NATIONAL COMMISSION ON FORENSIC SCIENCE

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

Views of the Commission
Facilitating Research on Laboratory Performance

Subcommittee
Human Factors
Status
Adopted by the Commission

Date of Current Version	18/08/16
Approved by Subcommittee	21/08/16
Approved by Commission	13/09/16

Commission Action
On September 13, 2016, the Commission voted to adopt this Views Document by a more than two-thirds majority affirmative vote (100% yes, 0% no, 0% abstain)

Note: This document reflects the views of the National Commission on Forensic Science and does not necessarily represent the views of the Department of Justice or the National Institute of Standards and Technology. The portion of the document directly labeled "Views of The Commission" represents the formal Views of the Commission. Information beyond that section is provided for context. Views documents do not request specific action by the Attorney General, and thus do not require further action by the Department of Justice upon their approval by the Commission. The National Commission on Forensic Science is a Federal Advisory Committee established by the Department of Justice. For more information, please visit: https://www.justice.gov/ncfs.

Views of the Commission

It is the view of the National Commission on Forensic Science that:

1. Additional research is needed to assess the performance of forensic science laboratories on routine analytic tasks such as comparison of samples to determine whether they have a common source.
2. Studies should be conducted by introducing known-source samples into the routine flow of casework in a blinded manner, so that examiners do not know their performance is being studied.
3. Government agencies should facilitate research of this type by funding pilot research programs.
4. Government agencies should facilitate research of this type by developing (or funding the development of) sets of test samples that can be used to carry out research. The research

test sets should be designed with the assistance of practitioners, statisticians, and experts on research methodology to facilitate studies that address important scientific questions.

5. Government agencies should identify and revise any regulations or memoranda of understanding regarding access to databases that prohibit, or appear to prohibit, access to these databases for the purpose of conducting research on laboratory performance.
6. To avoid unfairly impugning examiners and laboratories who participate in research on laboratory performance, judges should consider carefully whether to admit evidence regarding the occurrence or rate of error in research studies. If such evidence is admitted, it should only be under narrow circumstances and with careful explanation of the limitations of such data for establishing the probability of error in a given case.

Background

Forensic science plays a vital role in the criminal justice system. It is therefore essential that forensic science service providers take steps to assure the accuracy and reliability of their analyses and the overall quality of their work.[1] Crime laboratories should strive to be *high reliability organizations*.[2]

It is the view of the Commission that additional research is needed to assess the performance of forensic laboratories on routine analytic tasks. Too little is known about the reliability of examiners' judgments and accuracy of their conclusions. Too little is known about the circumstances under which their accuracy will be impaired.

In an ideal research program, the bench-level analysts, whose judgments are critical to the test results, will not know they are being studied. The research samples will be incorporated into the routine flow of casework in a manner that makes them indistinguishable from other samples being examined by the laboratory.[3]

---

[1] A forensic analysis is *accurate* when it produces correct conclusions; it is *reliable* when there is consistency within and among examiners in their interpretation of a given item or set of evidence. A method is *valid* when it has been shown to produce reliable and accurate results.

[2] Roberts, K. H. (1990). Some characteristics of one type of high-reliability organization. *Organization Science*, 1, 160–176; Weick, K. E., & Sutcliffe, K. M. (2007). *Managing the unexpected: resilient performance in an age of uncertainty (2nd ed.)*. San Francisco, CA: Jossey-Bass. The term "high reliability organization" (HRO) is used by management and organizational theorists to describe organizations that have succeeded in avoiding catastrophic failures in environments in which such failures are possible. Among the most important characteristics of HROs are their intensive focus on identifying potential sources of error and their efforts to evaluate their own performance in an open, forthright manner that avoids blaming and stigmatization.

[3] Informing someone that he or she is being tested can create what psychologists call demand characteristics that change the person's responses. Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist,* 17, (11): 776–783. doi:10.1037/h0043424. Individuals who know they are being tested may shift their threshold of decision in ways designed to make them look good. Paulhus, D. L. (1991). Measurement and control of response biases. In J.P. Robinson et al. (Eds.), *Measures of personality and social psychological attitudes*. San Diego, CA: Academic Press. Hence, performance testing will provide a more realistic picture of analytic performance if the analysts do not know they are being tested.

