J Forensic Sci, Jan. 2003, Vol. 48, No. 1
Paper ID JFS2002042_481
Available online at: www.astm.org

Joseph L. Peterson,[1] D. Crim.; George Lin,[2] M.S.; Monica Ho,[3] M.A.; Yingyu Chen,[1] M.A.; and R. E. Gaensslen,[1] Ph.D.

The Feasibility of External Blind DNA Proficiency Testing. I. Background and Findings*

ABSTRACT: We describe the origins, purposes, and findings of a national study to determine whether a large-scale program of blind proficiency testing in U.S. DNA laboratories is feasible and/or practical. Proficiency testing in clinical laboratories is reviewed, particularly as mandated by the Clinical Laboratory Improvement Acts and its role in the regulation of those laboratories. Proficiency testing in forensic urine drug testing labs is also briefly reviewed. Studies involving comparisons between open and blind proficiency testing are discussed. The clinical laboratory proficiency testing and regulation landscape provides the background for the DNA Act of 1994, and the congressional mandate to investigate blind proficiency testing in forensic DNA laboratories. Four models of blind proficiency testing are defined and discussed, along with the advantages and disadvantages of each and estimates of the costs of a large-scale program. The purposes of proficiency testing in a quality-assurance context are likewise discussed and related to the models and the arguments generally proffered for and against blind vs. open proficiency testing.

KEYWORDS: forensic science, DNA, proficiency testing, blind proficiency testing, quality assurance, quality control, DNA Act, Clinical Laboratory Improvement Act

Forensic DNA typing has emerged as an important and powerful tool in criminal identification and intelligence. DNA technology was first introduced into a criminal court in the United States in 1986 and has since become widely accepted by the criminal justice system and the courts. As with any technique or technology perceived as novel, concerns have periodically been raised in admissibility hearings about the validity and reliability of DNA analysis, and the extent to which quality assurance measures are adequate. The development and evolution of DNA technologies through three "generations" (RFLP, PCR—dot blot, and PCR STR) has prompted additional judicial inquiry as each new technique was introduced. In addition, the development of DNA technologies overlapped with the Supreme Court's Daubert vs. Merrell Dow Pharmaceuticals (1) decision, allowing the admissibility of DNA evidence to be tested under the new standard in jurisdictions that have adopted it perforce or voluntarily.

In this communication, we review the highlights of the development of proficiency testing as quality-assurance (QA) and quality-control (QC) tools in the clinical laboratory and in forensic science laboratories. QA measures in forensic DNA laboratories are reviewed, leading up to the DNA Act of 1994 and the congressional mandate for the present blind proficiency-testing feasibility project. Our analysis of proficiency testing in general, and of blind proficiency testing in particular, developed from the literature and from laboratory survey data, is summarized, and suggested as a conceptual framework for further discussion. We also review the major points of discussion and conclusions reached by this project's advisory board in making recommendations to the National Institute of Justice (NIJ) and, in turn, to Congress. A follow-up paper presents and discusses our experience with the construction and administration of actual blind proficiency tests on a smaller, experimental scale (2).

The overall project research was separated into two sequential phases, each comprising roughly two years. Emphasis in Phase 1 was on the comprehensive literature survey and the survey of forensic DNA laboratories and their practices. In Phase 2, there was more emphasis on the extent to which practices might already be in place that could provide the same information as blind tests, and on exploring less costly alternatives, such as audits and reanalysis. Different series of blind proficiency tests were set up and executed in each phase.

National Forensic DNA Review Panel

Upon providing the first two years of funding for this project, NIJ set up a national advisory panel to oversee the project, and to formulate recommendations on blind proficiency testing in DNA laboratories based on the project's findings. In the DNA Act of 1994 (3), the Director of NIJ was directed to certify to the joint Congressional Committees on the Judiciary within a year of the law's effective date that: (a) A national blind proficiency testing program was in operation; or (b) Such a program was not feasible; or (c) That a project was underway to establish such a program within two years of enactment. The Director of NIJ wished to have the advice of this advisory panel before making his certification to

[1] Professor and doctoral student, Department of Criminal Justice, and Professor, Forensic Science Group, Department of Biopharmaceutical Sciences, respectively, University of Illinois at Chicago, Chicago, IL.
[2] DNA Analyst, DOD DNA Identification Laboratory, Armed Forces Institute of Pathology, Rockville, MD.
[3] Doctoral student, Department of Sociology, Columbia University, New York, NY.

* This work was supported by Grant No. 96-DN-VX-0001 from the National Institute of Justice, Office of Justice Programs, U. S. Department of Justice. The opinions expressed and the conclusions reached are those of the authors, and do not necessarily reflect the views or official positions of the National Institute of Justice or the U.S. Department of Justice.

Received 6 Feb. 2002, and in revised form 21 June 2002; accepted 29 June 2002; published 13 Nov. 2002.

Congress. The panel consisted of twenty members, including the then-sitting DNA Advisory Board (DAB), plus representatives of the ABA Committee on Science and Technology, National District Attorneys Association, Home Office Forensic Science Service (UK), DHHS Registry Unit/CLIA, College of American Pathologists, and National Association of Criminal Defense Lawyers. The chairperson of the DAB acted as chair of the advisory panel. The panel met three times before formulating its recommendation to the NIJ Director. At that time, it had the benefit of the results of approximately the first year and a half of the project. The panel met a final time to review progress made during the second phase of the project and to consider whether it wished to change or amend its original recommendations to the NIJ Director.

Quality Assurance/Quality Control in Forensic DNA Testing Laboratories

Since DNA testing was introduced as an important laboratory technique used in the identification of persons in criminal cases, the forensic community has paid much attention to this area and invested considerable resources in it. Part of the effort has included trying to ensure the quality, integrity, and reliability of the DNA testing results. Open (declared) proficiency testing has always been one of the linchpins in the QA programs. Blind proficiency testing as a QA tool was raised during the congressional hearings leading up to passage of the DNA Act. Although it is practiced in a few labs and encouraged by various guidelines, it has not become a requirement for forensic DNA laboratories.

This project began with a detailed literature review to summarize and document the history of proficiency testing efforts, including significant regulatory and scientific landmarks that have placed proficiency testing as the centerpiece of most quality assurance programs within the clinical, medical, forensic urine drug testing, criminalistics and biological evidence analysis areas.

Rigorous QA and QC standards were established by consensus among laboratory analysts and oversight agencies (4). The first applications of DNA testing to disputed parentage and forensic identification were conducted by private, commercial laboratories, and they were the first to establish QA/QC programs for forensic DNA testing (4).

The first published set of quality assurance standards for DNA testing were those of the American Association of Blood Banks (AABB) (5). These standards were created for laboratories performing restriction fragment length polymorphism (RFLP) analysis in parentage tests and were recently extended to include several key provisions for polymerase chain reaction (PCR) based testing. The Technical Working Group on DNA Analysis Methods (TWGDAM), created in 1988, first published its "Guidelines for a Quality Assurance Program for DNA Analysis" in 1991 for the forensic science DNA typing laboratory community. The guidelines have undergone several revisions before the current version appeared in April 1995 (6).

The DNA Identification Act of 1994 established a framework for setting standards on quality assurance and proficiency testing in forensic DNA typing laboratories. The law created the DNA Advisory Board (DAB) whose members were appointed by the Director of the FBI. The law states, "The advisory board shall develop, and if appropriate, periodically revise, recommended standards for quality assurance, including standards for testing the proficiency of forensic laboratories, and forensic analysts, in conducting analyses of DNA" (3). In 1998, the National Institute of Justice created the National Commission on the Future of DNA Evidence. The group's objective is to develop policies that will maximize the value of DNA in the criminal justice system.

Development of QA/QC guidelines for forensic laboratories grew out of earlier experience with QA/QC guidelines and federal regulation of the clinical laboratories and of forensic urine drug testing laboratories.

