studies is noted. Possible challenges of such studies are also noted.

Board of Directors of the American Society of Crime Laboratory Directors

1. Replace the terms "validation;" "error rate estimation;" and "proficiency testing" with alternative terminology that is "consistent with the purpose of human factors research…"

We respectfully decline to adopt this proposed change. The comment suggests (without explanation) that the views document is using these terms in a manner inconsistent with the way these terms are used in "the forensic science industry." The subcommittee is not persuaded that the use of these terms in the views document varies from standard practice in forensic science and does not believe these terms will be misunderstood. Footnote 1 includes a clear explanation of the terms reliability, accuracy and validity.

2. An "acceptable" failure rate to the criminal justice system is not addressed.

The subcommittee is not convinced that there is a need to discuss in this document what rate of error might be deemed unacceptable by legal professionals. The purpose of the document is to encourage and facilitate additional research needed to understand how well bench-level forensic examiners perform when carrying out routine analytic procedures. While it is conceivable that such research might reveal rates of error on some tasks under some circumstances that are high enough that some legal professionals would find them unacceptable, it is not necessary to discuss where the threshold of acceptability might lie in advance of doing the research.

3. "What is the proposed consequence of failure to obtain the correct answer on a challenging test set? How will a failure by an examiner who is not deficient be differentiated from an examiner who is able to successfully pass a routine proficiency test but is deficient in training, diligence or skills?"

The views document takes care to warn against equating an error in a research study with incompetence and failure. Specifically it says:
Because errors are expected (and, indeed, desirable) in research of this type, it is imperative to avoid the naïve mindset that associates error of any sort with incompetence. That association might be appropriate for proficiency tests designed to ascertain whether practitioners have the minimal level of proficiency needed to do their jobs. It is inappropriate when evaluating the results of research studies that are designed specifically to determine when and where errors occur. In research of this type, the occurrence of an error should be viewed as a valuable opportunity for feedback, learning, and improvement. It is not necessarily an indication of deficiency in the training, diligence or skills of the individual who makes the error.
In order to clarify the issue of consequences, the revised views document now says explicitly: *Practitioners should not be punished or sanctioned for making errors in such studies.*

4. "…implementing blind performance testing into routine case work at a forensic science laboratory for validation, training, and improvement purposes is akin to collecting data from accidents and injuries that occur when common drivers are randomly introduced to high speed race tracks or icy mountain roads. This is an inefficient, dangerous, and scientifically inappropriate research technique."

The subcommittee is not convinced that this analogy is helpful. Errors that might occur during research studies in forensic laboratories have no dangerous consequences for the general public. They are analogous to the errors that pilots make while being tested in flight simulators. Simulator errors can be extremely valuable for learning to be a better pilot, and for exploring the circumstances under which pilot error might occur, but no one is actually injured. The same is true for errors that might occur in research studies testing the performance of forensic scientists.
No one would say that testing pilots in flight simulators to see how well they perform is "inefficient, dangerous and scientifically inappropriate." Such claims are equally unpersuasive when made against a proposal to study the performance of forensic scientists. Indeed, if there is a danger to be avoided, it is the danger that labs will fail to realize their errors in routine casework because they fail to undertake research on examiners' performance.

5. "The inability of an examiner to obtain the correct result, even under conceptually innocent circumstances, will be used against the examiner and the laboratory by the legal community, especially if monetary gain is involved. This is an untenable situation for practitioners and laboratories ensuring a climate of 'Gotcha' in which an air of constant liability exists."

The subcommittee has addressed this concern in the revised views document which has been modified to acknowledge this important concern (see note 26, and accompanying text). It also includes language urging courts to be circumspect in allowing evidence of errors in research

studies to be used to challenge examiner competence. While acknowledging this concern, the views document also notes:
It would obviously be a travesty if valuable research on the performance of forensic laboratories is not undertaken because forensic scientists fear the consequences should the results of that research become known.

6. "Any testing conducted should be done by an entity, whose methodology, sampling, and standards are approved by NIST (OSAC)"

The views document suggests that NIST be involved in the creation of research test sets. It also suggests that those undertaking such research consult with statisticians, practitioners and experts of research methodology to assure that the studies performed are scientifically rigorous and that they address appropriate questions. The subcommittee is not persuaded of the need to go beyond these suggestions in order to *require* that NIST or OSAC approve the entity conducting the research. For example, if the managers of a forensic laboratory wanted to cooperate with university researchers in designing a blind study of laboratory performance, it is unclear why permission of NIST or OSAC should be necessary to proceed with that research.

7. "The identity of the laboratory and analyst should not be documented."

The subcommittee appreciates this suggestion but feels that it is impractical. Laboratory managers would of course be aware of the identity of any analysts who participated in a research study and would therefore be able to ascertain who made errors, if errors occurred. Identifying those analysts would be necessary to provide them with feedback on their performance, which is one of the goals of the research. Laboratory managers would presumably have some discretion in how and whether they disclosed such information outside the laboratory, although that discretion might be limited by legal disclosure requirements.

8. "Any test should possess a conditional statement including a variation of the following language:

> "The results of this study are experimental in nature and do not reflect the error rate of X. These results do not represent the error rate for examiners in actual casework, and this testing was conducted to test the limits of specific human factors in an effort to further improve practices."

This disclaimer or similar language should be included, as a matter of OSAC/NIST approval."

For reasons discussed above under point 6, the subcommittee disagrees with the suggestion that OSAC/NIST approval should be required before a forensic laboratory undertakes research on the performance of laboratory personnel. The subcommittee similarly declines to include the suggestion that those who conduct such studies should be required to include the statement proposed in this comment, or any other pre-configured statement, when reporting the results of the studies. The way in which scientific researchers choose to report their findings has traditionally been left to the researchers themselves. While researchers often receive guidance from peer-reviewers and journal editors, as a matter of basic academic and scientific freedom the ultimate decision about what to say about the findings of a scientific study is left to the researchers themselves. The suggestion that individuals studying the performance of forensic scientists should be required to include the suggested boilerplate statement when reporting

findings is inconsistent with traditional scientific values and the subcommittee therefore declines to adopt it.

9. "A search of a challenging blind DNA profile in the Combined DNA Index System (CODIS) would be expected to result in multiple candidates. CODIS does not rank the candidate matches…. The amount of time needed to complete a review of the candidate matches for a blind performance test will be dependent on the number of candidates resulting from the search and the age of the case if casefiles are archived off-site from the laboratory."

The concern raised here is apparently that research involving the comparison of samples with CODIS might be time-consuming. This is a valid concern, but it is a concern that could and should be evaluated by the managers of each laboratory before deciding whether to undertake a particular type of study. If the laboratory managers concluded that the research would be too costly in time or opportunity cost, they could decline to participate in the research. The views document is not suggesting that laboratories be required to engage in research of this type. It is merely suggesting that the government take steps to encourage and facilitate such research, in order to make it easier for laboratories that wish to conduct studies to do so. The current version of the views document clarifies these points. The subcommittee believes the current document is adequate to address these concerns.

10. "Clarification is needed to determine if the time spent testing the[DNA] database is worth the time taken away from casework backlog reduction…Latent print database testing offers similar challenges. "

See response to item 9. The revised version of the views document clarifies that research would be at the discretion of laboratory managers who would be able to weigh such concerns against the benefits of research.