Blind studies of this type are difficult to conduct in laboratories where examiners communicate directly with detectives and have access to police reports and other information. To conduct such a study, laboratory managers need to enlist the support of police in preparing simulated case materials. The materials must be sufficiently realistic to pass as a real case. Furthermore, the police might need to provide other false information to the examiner to prevent the examiner from discovering that the case was simulated. Elaborate simulations of this type are feasible and have been conducted successfully to test the accuracy of DNA analysis,[4] but they are burdensome and expensive. They may also be problematic in other ways. Some observers may have qualms, for example, about involving police officers in the creation of false or simulated evidence.

Fortunately, blind studies are much easier to conduct in laboratories that employ a context management system to shield examiners from task-irrelevant contextual information. If bench-level examiners are typically exposed only to the specimens presented for analysis, along with any task-relevant information passed on by the case manager, the case manager can occasionally insert research samples without the examiner being able to distinguish the research samples from routine casework. The case manager would know it was a research sample; the examiner would not. Successful blind research programs of this type have been implemented in forensic laboratories.[5] The ability to implement blind research in this way is a secondary benefit that arises when laboratories adopt context management systems, as recommended by this Commission.[6]

Once a laboratory develops the ability to conduct blind studies, it becomes possible to do research on a variety of important issues by experimentally varying the nature of the research samples. Among the issues that could be addressed by research studies of this type are the following:

(1) Validation

In 2009, the National Academy of Sciences commented on the need for additional research on the performance of forensic laboratories. It called for research to "address issues of accuracy, reliability, and validity in the forensic science disciplines," saying that such research is needed to "establish the limits of reliability and accuracy that analytic methods can be expected to achieve as the conditions of forensic evidence vary."[7]

---

[4] Peterson, J. L., Lin, G., Ho, M., Ying, C., & Gaensslen, R. E. (2003) The feasibility of external blind DNA proficiency testing I. Background and findings. *Journal of Forensic Sciences*, 48(1), 21–31.

[5] For example, the Houston Forensic Science Center has been conducting blind testing in three of its disciplines: controlled substance, blood alcohol, and firearms analysis. It is planning to expand the blind testing program to latent print analysis and DNA analysis. This program is reportedly working well in Houston and is facilitated by the laboratory's adoption of the kind of context management system recommended by the Commission in the views document entitled *Ensuring that Forensic Analysis Is Based on Task-Relevant Information.* A similar program has been adopted by the Netherlands Forensic Institute.

[6] National Commission on Forensic Science. Views document on *Ensuring that Forensic Analysis is Based Upon Task-Relevant Information*. Adopted by the Commission December 8, 2015.

[7] National Academy of Sciences. (2009). *Strengthening forensic science in the United States: A path forward* (p. 23, Recommendation 3). Washington, DC: The National Academies Press, (hereinafter 2009 NAS report)

Introducing known-source samples into the normal flow of casework will provide valuable information on the reliability and accuracy of laboratory methods. For example, known-source latent prints might be submitted to a number of latent print examiners to determine how consistently they determine whether the prints are "of value for comparison" and how accurately they associate the prints with reference prints. By varying the nature of the latent prints (e.g., the size and clarity of the image, degree of distortion, substrate), researchers can study the effect of these variables on examiner performance and gain insight into circumstances in which performance is impaired.[8] Similar studies could be designed to explore the reliability and accuracy of other forensic disciplines as well.[9]

(2) Training and Improvement

To achieve and maintain optimal levels of performance on challenging tasks, people need feedback. They need information on how well they are doing; they need to be told when their judgments are sound and when their judgments are mistaken, incomplete, or otherwise suboptimal.[10]

For some challenging intellectual tasks, feedback is automatic. A pilot who makes a navigation error, for example, is likely to learn of the error quickly when the plane fails to reach the expected destination. For many of the most vital judgments made by forensic science practitioners, however, feedback is not automatically available or is incomplete and unreliable. This is particularly true for analytic judgments. An analyst evaluating whether a latent print was made by a suspect, for example, will rarely know with certainty the ground truth of the matter.[11]

---

[8] 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*. Hoboken, NJ: John Wiley & Sons. doi:10.1002/9780470061596.risk0452. ISBN 9780470035498. Research on how well forensic laboratories perform when processing highly challenging samples would have similar benefits.