Clinical Laboratory Regulation and Proficiency Testing

Currently, there are two federal regulatory programs for clinical laboratories, both administered by the Health Care Financing Administration (HCFA) of the Department of Health and Human Services (DHHS). The two programs are Medicare and Medicaid certification of facilities receiving reimbursement under these programs, and licensure of all clinical laboratories under the Clinical Laboratory Improvement Act of 1988 (CLIA '88). Additionally, the Food and Drug Administration (FDA) is another agency responsible for licensure and registration of facilities preparing, collecting, and shipping blood and blood products and for the approval of medical devices (7).

Medicare/Medicaid Regulatory Program

Federal authority over clinical laboratories began with the passage of the Social Security Act of 1965, which established a system for the payment of benefits for medical care for several categories of individuals, including the aged, financially needy, dependent children, and the disabled. The Health Care Financing Administration (HCFA) is primarily responsible for the administration of the Medicare program and for the provision of assistance to the states for the administration of the Medicaid program. The Medicare regulatory programs are based on standards developed by the Secretary of DHHS to assure the health and well-being of individuals to whom healthcare is being provided in a variety of inpatient and ambulatory settings, and include clinical laboratory testing.

Clinical Laboratory Improvement Act of 1967

The Clinical Laboratory Improvement Act of 1967 (CLIA '67), or the interstate licensure program, originated with the passage of the Partnership for Health Amendments (8) and was based on the decision by Congress to assure the quality of testing performed on specimens in the course of interstate commerce. CLIA '67 adopted the Medicare personnel standards and added quality control and proficiency testing standards (9). Due to the testimony of the Secretary of Health, Education, and Welfare in 1967 that a 25% error rate was common among clinical laboratories (10), some clinicians have contended that CLIA '67 was enacted mainly in response to misrepresentations of poor laboratory performance to legislators and to the public.

CLIA '67 required proficiency testing for all governmental agencies and private sector organizations concerned with laboratory regulation and accreditation (11). The Centers for Disease Control (CDC) was given the responsibility for the implementation and administration of the programs within the Public Health Service (PHS). Interstate laboratories were required to enroll and successfully participate in CDC's proficiency testing program, and additional standards on internal quality control, personnel, and record keeping were expected to be developed (9). The noteworthy difference in the regulatory process between CLIA '67 and the Medicare statute is successful participation in the CDC proficiency testing program. Unlike the Medicare programs, which had neither federal grading criteria in the regulations nor a definition of what constituted successful performance in proficiency testing for individual analytes or organisms, the specific grading scheme in the

regulations for the CDC PT program had approved several state and private sector programs in the mid-1980s.

In 1979, the PHS and HCFA signed an agreement that consolidated the administration of the Medicare and CLIA laboratory programs within HCFA. HCFA was responsible for the survey and certification and/or licensure of all clinical laboratories in both programs. More recently, the HCFA was given responsibilities for developing new regulations and the CDC assumed the responsibility for technical input and developing advances in proficiency testing. However, during this time, Medicare laboratories were not subject to the same national proficiency-testing program as laboratories under CLIA '67. In addition, the HCFA had not established minimally acceptable requirements for program content, challenges, frequency of test events, and grading criteria.

Clinical Laboratory Improvement Act of 1988

In order to respond to the problems existing in the Medicare regulatory programs and CLIA '67, the HCFA, CDC, state health officials, various private-sector organizations, concerned members of the laboratory industry, and the public decided to consolidate all of the CLIA '67 and Medicare/Medicaid laboratory requirements. CLIA '88 (12), which was based on four principles, namely personnel standards, quality control, quality assurance, and proficiency testing, was thus enacted in October 1988. A new set of guidelines, the so-called "March 14, 1990 final rules," consistent with the standards established under CLIA '88, was adopted on March 14, 1990 (13). The new guidelines explicitly defined grading practices and what constituted acceptable laboratory performance.

CLIA '88 mandates proficiency testing in all clinical laboratories (13). CLIA-certified and Medicare-approved laboratories are required to enroll in DHHS-approved proficiency testing programs for each specialty and subspecialty of service for which they seek certification (14). Generally, proficiency-testing programs provide five samples for each analyte or test three times per year. With few exceptions, the passing score is 80%. If a laboratory receives a failing score on a PT, the laboratory must take necessary actions to find, correct, and document any problems occurring in the test performance. Compared with Medicare/Medicaid regulatory programs and CLIA '67, new rules under CLIA '88 about proficiency testing are more explicit. The comparison among Medicare/Medicaid programs, CLIA '67, and CLIA '88 is shown in Table 1.

TABLE 1—*Comparison of clinical regulatory programs.*

Medicare/ Medicaid Programs	CLIA* 1967	CLIA 1988
• Designed for laboratories receiving Federal reimbursement for laboratory tests • No consistent grading of PT† (varied from state to state) • Licensure by specialty/subspecialty	• Designed for interstate commerce laboratories • PT grading criteria is standardized nationally • PT surveys are the basis of licensure	• Regulates all clinical laboratories with few exceptions • More explicit guidelines regarding PT • Licensure by individual tests, analyte, specialty and subspecialty • Criteria developed for PT providers • Standards developed for newer specialties

* CLIA, Clinical Laboratory Improvement Act.
† PT, proficiency testing.

Currently, requirements under CLIA '88 emphasize the increased importance of evaluating and achieving a passing score on specimens of known content, which are intended to be tested as if they were patients' samples, and serve as a measure of laboratory quality. With the passage of CLIA '88, standards were also developed in newer specialties such as cytogenetics, DNA probes, molecular genetics, and standards for other areas ranging from histocompatibility testing to cytology were updated. CLIA '88 requirements apply to all 150,000 hospital, reference, physical, and clinical laboratories in the U.S.

Quality Assurance/Quality Control and Proficiency Testing in Clinical Laboratories

The first proficiency-testing program can be traced back to 1946 when Sunderman distributed anonymous specimens to hospital laboratories to assess laboratory performance and to standardize results. The results varied widely among laboratories (15). The quality of work performed in those laboratories and the causes of analytical discrepancies existing among them were therefore examined. During the 1940s, the College of American Pathologists (CAP) was founded and instituted the first national proficiency survey called the Standard Solutions and Materials Program (16), which was similar to the one conducted by Belk and Sunderman. CAP organized, promoted, and further mandated proficiency testing as a criterion for laboratory accreditation, because PT programs were viewed primarily as a mechanism for a continuous, incremental improvement process.

By the late 1940s and early 1950s, compulsory PT was required by some professional societies, as well as some state and municipal governments. In the 1960s, proficiency testing had become a standard practice in clinical laboratories. Simultaneously, internal QC programs began to play an increasingly important role as laboratories perceived a need to insure accuracy in their analytical performance. The first program of this type was developed in 1967 by Joseph A. Peterson and involved 30 cooperating laboratories in the Colorado area (17). The serum pools were shared among many laboratories and the results were statistically reviewed by various organizations to determine the limits of acceptable performance (18). In 1969, Skendzel et al. (19) conducted a study that analyzed CAP surveys on laboratory performance over the past 6 years, and discovered that the coefficients of variation had narrowed by 50% or more for all the analytes Belk and Sunderman had studied except cholesterol.

With the passage of CLIA '67 and later CLIA '88, proficiency testing became mandatory in all clinical laboratories due to the belief that participating in PT could ensure the quality results (20). In fact, it has been demonstrated that mandated proficiency testing enhances overall quality of clinical laboratory testing, including turnaround time, accuracy of results, and training of laboratories, whereas self regulation has been found less effective in achieving these goals (21). Although CLIA's critics suggested that the mandated quality standards for PT might lead to a higher incidence of failed laboratories, it has not proven to be true; on the contrary, the PT performance data has strongly indicated that the overall quality of laboratories improved, which suggested that CLIA's mandated quality performance and standards for PT were achievable (22).

