J Forensic Sci, Jan 2003, Vol. 48, No. 1
Paper ID JFS2002043_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. II. Experience with Actual Blind Tests*

ABSTRACT: The background and goals of a national study to determine the feasibility of blind proficiency testing in U.S. forensic DNA laboratories are discussed. Part of the project involved designing and executing a series of fifteen blind proficiency tests. Execution included biological specimen donor recruitment and case evidence manufacturing. Simulated cases were submitted to DNA laboratories by law enforcement agencies and in some cases by other forensic-science laboratories. Replicate-manufactured evidence was submitted to reference laboratories to simulate the workings of a larger-scale program. Ten tests were straightforward, and essentially tested analytical ability. Five tests involved selecting on the basis of case facts appropriate bloodstains for typing from a bloodstain pattern. We describe in detail our experience in designing and conducting these blind proficiency test trials, and relate those experiences to the overall issue of blind proficiency testing as a quality-assurance tool in forensic DNA laboratories. In this feasibility test series, one blind test was detected by a laboratory, a second one was shown to the lab by law enforcement, and a third was never completed because of lapses in communication. Turnaround times were relatively fast in the independent/commercial labs and relatively slow in the larger public laboratories. Two cross-state case-to-case CODIS "hits" were "planted" among the first series of ten blind tests. One pair was detected. One member of the second pair went to a lab that was not CODIS-ready.

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

In an accompanying paper (1), the DNA laboratory blind proficiency-testing project's underlying background, and the findings and recommendations concerning a national blind proficiency-testing program, were described and discussed. In this paper, the construction and execution of a trial series of actual blind proficiency tests, and the results of our experience with them, are presented.

The study design was prompted in large part by the language of the DNA Act of 1994 (2). As noted in the prior paper, open (or declared) proficiency testing is a Technical Working Group on DNA Analysis Methods (TWGDAM)/DNA Advisory Board (DAB) quality assurance (QA) requirement for DNA testing laboratories. The ASCLD Laboratory Accreditation Board (ASCLD-LAB) also requires proficiency testing to be part of the QA program in accredited laboratories. Although blind proficiency testing is not required by any of the QA guideline-setting bodies, it is discussed in some of the documents and recommended by TWGDAM: it is "highly desirable" for the DNA laboratory to participate in a blind proficiency testing program that "realistically simulates" actual casework in order to evaluate "all aspects of the laboratory examination procedure" (3). Mandatory external blind proficiency testing for forensic DNA testing laboratories, along with other recommendations, were proposed during a joint hearing on forensic DNA analysis in 1991 before the Subcommittee on Civil and Constitutional Rights of the U.S. Congress (4). Congress subsequently passed the DNA Identification Act of 1994 (2), which established a framework for setting standards on quality assurance and proficiency testing in forensic DNA typing laboratories. The law required the FBI Laboratory to engage in blind proficiency tests, and directed the National Institute of Justice to certify to the joint Committees (House and Senate) 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 present project was thus initiated to explore the feasibility of a national, blind DNA proficiency-testing program in public and private forensic science laboratories. A principal component of the project was exploration of the feasibility of actual testing by setting up and administering a small, test series of blind proficiency tests in forensic DNA laboratories.

The research was separated into two sequential phases, each comprising approximately two years. Emphasis in Phase 1 was, initially, on a comprehensive literature review and survey of forensic DNA laboratories and their practices. In Phase 1, we also attempted two cycles of blind tests, each cycle consisting of four tests in nonfederal laboratories and one in the FBI Laboratory (in collaboration with a separate contractor). 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, on exploring less costly alternatives, such as audits and reanalysis, and on the set-up and execution of somewhat more complicated blind tests. In that phase, five tests were administered to forensic DNA laboratories.

[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.

The nationwide survey of forensic DNA laboratories in Phase 1 determined the procedures laboratories employed in the routine handling, processing, and reporting of evidence; the law enforcement agency survey explored whether the police agencies were willing to participate in the study. The goal of the Phase 2 survey examined laboratories' sample retention practices and the frequency, nature, and outcomes of internal and external audits. In both phases, surveys were used to ascertain which laboratories were willing to participate in the actual testing. Surveys of defense attorneys and defense experts explored the nature and percentage of their DNA cases, which had been subjected to re-testing and various types of reviews.

All the proficiency tests in this feasibility project were designed to make the analysts believe they were actual cases. All were either sexual assault or blood transfer cases. CODIS issues were considered, and two potential cross-state CODIS matches were designed into the Phase 1 tests.