[9] The 2009 NAS report recognized the importance of such research:

> Although a long-term research agenda will require a thorough assessment of each of the assumptions that underlie forensic science techniques, ***many concerns regarding the forensic science disciplines can be addressed immediately through studies in which forensic science practitioners are presented with a standardized set of realistic training materials that vary in complexity.*** Such studies will not explore the components of the decision process, but they will permit an assessment of the extent to which skilled forensic science practitioners will reach the same or similar conclusions when presented with the types of materials that lead to disagreements. (p. 189) (emphasis added).

[10] These principles are widely recognized in personnel management. *See e.g.*, the statement on the importance of providing feedback to employees by the United States Office of Personnel Management: https://www.opm.gov/policy-data-oversight/performance-management/performance-management-cycle/monitoring/feedback-is-critical-to-improving-performance/

[11] Although it has been argued that the results of latent print analysis are tested through the adversarial process of the trial itself, *see e.*g., *US v. Havvard*, 260 F.3d 597 (7th Cir. 2001), this kind of "testing" is woefully inadequate from a scientific perspective. Risinger, D. M., & Saks, M. J. (1996) Science and non-science in the courts: Daubert meets handwriting identification expertise, *Iowa L. Rev*, 82, 21, at 33–34 and 41 fn. 100. One cannot draw reliable inferences about the accuracy of a forensic scientist's conclusions in a criminal case from the conviction or acquittal of a defendant; indeed, such arguments are tautological to the extent that the forensic scientist's conclusions contributed to the legal outcome that provides the "test" of its validity.

There is only one consistently effective way to provide valid feedback to analysts on the accuracy of their performance—they must be tested using samples for which the ground truth is known.[12] Research programs of the type discussed here have the potential to provide that feedback. After completing the analysis of a research sample, examiners can be told about its source, allowing them to assess their performance and learn from any errors.

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 reliability and accuracy of judgments in such cases but will also provide feedback that will help examiners improve their decision making in such cases.

(3) Error Rate Estimation

To evaluate the probative value of a forensic scientist's conclusions for proving a proposition (e.g., that a latent print was made by a particular individual), it is important to know the rate at which laboratory analysis produces erroneous conclusions. The 2009 NAS report declared: "The assessment of the accuracy of conclusions from forensic analyses and the estimation of relevant error rates are key components of the mission of forensic science."[13] Research involving known-source samples that are processed as part of routine casework would provide valuable information about the rate at which errors occur in various types of analysis.[14] Testing for the purpose of error rate estimation should involve samples designed to replicate those routinely encountered in casework.[15] It would be misleading to attempt to estimate error rates from performance on samples designed to be unusually challenging and difficult. However, it might well be useful to estimate the rate of error separately for different types of comparisons or classes of samples if the level of difficulty (and hence the expected rate of error) differed for different types of cases or different types of samples.

(4) Quality Control and Quality Assurance

In clinical medicine, routine testing of laboratory performance has been a prominent feature of laboratory analysis since 1988, when Congress passed the Clinical Laboratory Improvement Amendments (CLIA). CLIA has been credited with bringing about remarkable improvement in the quality of clinical testing; it brought an end to scandals that previously plagued the field of

---

[12] Thorndike, E. L. (1927). The law of effect. *American Journal of Psychology, 39*, 212–222; Ammons, R. B. (1956). Effects of knowledge of performance: A survey and tentative theoretical formulation. *Journal of General Psychology, 54*, 279–299; Annett, J. (1969). *Feedback and human behaviour*. Harmondsworth, Middlesex, England: Penguin Books. Arkin, R. M., & Walts, E. A. (1983). Performance implications of corrective testing. *Journal of Educational Psychology, 75*, 561–571; Causer, J., Barach, P., & Williams, .A.M. (2014). Expertise in medicine: using the expert performance approach to improve simulation training. *Medical Education*, 48(2): 115–23.

[13] 2009 NAS report, at 122.