Proficiency Testing in Forensic DNA Laboratories

Testing the validity and reliability of scientific test results in the crime laboratories is as important as that in the clinical laboratories. Until the mid 1970s, however, there were virtually no procedures

for that purpose. In 1974, the Forensic Sciences Foundation (FSF) conducted a study on developing a proficiency testing program for crime laboratories. That study found serious problems in the examination and interpretation of results for several types of specimens (23). A combination of greater resources to these laboratories, improved education and training opportunities, implementation of accreditation and certification programs, as well as proficiency testing and quality assurance programs were therefore suggested (24).

The British Forensic Science Service (FSS) has long been cited for its demanding quality assurance standards, including proficiency testing, beginning in 1969. Margaret Pereira has noted that FSS's quality assurance programs included both 1) open (declared) samples, and 2) blind trials that enter laboratories disguised as genuine cases (25). Although blind trials are much more difficult to construct, she comments their advantage is that they test "the whole system," from receipt of evidence and quality of scientific work to the laboratory report as well as the time required to complete the case. At the Banbury Conference in 1988, David Werrett of the Home Office Central Research Establishment said that the British had established undeclared DNA trials (26).

Proficiency Testing Standards

The ASCLD Laboratory Accreditation Board (ASCLD-LAB) requires laboratories seeking accreditation to establish and maintain a "quality system." Proficiency testing is cited as an "integral component" of QA programs and requires laboratories to subscribe to an external proficiency test provider if they seek to gain and retain accreditation (27). Recognizing the importance of proficiency testing in forensic laboratory quality assurance, the TWGDAM "Guidelines for a Quality Assurance Program for DNA Analysis" devotes an entire section to proficiency testing, stating that "Participation in a proficiency testing program is a critical element of a successful QA program and is an essential requirement for any laboratory performing forensic DNA analysis" (6). It also mentions that it is "highly desirable" for the DNA laboratories to participate in a blind proficiency testing program that "realistically simulates" actual casework (6).

As a result of the increasing number of quality assurance programs requiring proficiency testing, the TWGDAM Quality Assurance Subcommittee joined with the DNA Proficiency Review Committee (PRC) of ASCLD-LAB to produce *Guidelines for DNA Proficiency Test Manufacturing and Reporting* (1994) to set standards for commercial providers of DNA test samples (28). These guidelines also set standards for the personnel, facilities, and procedures used by the manufacturers, along with quality control procedures they have to follow in manufacturing PT specimens. Recently, ASCLD announced those PT providers that have been approved to service ASCLD-LAB accredited laboratories. Additionally, the American Board of Criminalists (ABC) certifies individuals based upon their educational background, experience, and performance on a written examination (29).

Purpose of Proficiency Testing

The Office of Technology Assessment (OTA) issued a report in 1990 which strongly endorsed the types of DNA testing that were being used in forensic laboratories and declared, "the forensic uses of DNA tests are both reliable and valid when properly performed and analyzed by skilled personnel" (30). *DNA Technology in Forensic Science* (often called "NRC I"), issued by the National Research Council (NRC) in 1992, agreed with the OTA findings. NRC recommended regularly scheduled proficiency testing as a way of measuring laboratory error rates and evaluating whether and how laboratories have taken corrective action to reduce errors (31). A few years later in 1996, the NRC in a report entitled *The Evaluation of Forensic DNA Evidence* (often called "NRC II"), changed its stance on the goal of proficiency testing, stating that it ". . . is not designed to measure error rates," but ". . .is one of the best ways of ensuring standards and. . .should be used to improve laboratory performance by identifying problems that need to be corrected" (32).

Recent Commercial Proficiency Testing

The DNA Identification Act of 1994 (3) established a federal framework for setting standards on quality assurance and proficiency testing. The law created the DNA Advisory Board (DAB) whose members were appointed by the FBI Director. The board developed a set of standards for DNA testing that replaced the then-existing TWGDAM Guidelines. Later, when the DAB ceased operation (as provided for in the law), TWGDAM was resurrected as the QA/QC guideline-setting entity, and the then-current DAB guidelines again became TWGDAM guidelines.

Currently, DNA proficiency test trials are available from Collaborative Testing Services (CTS), the College of American Pathologists (CAP), the Serological Research Institute (SERI), Cellmark Diagnostic's International Quality Assurance Survey (IQAS), and the Spanish and Portuguese Working Group (GEP) of the International Society for Forensic Genetics (ISFG).

In 1987–1988, the California Association of Crime Laboratory Directors organized proficiency trials, which used simulated DNA evidence samples, for three commercial facilities (33). The American Association of Blood Banking (AABB) also started a DNA proficiency-testing program by adding a DNA module to their 1991 Parentage Specimen Program (PSP). Collaborative Testing Services (CTS), in conjunction with the American Society of Crime Laboratory Directors and the Forensic Sciences Foundation, began one of the first DNA PT programs by adding a DNA module to its physiological fluids offerings within its forensic laboratory testing program in 1991. In 1993, the College of American Pathologists started proficiency testing for both the forensic and parentage laboratories under its forensic identification survey, beginning with 41 participants and growing to 80 within a year (34).

Generally, each sample pack consists of bloodstains and/or semen stains of which there is a "crime scene" stain and a combination of suspect and victim stains. The objective of these proficiency tests for forensic DNA laboratories is to correctly include or exclude suspect/victim stains from the crime scene stain. Each of these proficiency tests allows its participants to report information pertaining to methodology, band-sizing data from RFLP analysis, and discrete typing results from PCR-based DNA testing. In February 2000, the Bureau of Justice Statistics published survey results based on replies from 108 forensic laboratories performing DNA testing in the U.S., and found that while some laboratories only required proficiency tests once a year, most required the tests every six months (35).

Factors Relating to Proficiency Test Performance

Duration of Participation

A long accepted quality assurance maxim holds that "anything improves if you measure it" (36). Hansell and Haven (37) first showed long-term improvements in interlaboratory agreement in

the CAP Ligand Assay surveys from 1972–1978. In Data ReCAP, 1970–1980, researchers for the CAP showed that interlaboratory agreement improved markedly for most analytes over time (38). More recently, data was examined from CAP surveys from 1987 to 1993 in the areas of chemistry, hematology, immunology, and blood banking, and was found that laboratories with consistent participation show consistent and statistically significant improvement in performance for the first 3–4 years of proficiency testing (39).

Personnel Qualifications

A review of personnel standards was conducted in 1996 by Peddecord et al. (40) in which the relationship between laboratory personnel regulations and laboratory performance was examined. By utilizing proficiency test results as a measure of laboratory performance relative to personnel regulations, it was determined that better PT results were usually associated with higher personnel qualifications.

Laboratory Environment

Stull et al. (41) observed the aggregate rates of satisfactory event performance for all regulated analytes, tests, and specialties were 97% in hospital and independent laboratories, and 91% in all other testing sites. Better PT performance has also been positively related to increased test volume for certain analytes such as cholesterol (42) and glucose (43). Shehangian (44) determined that increased institutional size and laboratory workload have also been generally related to improved PT performance and less variation in chemistry, bacteriology, parasitology, and qualitative hematology.

Testing Methodology and Automation

Another factor that should not be ignored is advances in testing methodology and automation. In physician office laboratories, automation was related to increased precision and reduction in error rate by a factor of 1.5–3 (45).

Quality Control Procedures

Positive relationships between better quality control practices and better proficiency test performance are also significant. Lawson et al. (46) showed that PT results are related to measures of performance in a laboratory's quality control system. Two recent studies (45,47) have concluded that improvement of laboratory performance was not the direct result of the PT process itself, but primarily due to two factors: (a) extensive education that was a key component of the larger QA/QC program, and (b) voluntary withdrawal from testing by laboratories displaying poor performance.