Methods

Surveys

In Phase 1, all laboratories performing forensic DNA testing in the U.S. were surveyed and asked their willingness to participate in the study. The mailing list was compiled from the National Institute of Justice DNA grantee list, and the FBI's 1995 Combined DNA Index System (CODIS) Survey laboratory register. A telephone call was made to every laboratory on the list to confirm that the laboratory was performing DNA analysis and to determine if there were any other laboratories in their states performing DNA analysis. Laboratories willing to participate (i.e., to be potential test candidates or potential reference laboratories) were required to indicate this willingness in writing. Laboratories and analysts were informed that they might be receiving a proficiency sample disguised as an actual case in the coming year. Those who did so were then asked to supply a list of law enforcement agencies and conduit laboratories that routinely submitted evidence to them for DNA testing. Those law enforcement agencies and conduit laboratories identified by the forensic DNA laboratories were also surveyed. Like laboratories, law enforcement agencies, and conduit laboratories willing to participate in the study had to return a written agreement to participate. A similar, but somewhat less elaborate procedure was followed in Phase 2, although every forensic DNA laboratory was re-surveyed. In Phase 1, 39 laboratories, 63 law enforcement agencies, and 9 conduit laboratories agreed to participate.

A "conduit laboratory" in the context of this project is a forensic laboratory that receives case evidence but does not itself do DNA typing. Following its intake procedures and preliminary evaluation and testing of case evidence, the conduit lab sends out samples of biological evidence to a larger or more specialized DNA testing lab. Conduit labs can be laboratories in a system that has only one or a few of its labs set up for DNA, or it can simply be a lab that did not yet have its DNA program up and running. Blind proficiency tests submitted through conduit laboratories are easier to manufacture because the target lab expects cuttings rather than whole evidence items. It is also easier to get proficiency samples into the target laboratory because of the cooperation of the conduit lab that knows what the target lab is expecting.

Blind Proficiency Test Manufacturing

We used the services of a contractor that had experience in manufacturing open proficiency test specimens. Accordingly, the contractor's personnel were familiar with and experienced in donor recruitment, specimen acquisition, storage and handling, careful attention to detail and rigorous quality control in proficiency test specimen preparation, and comprehensive record keeping. The detailed setup agreement with the submitter, and the case scenario, then enabled us to prepare detailed instructions for the manufacturer to put the biological evidence together. Complete specifications for a "case" included detailed instructions for the collection of specimens from donors, for use of those specimens in manufacturing the biological evidence, and for transmitting the items to the submitting entity. We also included a document to be sent to the submitting law enforcement agency or conduit lab along with the "case." In some instances, additional instructions for marking the evidence, etc., was included with this document.

There are ethical issues in recruiting donors for forensic lab proficiency tests, especially blind tests involving biological evidence where DNA profiles may be entered into CODIS databases. Anonymizing donor identities to those who are actually making up the tests is one strategy, probably an advisable one, for avoiding problems in some unlikely but possible scenarios where a biospecimen donor gets involved in a criminal case while his or her profile is in CODIS as a result of being a volunteer donor. Furthermore, there needs to be a reliable mechanism for removing proficiency test biospecimen donor DNA profiles from databases. Any biomedical laboratory or university doing research would be required to submit a detailed protocol for approval to an Institutional Review Board (IRB) which enforces the regulations for the protection of research subjects. The board would also have to approve a written consent form for subjects. Under some circumstances, an IRB will approve a consent that is administered to a volunteer, but does not have to be signed. This strategy can also be used to anonymize subjects to researchers. Because these boards tend to govern the conduct of subjects in *research*, it is not clear how these ethical matters might be handled if a proficiency test manufacturer were not a research facility subject to Office for Human Research Protections (OHRP) in the U.S. Department of Health and Human Services (DHHS).

Biospecimens from volunteer donors as a matter of course should be screened for bloodborne pathogens, and positive specimens should not be used in manufacturing proficiency tests. A policy is also needed to govern the notification of a volunteer donor about any positive bloodborne pathogen results in his/her specimen.

In a proficiency test, it must be decided in advance how an acceptable result will be defined or determined. One approach is the use of reference labs. Another is the use of consensus results. Either way, it is essential that test specimens be prepared as close to identically as humanly possible. Test items should have identical biological specimens, in the same locations, and containing the same quantities of DNA. For blood specimens, white cell counts should be checked to insure they are within normal range. For semen, specimen quantities in or on blind test items should be adjusted such that like specimens contain close to the same number of sperm cells. Similarly, we requested that female vaginal swab donors not collect more than 4–6 swabs in a 24-hour period to help insure some consistency in the quantities of exfoliated epithelial cells. If saliva specimens were to be used in the preparation of blind test items, attention would have to be given to the cell quantity in them to help insure uniformity.