[14] See Koehler , J. J. (2016). Forensics or Fauxrensics? Ascertaining Accuracy in the Forensic Sciences (August 1, 2016). Available at SSRN: http://ssrn.com/abstract=2773255 or http://dx.doi.org/10.2139/ssrn.2773255

[15] *Id*.

medical testing. [16] CLIA requires medical laboratories to participate in routine proficiency testing using samples supplied by designated test providers. The proficiency test samples must be tested in the same manner as patient specimens, at the same time as patient specimens, by the same personnel who routinely test the patient specimens, and using the same test system that is routinely used for the patient specimens.[17] Feedback from testing has helped laboratory managers identify weaknesses in laboratory systems, identify weaknesses in training or preparation of staff, and detect problems with equipment and procedures. In light of the vital role testing has played in both quality management and quality improvement in medical laboratories,[18] it seems likely that a similar program of testing could prove helpful in forensic science. [19]

At the present time, most forensic laboratories require analysts to take periodic proficiency tests. These tests are valuable for assuring that analysts have the minimal level of competency needed to perform at a satisfactory level, but current proficiency testing programs do not achieve many of the important benefits that could be achieved with more comprehensive research programs of the type discussed here. Proficiency tests involve relatively few samples; analysts know they are being tested (which may cause them to perform differently during proficiency tests than when performing casework); and the tests are typically designed to be relatively easy for a competent analyst to pass.[20] Hence, proficiency tests have limited value for establishing the limits of reliability and accuracy that analytic methods can be expected to achieve as the conditions of forensic evidence vary.[21] These tests provide little useful feedback to forensic analysts on the limits of their expertise when dealing with difficult cases or marginal evidence and hence have little value for helping experienced analysts improve their skills. They are not designed to determine error rates.[22] And because proficiency tests are designed to test the competency of

---

[16] Westgard, J. O., & Westgard, S. A. (2006). The quality of laboratory testing today: An assessment of metrics for analytic quality using performance data from proficiency testing surveys and the CLIA criteria for acceptable performance. *American Journal of Clinical Pathology*, 125(3), 343–354. http://dx.doi.org/10.1309/V50H4FRVVWX12C79

[17] A convenient summary of CLIA testing requirements can be found in a brochure prepared by the Center for Medicare and Medicaid Services: https://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/Downloads/CLIAbrochure8.pdf

[18] Westgard, J. O., & Westgard, S.A. (2006), op cit.

[19] Forensic laboratories often need to deal with a broader range and quality of samples than medical laboratories, but that does not diminish the need for research on laboratory performance. Indeed, the challenge of dealing with a variety of sample types may makes research on laboratory performance even more important in forensic laboratories.

[20] The president of Collaborative Testing Services, an organization that provides test samples that are widely used for proficiency testing in forensic laboratories, told the Commission during its seventh meeting (August 10, 2015) that he has been under commercial pressure to make proficiency tests easier.

[21] Proficiency tests are all too often analogous to the driving tests one must pass to obtain a driver's license. They test whether drivers possess a minimal level of competency, not how well they can drive under challenging conditions. The typical test required to obtain a driver's license would be useless for assessing how well drivers perform on a high-speed race track, or when it would be safe for a driver to negotiate icy mountain roads. These tests serve a valuable function by weeding out truly incompetent drivers, but they do not provide typical drivers the kind of feedback they need to improve their skills, nor do they provide insights into when, due to challenging conditions, it becomes unsafe even for a competent operator to drive.

[22] Indeed, Collaborative Testing Service specifically warns that its proficiency tests were not designed and should not be used for error rate estimation. Collaborative Testing Services, Inc. CTS Statement on the Use of Proficiency Testing Data for Error Rate Determinations, March 30, 2010, pg. 3,

individuals, rather than explore variables that may affect accuracy, performance on these tests provides little insight into the circumstances in which accuracy may be impaired.[23] Thus, although proficiency testing serves an important purpose, it is no substitute for the kind of research programs discussed here.