Limitations of Proficiency Testing

Although an empirical study (44) showed that there is a positive relationship between PT performance and other putative quality indicators of laboratory performance, there are limitations to the usefulness of PT data. These limitations include: (a) incomplete testing of the total testing process (TTP), (b) special treatment of PT materials, (c) the "matrix effect," and (d) how PT performance criteria are used.

Incomplete Testing of the Total Testing Process

Due to proficiency testing materials originating from a different source than patient specimens, PT samples enter the testing process at the pre-analytical phase of the total testing process (TTP) rather than at the beginning of the TTP. Therefore, open proficiency testing only assesses the analytical stage of the TTP.

Special Treatment of Proficiency Test Materials

A survey conducted by Cembrowski and Vanderlinde (48) found that various practices were used by laboratories to improve performance on PT specimens, including replicate analysis, sending the PT sample to a designated analyst, analyzing PT specimens immediately after standardization and quality control, and delaying analysis until the analytic process was optimal. Therefore, the results from a proficiency test may be not truly representative of routine performance of a laboratory. However, it should be noted that these issues have been addressed under CLIA '88 guidelines—laboratories must attest to the fact that no special treatment is given to proficiency test samples.

Matrix Effect

Proficiency test specimens are typically manufactured samples that simulate patient specimens; due to their dissimilarity, PT results can be difficult to evaluate and control (49). Specifically, factors related to error are the confounding effects of "fluid-matrix" caused bias, method instrument bias, and deviations from methods associated with analyzing PT specimens. Proficiency test specimens are suspended in solutions (a "fluid-matrix") to approximate clinical and biological conditions. Clinicians have observed a "matrix effect" in which the fluid-matrix may destabilize the PT specimens over time and/or cause interference in instrument readings. The true value of the proficiency testing may, therefore, be biased.

Proficiency Test Performance Criteria

Although proficiency testing has limitations mentioned above, it is still important to establish true values to ensure the basis of uniform standards in clinical laboratories. In response to these potential problems, the clinical laboratory community has developed four basic means by which proficiency test true values may be established: (a) consensus values or peer group statistics after appropriate outlier exclusion, submitted by participating laboratories, (b) analysis of specimens by definitive methods or protocols correlated to definitive methods, (c) referee laboratories, and (d) documentation of the composition of the specimen by design and method of manufacture by the manufacturer.

One of the disadvantages of using peer group statistics is that systematic or random errors specific to the methodology would not be taken into account (50). In addition, it is possible that laboratories might standardize and calibrate on a biased consensus result. Therefore, in order to produce a more accurate assessment of the overall quality of laboratory testing in an interlaboratory survey, a proficiency-testing program should involve participation by a wide spectrum of laboratories representing all levels of performance.

The second means, analysis of specimens by definitive methods or protocols correlated to definitive methods, can also be problematic. The National Bureau of Standards (NBS), now known as the National Institute of Standards and Technology (NIST), has recognized the need for assigning definitive values, as opposed to consensus values, to the analytes in interlaboratory surveys and has determined the practical analytic goals for accuracy. These definitive values may be determined through the use of exacting protocols, state of the art equipment, and methodologies (51). However, these methods are often slow, tedious, costly and may necessitate development of new techniques (52).

In a study conducted by AuBuchon (53), all the methods for the analysis of alanine aminotransferase (ALT) were compared, and only small differences were found among analytical methods. The NIST has, however, developed definitive methods for use with certain analytes. The number of analytes for which definitive methods are available is small, but for those analytes that do not yet have definitive analytic methods, survey-verified grand consensus mean values often come very close to true values (54–57).

Comparing Blind and Open Proficiency Testing

Proficiency testing (and tests) are called open, or declared, if the specimens are known to be PT specimens, and the laboratory staff know they are being tested. Most proficiency testing in clinical settings is open. It has been said that open PT is not representative of the routine performance of laboratories. Before the passage of CLIA '88, one study found that proficiency test specimens received special treatment (48); another study also documented evident collusion on proficiency tests among physician office laboratories in a small geographic area (58). For these reasons, it has been suggested that if an unknown or blind PT specimen were submitted to a laboratory in the guise of a routine specimen and was not detected, it could not receive special attention. The results of a blind PT would, therefore, be a "truer" measure of laboratory performance.

Urine Drug Testing and Toxicological Analysis

Currently, some proficiency testing programs, such as the Department of Defense's proficiency testing program for forensic urine drug testing (FUDT) (59) and HIV testing, are blind. The first studies were qualitative in nature (60–62). The first comparative analysis of proficiency test results was conducted in 1976 and looked at data from two blind vs. open trials, which occurred in 1973 and 1975 (60). Participating laboratories were given a set of open PT specimens and, simultaneously, were given an identical set of specimens submitted through hospital administrators or physicians and disguised as ordinary patient specimens. The findings showed that the laboratories detected a larger percentage of the drugs in the open samples than in the blind samples, due perhaps to the labs' willingness to report analytes that failed to meet normal cutoffs or perhaps to laboratories giving the PT specimens more attention.

A study in which 13 laboratories were evaluated with blind proficiency test specimens from 1973–1981, compared with CDC open PT test data from 1979–1981, showed that blind PT samples resulted in a lower correct response rate and higher rate of false negatives (61). Another study examined CDC proficiency test data, including blind and open PT, from 1978 to 1980 in the areas of drug monitoring, drugs of abuse, chemistry profile, and blood lead (62). The researchers found that blind PT scores were "27 percentage points lower than the mailed cumulative averages" and that ". . . each laboratory's blind proficiency testing performance was rated unacceptable" (p. 1366, 62).

Quantitative studies have also been conducted using blind samples spiked with various concentrations of analytes (63,64). The results were similar to the original qualitative studies in that an astounding number of laboratories failed to identify a compound or a false identification were reported. Similarly, a large number of reported quantitative results were outside an admittedly arbitrary acceptable range of the target value (2 standard deviations from the mean or coefficient of variation >= 15% from target value).

In 1987, the American Association of Clinical Chemistry (AACC) evaluated the ability of subscribers to the AACC Toxicology Survey Plus program to assess accurately the presence or absence of five drugs of abuse, namely cannabinoids, cocaine, morphine, methamphetamine, and phenylcyclidine (65). The researchers determined that urine drug testing "can produce accurate results." The AACC repeated the study in 1989 by supplying participating laboratories with blind specimens. The blind results were comparable to the open PT results, and the overall accuracy was 97%, the false negative rate was 2.36%, and there were no false positives. Although there was a slightly higher rate of false negatives, the investigators still concluded that urine drug testing "can be accurate."

The results from these studies suggested that laboratories generally performed better when the staff knew they were being tested. It has been postulated that for certain analytes, the employment of less sensitive testing could be the cause of lower correct response rates on blind tests (61,64). If a particular test method is not specified, the laboratory may opt for a less expensive test or method, which may be less sensitive. The desirability of blind proficiency testing is supported by the findings from these that suggest it provides a more realistic test of lab performance on routine specimens.

Clinical Chemistry

Glenn and Hathaway in 1979 examined data from a hospital chemistry laboratory's QC program in which specimens were resubmitted as patient specimens (66). The study not only examined analytical results, but also pre- and post-analytical errors. In contrast to the FUDT PT studies, the study found that analytical results from the blind specimens were comparable to the open samples. However, the researchers found that blind quality assurance was useful in detecting problems in the pre- and post-analytical phases.

Parsons et al. (67) conducted a 3-year observational study that consisted of two phases to assess the performance of open and blind proficiency testing in clinical laboratories conducting blood lead analysis. In Phase 1, 22 certified clinical labs received open PT samples from two providers, whereas in Phase 2, only a single provider distributed PT samples to 24 labs. The blind samples were also introduced to both groups of laboratories at approximately the same time as the declared samples. Researchers found that, in both phases of the testing, only 4.6% of open PT results were unacceptable, whereas 17.7% of blind PT results were unacceptable under CLIA '88 criteria. They suggested that despite the statistically significant differences on the results of open and blind PT, the differences in most cases were "clinically insignificant and would not likely change PT accreditation" (p. 330, 67).

Human Immunodeficiency Virus (HIV) Testing

In 1988, the Department of the Army evaluated HIV testing laboratories that were part of its total quality assurance program (68). Each of the participating laboratories was rated on eight criteria including open and blind testing. The laboratories, additionally, participated in internal and external proficiency panels for a period of 12 months. It was found that blind PT results were nearly as good as open PT results, namely 99.6% correct response rate in blind versus 100% in open.

The first study in which human sera were used was conducted by the CDC and the Association of Schools of Public Health to develop a method for establishing a blind proficiency testing system for HIV testing (69,70). Analytic and nonanalytic issues were examined and the researchers found that test results were of ". . . high accuracy and relatively few errors attributable to laboratory performance." The few analytic errors were all false negatives. The study concluded that blind PT had been most useful in identifying non-

analytic problems and that although blind PT does provide a more valid measurement of routine performance levels, the "complexity and expense limits blind proficiency testing as an external quality assurance tool."

Schalla et al. in 1996 also conducted a blind versus open performance evaluation involving HIV detection (71). The specimens in this study were split into two parts, one as a patient sample (blind PT) and the other was sent to the CDC. The CDC split this specimen three ways: one was sent as an open PT to the same laboratory, another was tested by the CDC which served as the reference laboratory, and the third was frozen and tested in the event that both the open and blind sample tested by the target laboratory disagreed with the reference laboratory result. Of the 6,967 pairs of split specimens, there were 61 (0.88%) discrepancies between the reference laboratory and target laboratory. Of the 25 inaccurate results obtained by the testing laboratories, 14 involved blind samples only, 9 involved open samples, and 2 involved both.

Conclusions about PT in Clinical Laboratories

Proficiency testing is used in laboratories to serve as a mechanism for self-improvement and to assess quality performance. The process involves interlaboratory comparisons of PT data and/or identifying problems that cause error within the laboratory. Numerous studies have shown consistent and statistically significant improvement on proficiency tests in laboratories over time. This may be due to advances in quality control procedures, testing methodologies, and automation or simply the fact that heightened vigilance associated with the PT program itself has caused laboratories to improve their quality in testing. However, open PT focuses primarily on the analytic stage of the testing process; the errors that may occur during pre- and post-analytic process are not captured. In addition, special treatment may be given to those specimens in open proficiency tests.

Insuring the quality of laboratory testing led to the passage of the Clinical Laboratory Improvement Acts (CLIA) of 1967 and 1988 and the DNA Identification Act of 1994, which mandated proficiency testing for all clinical and forensic DNA laboratories, with only a few exceptions. In the regulated clinical laboratories, proficiency testing has become the centerpiece of quality performance measurement. For this reason, it may have become more a regulatory tool and less a means of self-improvement in those labs.

Due to the limitations open proficiency testing has, it has been suggested that blind PT would be a "truer" measure of laboratory performance. The TWGDAM "Guidelines for a QA Program for DNA Analysis" also note that it is "highly desirable" for the DNA laboratory to participate in a blind DNA proficiency testing program annually. In many comparative studies, qualitative results consistently showed that blind PT samples resulted in lower correct response rates and, in particular, higher rates of false negatives, and that compared with open proficiency testing, the improvements in blind PT generally lagged over time.

There are still conflicting opinions on the uses, limitations, and advantages of proficiency testing. Despite the empirically established advantages of blind PT over open PT as a QA tool, the complexity and costs of blind PT prevented either of the CLIA statutes or the regulations derived from them from requiring blind testing.

Approaches to Blind Proficiency Testing in Forensic DNA Laboratories

As noted, the suggestion that blind PT be explored as a QA measure in forensic DNA laboratories grew out of congressional hearings that eventually led to passage of the DNA Act. It is practiced in a few labs, and encouraged by various guidelines, but it has not been made a requirement for forensic DNA laboratories by the law or by any prior subsequent consensus QA guidelines. It was one principal purpose of this project to explore the feasibility of a national forensic DNA blind proficiency testing program.

Blind Proficiency Testing Models

Early in the project, we tried to formulate models of the different ways blind proficiency testing could be implemented. Four models appear to capture every possibility:

1. Blind/LE. In this model, denoted Blind/LE for "Blind/Law Enforcement," no one in the target laboratory (the laboratory that is to be tested) knows anything about the test or when it might be submitted. The blind PT is fully disguised as a routine case, and is submitted to the laboratory by a law enforcement agency in the normal way. This model requires the greatest amount of planning and biospecimen manufacturing effort because the blind test must appear in every way to be a routine case. Success in this model relies on the cooperation of a law enforcement agency willing to carry out the submittal of the "case" and willing to be part of the deception of the target lab.
2. Blind/CL. In this model, denoted Blind/CL for "Blind/Conduit Lab," no one in the target laboratory knows anything about the test or when it might be submitted. This model can be used in circumstances where a target lab is accustomed to receiving DNA cases from other forensic laboratories. One example is centralized DNA testing labs in multi-laboratory systems where some of the system's labs take cases in and do initial processing, but send specimens for DNA typing to a central lab. Another example is independent/commercial labs. This model is easier for the test manufacturer because the target lab is not accustomed to receiving an intact case with all its evidence. Selected specimens, cuttings, and appropriate exemplar specimens are sufficient. Success in this model requires cooperation of a conduit laboratory willing to submit appropriate specimens from the "case" and to be party to the deception of the target lab.
3. Blind Analyst. In this model, only the DNA analysts in the target laboratory need to be kept in the dark about the test and when it may be submitted. Laboratory administrators and/or QA coordinators may be involved in the design, planning and even manufacture of the test. This type of blind PT is practiced in a few laboratories.
4. Random Audit/Reanalysis. This model, often simply called "random reanalysis," consists essentially of blind testing an analyst by re-examining the case and the evidence. The re-examination might consist of a detailed review of all the work (audit), or of an audit as well as a reanalysis of the specimens. The re-examination may be done by another analyst in the same lab, another analyst in the lab system, or an analyst or auditor fully external to the lab. This model is the easiest for the test administrator in that it requires no case or specimen manufacturing. It does require that the case evidence still be available and accessible, and that there be sufficient remaining specimen for re-typing if that is to be a part of the reanalysis. A few laboratories and lab systems practice this type of blind proficiency testing.

There are advantages and drawbacks to each of these models. They vary in complexity and in the estimated cost of a large-scale program. Other considerations in choosing one model over another include how an "acceptable" response will be determined, whether

it is desirable to have responses to the same test materials from a large number of target laboratories to enable comparison of laboratory performance, and the extent to which the CODIS system might be an impediment to large scale testing. There are several ways of defining the "answer" to a proficiency test. DNA testing at present uses loci expressing discrete genetic types, so it should be quite trivial to get the specimens typed by one or two "reference" labs. In the RFLP era, though, there were reasons to want to compare laboratory bandsizing data (72,73). There can be more subtle issues surrounding the definition of an acceptable response. Suppose that most labs have adopted the thirteen core CODIS loci as the basis of their typing protocols, as is the case in much of the U.S. Then assume that a small laboratory, not connected to CODIS, still uses HLA-DQA1 and PM typing to screen its evidence specimens. If that small lab participated in a blind test, and correctly typed the specimens, might it be said that the lab's response was unacceptable because they did not perform state-of-the-art DNA typing? A second consideration involves whether an oversight committee might want to see the results of typing replicate evidence by many labs to have a basis of comparison for all the labs in the program. This requirement could not easily be met by a random audit/reanalysis model. Another consideration is the CODIS system, especially as more and more labs participate and backlogs are reduced. The problem might be avoided at least to an extent by judicious selection of case facts and specimens. But to the extent that PT specimen types are entered into CODIS, several problems are created. One is that the first few labs who complete the case will quickly figure out that the specimens are likely coming from PTs, possibly compromising the tests in many other labs. Another problem is the protection of specimen donors whose profiles may be entered into CODIS. There must be a way for these profiles to be purged. If the specimen donor pool was small, the DNA labs would pretty quickly figure out which profiles were always seen in blind PTs.