Challenges in Implementing Research on Laboratory Performance

(1) Developing Suitable Sets of Research Samples

To implement rigorous research programs of the type discussed here, forensic laboratories will need access to suitable sets of evidentiary specimens for which ground truth regarding origin is known. To study the performance of latent print examiners, for example, laboratory managers will need access to latent prints of known origin that can be inserted into routine casework and compared with same-source and different-source exemplar prints. Similarly, to study the performance of DNA analysts, laboratory managers will need known-source biological specimens, and to study firearms examiners, they will need known-source bullets and shell casings.

The research sample sets should be designed with the help of forensic science practitioners, statisticians, and experts on research methodology to assure that the studies in which they are used are methodologically sound and address important questions about laboratory performance. Different test sets will be needed according to the goals of the study. To achieve some goals (e.g., establishing limits of validity; providing feedback on performance in difficult cases), the research sets would need to be extremely challenging; for other goals, the research sets would need to replicate routine casework.[24]

---

http://www.ctsforensics.com/assets/news/CTSErrorRateStatement.pdf ("The design of an error rate study would differ considerably from the design of a proficiency test. Therefore, the results found in CTS' Summary Reports should not be used to determine forensic science discipline error rates").

[23] For example, there were a startling number of errors on the most recent CTS proficiency test of latent print examiners, but the design of the test makes it difficult to ascertain what went wrong and hence reveals little about the circumstances under which such errors are likely (and unlikely) to occur. Four hundred thirty-one latent print examiners took this test, which required that they compare 12 latent prints with reference prints from 4 individuals. Thirty seven of the participants incorrectly identified at least one of the latent prints as coming from the wrong finger (there were 51 false identifications in all); 11 participants mistakenly excluded a finger that was the true source of a latent print, and 4 of those also mistakenly identified it as coming from a different finger. Collaborative Testing Services, Inc. Forensic Testing Program, Latent Print Examination Test No. 16-515/516 Summary Report. May 11, 2016.

[24] Tests sets could be designed to examine a variety of important issues. For example, an important issue in latent print analysis is the ability of analysts to deal with distortions of latent prints that might arise from pressure, torsion, double-taps, and other mechanisms. Test sets of known-source latent prints that incorporate various types and degrees of distortion would, when introduced into the flow of casework, help laboratories see how well analysts recognize and deal with distortion. Research on examiners' performance with such specimens could help identify conditions under which a particular kind or degree of distortion can lead to inaccuracy. It could also provide invaluable feedback to analysts on their performance when dealing with the challenge of analyzing distorted latent prints, helping them to improve their skills.

Development of suitable sets of research samples is a time-consuming and expensive task that will exceed the resources of many forensic laboratories. Laboratories may be able to cooperate to share this burden. Known-source latent print images, for example, could be prepared by one laboratory and shared electronically with other labs—creating efficiencies through inter-laboratory cooperation. Nevertheless, it is unrealistic to expect forensic laboratories themselves to bear the entire burden of creating such samples. Assistance from governmental agencies is needed.

It is the view of the Commission that a government agency, such as the National Institute of Standards and Technology (NIST), should play a leading role in creating test sets for research on laboratory performance. This is a function that will be most efficient if handled in a centralized manner by an agency with expertise in testing. NIST has made valuable contributions to forensic DNA testing by providing mixed biological samples to laboratories for proficiency testing. In the view of the Commission, it would be desirable for NIST to expand its efforts in this arena to include the creation of test sets for other types of research on laboratory performance. NIST (and other government agencies) should also consider funding the creation of research test sets by private vendors and research organizations.

(2) Research on the Accuracy of Database Searches

In some forensic science disciplines, the processing of blind research samples may entail searching for matching samples in local, state, or national databases. For example, a DNA analyst might search various government DNA databases looking for profiles that match the profile of a research sample. A latent print analyst might search databases of fingerprints looking for a match to a latent print that was submitted as a blind research sample. Allowing analysts to conduct such searches will provide valuable feedback on the overall operation of database systems. If a same-source reference sample is present in the database, these searches will test how often it is found and provide insight into circumstances under which same-source reference samples are missed or are given a low ranking by search algorithms. These searches will also provide information on the risk of finding nonsource reference samples that are similar enough to be misidentified as potential sources. At present, very little information is available about the sensitivity and specificity of identification of samples from databases *as those operations are performed in actual casework*. Feedback on this issue from research studies will provide valuable insights that can be used to make these systems more effective and efficient.
Before database systems are studied in this manner, it may be necessary to change some of the rules that currently govern the purposes for which crime laboratories may gain access to governmental databases.[25] These rules (as currently written) may not authorize forensic laboratories to access databases in the manner described here for the purpose of quality assurance and research, or may create uncertainty about whether databases may be used in this way. Crime laboratories will also need to take steps to assure that specimens created for research purposes are not permanently entered into databases in a manner that might create the false impression that these samples are associated with a crime scene. For example, laboratories might require that entry of samples into a database be authorized by section managers or quality