Models Tested in This Study

Details of our experience with actual blind proficiency testing in forensic DNA laboratories are the subject of a companion paper (2). Our objective was to test the "blind/LE" and "blind/CL" models because these were the most challenging. One blind/LE test was detected by a preliminary examiner, but was sent through blind to the DNA section, thus unintentionally becoming a single "blind analyst" trial. "Random reanalysis" was discussed in some detail with some laboratories and laboratory systems that use it as a QA tool.

Based on our experience and estimates of scale-up economies where relevant, we tried to estimate the costs of a large-scale national program under the different models.

Advantages and Disadvantages of the Different Models

There are several advantages and disadvantages aside from cost to the different blind PT models. They are summarized in bullet point form in Table 2. The Blind/LE model is most like actual casework for most laboratories, and the most complicated of the test models in terms of planning, evidence manufacturing and execution. The Blind/CL model is easier in terms of planning and evidence manufacturing, but primarily tests analytical skills. The same things can be said about Blind Analyst in comparison with Blind/LE. Random reanalysis (random audit with or without reanalysis) can test the whole system, but relies on the ability to select representative cases at random from a pool of worked cases where the evidence is still available, and where the biological evidence has not been totally consumed.

TABLE 2—*Advantages and disadvantages of the blind PT models*.

Model	Advantages	Disadvantages
Blind/LE	• Tests the whole system • Close to real casework situation • Large scale performance comparisons possible	• Complicated; elaborate planning • Risk of detection of blind test • Risk of revelation by LE
Blind/CL	• Evidence easier to manufacture • Less complex than Blind/LE • Large scale performance comparisons possible	• Does not test the whole system • Risk of revelation by CL
Blind Analyst	• Less complex than blind/LE • Easier to plan and execute than the above • Large scale performance comparisons possible	• Cannot test the whole system • Risk of detection or revelation
Random Reanalysis	• Can test the whole system • No evidence manufacturing	• Large scale performance comparisons not possible • "Randomness" may be difficult to achieve in case selection

Only the Blind/LE and Random Reanalysis models can test what is often called "the whole system." This term usually means that the criminalistics judgment abilities of the lab can be tested by introducing sufficient complexity in the case so as to require judgment, that the analytical abilities of the lab are tested, and that the interpretation of results is tested as well. In addition, intake procedures, record keeping, completion of forms, documentation, report writing, and the syntax and accuracy of conclusions can be assessed. In this context, it is important to have a pre-defined way of deciding what will be considered acceptable responses by target laboratories.

Estimated Costs of a Large-Scale Program

Our cost estimates for a large-scale, national blind proficiency testing program in forensic DNA laboratories are shown in Table 3. The estimated cost depends on the proficiency testing model. It was necessary to make some reasonable assumptions in arriving at these cost estimates. The blind proficiency testing feasibility trials in this study were all external, using the Blind/LE or Blind/CL models, and the test administration and manufacturing was centralized. Cost models based on decentralized or local manufacturing and/or test administration could give different estimates. We assumed that 150 laboratories would be tested either once or twice per year. We based our larger scale projections on an estimate of the cost per test. We calculated these costs from our limited blind test preparation, manufacture and administration. We also obtained cost estimates from a government agency test provider and from a private, commercial test provider for comparison. The cost estimates from this study are probably conservative. We included a

TABLE 3—*Cost estimate summary.*

Blind Proficiency Test Program Model	Estimates Extrapolated from This Project*		
	Cost/Test	One Test Per Year Total	Two Tests Per Year Total
Blind/LE, Blind/CL	$3,500	$535,000†	$814,000‡
Blind Analyst	$2,000	$310,000†	$630,000§
Random Reanalysis‖	$2,000–3,450	$330,000–517,500	$660,000–1,035,000
	Estimate from a Government Agency Test Provider		
Blind/LE, Blind/CL	$10,000	$1,510,000	$3,020,000
	Estimate from a Commercial Test Provider		
Blind/LE, Blind/CL	$3,400	$520,000	$1,050,000
Blind Analyst	$1,400	$220,000	$450,000

* All values are in US dollars.
† Includes costs of one proficiency test review meeting.
‡ 150% of one-test-per-year costs and includes two proficiency test review meetings.
§ Includes two proficiency test review meetings.
‖ The low-end figure does not include reanalysis of the biological evidence.

20% fringe benefit rate for personnel in the calculations as well as the costs of a proficiency test "oversight committee" (ten people) meeting.

It is likely that economies would be realized under a two test per lab per year requirement, i.e., it would probably not cost twice as much to do two tests per year as it did to do one per year. We projected the costs of doing two tests per year at 150% of the one-test-per-year cost, and we allowed for two meetings of the hypothetical proficiency test "oversight committee."

One might project that costs would decrease over time if a large-scale program was implemented and sustained. Over time and with experience, it would become easier to set up tests. Fewer person-days effort on the part of the test coordinators could be required, and travel might become less necessary. Some of these "savings" might never be realized, however, because of the continuous changes in personnel assignments in law enforcement agencies, and because inflationary pressures might simply offset any savings.

It is reasonable to assume that the costs of running a program under the Blind Analyst model would be less than those involved under the fully blind models. While laboratory administration is involved in planning the tests, we assumed that there still would be an external test coordinating entity tending to the details, manufacturing, and transmittal to the labs.

The cost estimates for a program under the random reanalysis model assume that the entire reanalysis is conducted by an entity external to the laboratory (and external to the laboratory system, if applicable), to be consistent with our definition of an "external blind proficiency test." It was assumed that the process would require 2 person days effort by an "auditor" (charging $500/day) and about 3 person-days effort by an "analyst" (charging $350/day). Included in the estimate were $1,200 travel costs and $200 consumables costs. The "auditor" would have to visit the target lab to review candidate cases for reanalysis, choose one or more, then gather all the information, records, and evidence. The 3 person-day estimate for an analyst is based on what it might take to reanalyze a case involving a series of specimens and the thirteen core CODIS STR loci. These assumptions give a cost per test estimate of $3,450. If the audit were conducted without any reanalysis, the cost per test decreases to $2,200. At the low end (where an audit does not include reanalysis of the biological evidence), these costs are roughly comparable to the estimates given for the Blind Analyst model. At the high end, they are comparable to those for the Blind/LE model. There is no requirement under this model for a national proficiency test-coordinating group, but there is likewise nothing that precludes having one. Funding for the operation of such a group was not included in the cost estimate for Random Reanalysis.

The Case For and Against Blind (vs. Open) Proficiency Testing

The case for blind (instead of, or in addition to, declared proficiency testing) is based on the information that might be obtained from blind proficiency tests that is not available from declared proficiency tests. Two points are generally cited in support of blind over declared proficiency testing.

First, there is some evidence from the clinical PT literature that examiners on the whole perform better in open proficiency tests because they know they are being tested. In regulated clinical laboratories, open proficiency testing is a major, if not the sole, criterion for a laboratory retaining its license to do particular tests. Here, analysts must also attest to the fact that they used their "standard" procedures in examining the sample, and that they did not collaborate with other analysts/laboratories. As such, proficiency tests are used more as regulatory tools in this environment than as educational devices designed primarily to improve quality. It is unclear to what extent this fact is responsible for the better performance in declared vs. blind tests. In the context of forensic DNA laboratories, it must be kept in mind that while declared proficiency tests can be directed to individual examiners (to the extent that laboratory division of labor permits it), blind proficiency tests submitted through LE agencies or CLs cannot. The Blind Analyst and Random Reanalysis models do permit particular examiners to be singled out for testing to the extent possible in an individual laboratory.

Second, it is often said that blind proficiency testing tests "the whole system" whereas declared proficiency testing primarily tests the ability to obtain acceptable analytical results, and maybe an acceptable interpretation of the results. By "the whole system" is usually meant all the steps and record keeping that go into case intake, sorting and selection of items for analysis, screening or preliminary tests, DNA analysis itself, interpretation of the results and prepara-

tion of a report. To the extent, therefore, that it is desirable to test all these aspects of the forensic lab analysis "system," in addition to the analytical results as part of an ongoing QA program, blind proficiency testing would be required with some suitable frequency.

Compared with declared testing, blind proficiency testing is complicated and expensive under either the fully external Blind/LE or the Random Audit/Reanalysis models. These are the only ones that test "the whole system." Much of the same information obtainable under the external "blind/LE" model is probably obtainable by Random Audit/Reanalysis. However, if random reanalysis is done externally, it is not materially less costly according to our estimates than fully external "blind/LE." Further, there are issues of case and evidence availability for audit/reanalysis, and the randomness of selecting cases. Finally, the type of interlaboratory comparison data available under a blind/LE model where "evidence" is manufactured is not available under random reanalysis.

The Recommendations of the National Forensic DNA Review Panel to the Director of the NIJ

The Review Panel had extensive briefings from the project staff about the background information and about the results of and actual experience with blind test trials. The purpose of this project, as stated in the DNA Act and in the original charge to the panel, was to make a recommendation on the "feasibility" of blind PT in forensic DNA laboratories. "Feasibility" could mean "possible," or it could mean "possible and practicable." The results of our study clearly show that blind PT is possible. And blind testing is and has been done in some labs independent of this project. The panel formulated its recommendation to the NIJ Director based on the second interpretation of "feasibility," i.e., is it practical in terms of all the problems, issues, complexities and costs. The final recommendation was three-pronged:

1. The accreditation system and associated quality assurance guidelines of the DNA Advisory Board need to be given the opportunity to take hold.
2. It is recommended that the DNA Advisory Board generate guidelines for more stringent external case audits for use by ASCLD-LAB, or another relevant accrediting body, as part of the accreditation process. The external case audits should be conducted regularly and serve as a measure of how well accreditation and its associated requirements are working in a quality assurance context.
3. In the extreme, blind proficiency testing is possible, but fraught with problems (including costs), and it is recommended that a blind proficiency testing program be deferred for now until it is more clear how well implementation of the first two recommendations are serving the same purposes as blind proficiency testing.

In connection with this recommendation, there was discussion about the fact that lab accreditation had already been made a requirement for forensic DNA labs, although no time limit was imposed. Lab accreditation should result in improved QA because accredited labs have to follow more stringent requirements, and are subject to audits and inspections. Part of the thinking behind the recommendations was to give widespread lab accreditation time to take effect. The second point of the recommendation suggested that audits and/or reanalysis be implemented as a means of testing how things are going as more and more laboratories become accredited.

Another line of discussion in connection with the recommendations had to do with the extent to which a random audit/reanalysis model for blind PT might provide virtually the same information from a QA point of view as what we have called the blind/LE or blind/CL models. The latter are more complicated. Another aspect of this thinking involved whether there is already scrutiny of DNA cases by defense counsel and their experts. Such scrutiny could be at least as thorough as a formal random audit/random reanalysis program. Part of the effort in Phase 2 of this project had to do with determining the extent to which there was review and scrutiny of DNA cases, and actual reanalysis of evidence, by defense counsel and defense experts.

References

1. *Daubert vs. Merrell Dow Pharmaceuticals*, 113 S. Ct. 2786, 1993.
2. Peterson JL, Lin G, Ho M, Chen Y, Gaensslen, RE. The feasibility of external blind DNA proficiency testing. II. Experience with actual testing. J Forensic Sci 2003;48(1):1–9.
3. Public Law No. 103-322. DNA identification act of 1994, subtitle c of the violent crime control and law enforcement act of 1994. September 13, 1994.
4. Anderson TD, Roby RK, Weedn VW. Analysis of the data from the College of American Pathologists — Forensic and parentage identity DNA proficiency testing programs. Armed Forces Institute of Pathology. Rockville, MD, 1995.
5. AABB. Standards for parentage testing laboratories, American Association of Blood Banks, Bethesda, MD, 1991.
6. Technical Working Group on DNA Analysis Methods (TWGDAM). Guidelines for Quality Assurance Program for DNA Analysis. Crime Laboratory Digest 1995;22/2:21–43.
7. Edinger SE. The Medicare and clinical improvement act of 1967 proficiency testing requirements and its relationship to the private sector. Arch Pathol Lab Med 1988;112:357–62.
8. Public Law No. 90-174. Clinical laboratory improvement act of 1967. Code of Federal Regulations, Title 42, Part 72, 1967.
9. Dept. of Health, Education, and Welfare, Public Health Service, Centers for Disease Control; Clinical laboratory improvement act of 1967. Part F. Title III, Public Health Service Act. Section 353, 1967.
10. Boeckx RL. Government regulation: can it guaranty quality in the clinical laboratory? Discussion Session I and Discussion Session II. Clin Chem 1992;38/7:1218–50.
11. Peddecord MK, Hammond HC. Clinical laboratory regulation under the clinical laboratory improvement amendments of 1988: can it be done? Clin Chem 1990;36/12:2027–35.
12. Public Law No. 100-578. 100th Congress. Clinical laboratory improvement amendments of 1988. Stat 42 USC 201. H.R. 5471. Oct 31, 1988.
13. Health Care Financing Administration. Medicare, Medicaid and CLIA Programs: Revision of laboratory regulations; final rule with request for comments. Fed Reg 1990a;55:9534–610.
14. Health Care Financing Administration. Medicare and laboratory certification programs; enforcement procedures for laboratories. Fed Reg 1991;56:13430–45.
15. Belk WP, Sunderman FW. A survey of chemical analyses in clinical laboratories. Am J Clin Pathol 1947;17:853–61.
16. Wagner LR. The College of American Pathologists, 1946–1996: Laboratory standards. Arch Pathol Lab Med 1997;121:536–541.
17. Copeland BE, Rosenbaum JM. Organization, planning, and results of the Massachusetts Society of Pathologists Regional Quality Control Program. Am J Clin Pathol 1972;57:676–88.
18. Glenn GC. College of American Pathologists Conference XIII on the evaluation of proficiency testing results for quantitative methods in relation to clinical usefulness: concluding remarks. Arch Pathol Lab Med 1988;112:471–2.
19. Skendzel LP, Hanson DJ, Cirin W. The 1949 College of American Pathologists survey revisited. Am J Clin Pathol 1970;54:493–5.
20. Laessig RH, Ehrmeyer SS, Lanphear BJ, Burmeister BJ, Hassemer DJ. Limitations of proficiency testing under CLIA '67. Clin Chem 1992a;38:1237–44.
21. Boone DJ. Evaluating laboratory performance: historical and governmental perspectives. Arch Pathol Lab Med 1988;112:354–6.
22. Ehrmeyer SS, Laessig RH. Effect of legislation (CLIA'88) on setting quality specifications for US laboratories. Scand J Clin Lab Invest 1999;59:563–8.
23. Peterson JL, Fabricant EL, Field KS, Thornton JI. Crime laboratory proficiency testing research program. U. S. Government Printing Office, Washington, D.C., 1978.