[25] These rules are found in a variety of sources, including the state and federal legislation that authorized the creation of the databases and in memoranda of understanding (MOU's) between agencies that operate the databases (e.g., the FBI) and the laboratories and law enforcement agencies that are granted access to the databases.

assurance personnel who know which samples are from real crimes and which are research specimens.[26]

It is the Commission's view that any rules regarding database access that preclude the kind of quality assurance testing program discussed here should be changed to allow such testing to proceed. Because the key rules at issue are contained in memoranda of understanding between the FBI and the various agencies it serves, government agencies should identify and revise any language or agreements that prohibit, or appear to prohibit, access to these databases for the kind of performance and quality assurance testing discussed herein.

(3) Avoiding Misuse of Research Findings

For research on laboratory performance to achieve its goals, some of the research samples must be sufficiently challenging to induce errors. One cannot successfully study the strengths and limitations of a system without knowing the circumstances under which the system fails. To identify the boundary conditions within which forensic assessments are accurate, and beyond which accuracy suffers, it is necessary to test performance under marginal conditions—to push the boundaries until accuracy drops off. By analogy, if one wishes to study the strength of various types of chain, one must stress the chains until they break. Little of value will be learned from a study of chain strength in which the chains rarely if ever break. In this context, a low rate of failure indicates an inadequate study, and frequent failure is the hallmark of a successful, informative study.

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. Practitioners should not be punished or sanctioned for making errors in such studies.

It should also be recognized that the rate of error when processing challenging research samples may have little relevance for predicting the rate of error in more routine forensic tests on samples that are easier to analyze. The rate of error when analysts are pushing the boundaries of their expertise tells us little or nothing about the probability of error when they are well within those boundaries. Performance is almost certain to be better in easy cases than in hard cases. Forensic laboratories may be reluctant to undertake research of the type discussed here if errors that occur during the studies are used unfairly to impugn the reputation of the laboratory or

[26] This step would not prevent analysts from searching samples against a database while remaining blind to the source of the samples. It is possible to search both latent fingerprints and DNA profiles against government databases without making the item a permanent part of the database. Latent prints, for example, must be registered before they become a permanent part of the FBI's fingerprint database, which was formerly known as the Integrated Automated Fingerprint Identification System, or IAFIS, and is currently known as the Next Generation Identification System. Laboratories can search latent prints against the database without "registering" them; and prints that are "registered" can later be "unregistered," which results in their removal for the system.

examiners. Accordingly, the Commission urges state and federal judges to consider carefully circumstances under which information about errors in research studies should be admissible in the courtroom. Although the results of such research will be valuable and enlightening on a number of important issues, it would be misleading to equate the rate of error on research samples designed to be highly challenging with the rate of error for cases in general or with the probability of error in a specific case, particularly if the case involved relatively easy or straightforward analysis. Consequently, if the results of performance testing are admitted as evidence in the courtroom, it should only be under narrow circumstances, and with careful explanation of the limitations of such data for establishing the probability of error in a given case.[27]

(4) Funding of Pilot Programs

To facilitate development of research programs of the type described here, funding should be provided 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. Funding agencies should both support trial efforts and provide incentives to encourage laboratories to undertake these efforts.

[27] The Commission recognizes that the results of performance testing may fall within the government's disclosure obligations under *Brady v Maryland, 373 U.S. 83 (196*3). But the right of defendants to examine such evidence does not entail a right to present it in the courtroom in a misleading manner. The Commission is urging that courts give careful consideration to when and how the results of performance testing are admitted in evidence, not that courts deny defendants access to evidence that they have a constitutional right to review.