24. Peterson JL, Markham P. Crime laboratory proficiency testing results, 1978–1991, Parts 1 and 2. J Forensic Sci 1995;40:1009–29.
25. Pereira M. Quality assurance in forensic science. Forensic Sci Int 1985;28:1–6.
26. Werret DJ. Remarks at Banbury Conference in DNA technology and forensic science, Cold Spring Harbor, New York, November 1988;240.
27. ASCLD. Laboratory Accreditation Board Manual. American Society of Crime Laboratory Directors-Laboratory Accreditation Board. April 2001;45–7.
28. Technical Working Group on DNA Analysis Methods (TWGDAM). Guidelines for DNA proficiency test manufacturing and reporting. Crime Laboratory Digest 1994;21/2:27–32.
29. Peterson JL, Gaensslen RE, Lin G, Ho M, Chen Y. Developing criteria for model external DNA proficiency testing—final report. Chicago, IL: National Institute of Justice, 2001;30–1.
30. Office of Technology Assessment (U.S. Congress). Genetic witness: forensic uses of DNA tests (OTA-BA-438), Washington, DC: Government Printing Office (GP0052-003-01203-1), 1990.
31. National Research Council, DNA Technology in Forensic Science, Washington, DC. National Academy Press, 1992.
32. National Research Council, The Evaluation of Forensic DNA Evidence, Washington, DC: National Academy Press, 1996.
33. Kuo M (Chair). California Association of Crime Laboratory Directors: DNA Committee Report to the Board of Directors, Report #6, Orange County Sheriff Coroners Crime Laboratory. Santa Ana, CA, 1988.
34. Reeder D. DNA testing survey takes hold in crime laboratories. CAP Today. April l995.
35. Steadman GW. Survey of DNA crime laboratories, 1998. U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics 2000:1–8.
36. Hammond HC. Federal regulation of clinical laboratories: a prospective public policy analysis. Arch Pathol Lab Med 1988;112:363–7.
37. Hansell JR, Haven GT. Changes in level of precision of common ligand assays during a seven year interval. Amer J Clin Pathol 1979:72:320–5.
38. Elevitch P, editor. Data ReCAP 1970–1980, Northfield, III: College of American Pathologists; 1981.
39. Tholen, Lawson NS, Cohen T, Gilmore B. Proficiency test performance and experience with College of American Pathologists' Programs. Arch Pathol Lab Med 1995;119:307–11.
40. Peddecord KM. A critical review of personnel standards. Clin Chem 1996;42:790–6.
41. Stull TM, Hearn TL, Hancock JH, Handsfield JH, Collins CI. Variation in proficiency testing performance by testing site. JAMA 1998;279: 463–7.
42. Erickson R, Driscoll C, Dvorak L, et al. Factors related to accuracy in office cholesterol testing. J Fam Pract 1991;33:457–61.
43. Phillipou G, Farrant RK, Phillips PI. Computer based quality assessment of hospital capillary blood glucose monitoring. Diabet Med 1990;7:234–7.
44. Shehangian S. Proficiency testing in laboratory medicine: uses and limitations. Arch Pathol Lab Med 1998;122:15–30.
45. Bloch MJ, Cembrowski GS, Lembesis GJ. Longitudinal study of error prevalence in Pennsylvania Physicians' Office Laboratories. JAMA 1988;260/2:230–5.
46. Lawson NS, Gilmore BF, Tholen DW. Multiprogram characterization or laboratory bias, precision, and total error proposal for improved assessment with shared external and expanded internal quality control pools. Arch Pathol Lab Med 1988;112:454–61.
47. Strand CL. Proficiency Testing: One important component of continuous quality improvement. Am Clin Pathol 1994;90:393–4.
48. Cembrowski GS, Vanderlinde RE. Survey of special practices associated with College of American Pathologists Proficiency Testing in the Commonwealth of Pennsylvania. Arch Pathol Lab Med 1988;112:374–6.
49. Duckworth JK. Proficiency testing: its role in a voluntary clinical laboratory accreditation program. Arch Pathol Lab Med 1988;112:346–8.
50. Rej R, Jenny RW. How good are clinical laboratories? an assessment of current performance. Clin Chem 1992;38/7:1210–7.
51. Welch MJ, Hertz HS. The how and why of an accuracy base for proficiency testing. Arch Pathol Lab Med 1988;112:343–5.
52. Glenn, personal communication 1996.
53. AuBuchon JP. Mathematical model for comparing proficiency testing results across analytical methods. Clin Chem 1991;37:265–9.
54. Erhmeyer SS, Laessig RH. An evaluation of the ability of proficiency testing programs to determine intralaboratory performance: peer group statistics vs. clinical usefulness limits. Arch Pathol Lab Med 1988;112: 444–8.
55. Gilbert RK. Accuracy of clinical laboratories studied by comparison with definitive methods. Am J Clin Pathol 1978;70:450–70.
56. Grannis GF. Studies of the reliability of constituent target values established in large interlaboratory survey. Clin Chem 1976;22:1027–36.
57. Hartmann AE, Naito HK, Burnett RW, et al. Accuracy of participant results utilized as target values in the CAP chemistry survey program. Arch Pathol Lab Med 1985;109:894–903.
58. Boone DJ, Hearn T, Lewis S. Assessment of extent to which laboratories compare results before reporting in national laboratory performance surveys, abstracted. Clin Chem 1985;31:115.
59. Smith ML, Bronner WE, Shimomura ET, Levine BS, Froede RC. Quality assurance in drug testing laboratories. Clinics in Laboratory Medicine 1990;10:503–16.
60. Lamotte LC, Guerrant GO, Lewis DS, Hall CT. Comparison of laboratory performance with blind and mail distributed PT samples. Pub Health Rep 1977;92:554–60.
61. Hansen HJ, Caudill SP, Boone JD. Crisis in drug testing: results of a cdc blind study. JAMA 1985;253:2382–7.
62. Boone JD, et al. Laboratory evaluation and assistance efforts: mailed, on-site and blind PT surveys conducted by the Centers for Disease Control. AJPH 1982;72:1364–8.
63. Mason MF. Some realities and results of PT of laboratories performing toxicological analyses. J Anal Toxicol 1981;5:201–8.
64. Davis KH, Hawks RL, Blanke, RV. Assessment of laboratory quality in urine drug testing: a proficiency testing pilot study. JAMA 1988;260:1749–54.
65. Frings CS, White RM, Battaglia DJ. Status of drugs-of-abuse testing in urine under blind conditions: an AACC study. Clin Chem 1989;35/5: 891–4.
66. Glenn GC, Hathaway TK. Quality control by blind sample analysis. Am J Clin Pathol 1979;72:156–62.
67. Parsons PJ, Reilly AA, Esernio-Jenssen D, Werk LN, Mofenson HC, Stanton NV, et al. Evaluation of blood lead proficiency testing: comparison of open and blind paradigms. Clin Chem 2001;47/2:322–30.
68. Damato J, et al. The department of army quality assurance program for human immunodeficiency virus antibody testing. Laboratory Medicine 1988;19:577–80.
69. Peddecord KM, et al. Analytic results of HIV-1 testing using blind proficiency testing. Clin Lab Sci 1992b;5:165–71.
70. Peddecord KM, et al. Identification of nonanalytic issues in HIV- 1 antibody testing using blind proficiency testing. Clin Lab Sci 1992c;5: 172–6.
71. Schalla WO, Blumer SO, Taylor RN, et al. HIV blind performance evaluation: a method for assessing HIV-antibody testing performance. Proceedings of 1995 Institute on Critical Issues in Health Laboratory Practice. Pioneers in Laboratory Research. Atlanta, GA. Centers for Disease Control and Prevention 1996; Abstract.
72. Mudd JL, Baechtel FS, Duewer DL, Currie LA, Reeder DJ, Leigh SD, et al. Interlaboratory comparison of autoradiographic DNA profiling measurements. 1. data and summary statistics. Anal Chem 1994;66/20: 3303–17.
73. Duewer DL, Currie LA, Reeder DJ, Leigh SD, Liu H-K, Mudd JL. Interlaboratory comparison of autoradiographic DNA profiling measurements. 2. Measurement uncertainty and its propagation. Anal Chem 1995;67/7:1220–31.

Additional information and reprint requests:
Joseph L. Peterson, D. Crim.
Department of Criminal Justice (M/C 141)
University of Illinois at Chicago
1007 West Harrison
Chicago, IL 60607-7140