In proficiency tests requiring clothing, the ideal approach is to buy the needed items of old, worn clothing from the volunteer biospecimen donors. If blind tests were to be manufactured on a larger scale, this practice might be impractical. A second method is

acquisition of the required items from various used clothing outlets. Clothing obtained from the volunteer biospecimen donor does not have to be washed unless the case scenario calls for it. Second hand clothing obtained from outlets should be washed thoroughly before use to avoid introducing unknown and unexpected biospecimens and DNA profiles into the tests. In sexual assault scenarios where the case requires the submission of the complainant's panties, there is no workable alternative to obtaining panties actually worn by the biospecimen donor; that is, this item cannot be manufactured in a sufficiently convincing way to fool an experienced forensic biologist.

Planning the number of biospecimens needed, and thus the numbers of donors needed, as well as the quantities of clothing, sexual assault evidence collection kits, etc. depends on planning the type of case that will be used to house the blind test. This plan must be very detailed, and must be forged in close collaboration and cooperation with the jurisdiction of the target laboratory. Every jurisdiction and every lab have their standard evidence collection procedures, requirements, packaging, labeling, and so forth. In many places, sexual assault evidence collection kits are jurisdiction or lab specific, and all the specimens expected in a normal case must be provided. For example, if complainant pubic hair combings are always collected along with clipped knowns, this practice must be followed in constructing the blind test. In some scenarios, the race or ethnic origin of the hypothetical victims and suspects may be important. Cases manufactured around those scenarios can require that volunteer donors of specific ethnicity be recruited, depending on the types of specimens that must be submitted.

We followed a practice of providing or including the minimum number of items that would be expected in a case in a particular jurisdiction without drawing attention to it. This strategy minimizes the number of specimens and items that must be handled, and thus reduces the possibility of errors in manufacturing and test preparation. Generally, we did not manufacture more than ten blind "cases" in a single work session (usually about 4–6 hours).

Blind Proficiency Test Feasibility Trials

In Phase 1, there were 39 potential participating target labs among the 94 respondents. We selected 8 of the 39 based on their being: (a) reasonably representative of the different types of laboratories; (b) distributed over a large geographical area and range of jurisdictions; and (c) accessible through a conduit laboratory or law enforcement agency that would cooperate with us in submitting a blind proficiency test case.

A TWGDAM-approved DNA proficiency test manufacturer that agreed to manufacture the proficiency tests to researchers' exact specifications was employed (5). This manufacturer had its own well-developed quality-assurance protocols in place for the manufacture of tests. It also had a pool of potential biological-evidence donors. The FBI Laboratory, which was required by the DNA Act to engage in blind proficiency tests, made arrangements with a separate contractor to submit cases to them. That contractor worked with our research team to submit two blind cases contemporaneously with this study, though technically not a part of it. Thus, the Phase 1 tests consisted of ten blind cases for submittal to DNA laboratories, including two cases submitted to the FBI Laboratory. We also selected two reference laboratories from among the pool of potential participants, based on their reputation for outstanding work. Phase 1 tests were manufactured and administered in two cycles, five tests at a time.

Blind proficiency test sample donors were recruited by the manufacturing contractor. Approval of the project, and for the use of human subjects in research, was obtained from the University of Illinois at Chicago Institutional Review Board (IRB). It was necessary to employ a detailed informed consent statement, because of the possibility of donor DNA profiles ending up in local, state and/or national CODIS files until the tests could be completed and closed out. We took extra precautions in this respect, requiring participating laboratories to certify in writing that any profiles related to blind proficiency tests had been removed from databases or data banks. Informed consent was obtained by the manufacturing contractor, and the donors were anonymized to the UIC researchers. We were additionally advised by NIJ that the donor identities enjoyed further protection under the U.S. Code of Federal Regulations, such that neither our contractors nor we could be compelled to release them. Another issue presented to, and ultimately approved by, the UIC IRB in the context of protecting research subjects was that we were engaging in deception of the laboratories, and asking law enforcement or conduit lab personnel to be party to that deception.

Five volunteer donors, three females and two males, contributed biological specimens to the Phase 1 blind proficiency tests. Detailed requirements were written for the nature and quantity of biological specimens required from each donor. Phase 1 tests included eight sexual assault cases and two blood transfer cases (see Table 5). The sexual assault cases included underwear and slides/swabs from the victim and tubes of blood from both the victim and suspect. The general case scenario given by police investigators to the laboratory had the alleged victim reporting the crime to police, being taken to a medical facility for examination, a rape kit/evidence being taken, and a suspect having his blood drawn at a hospital or clinic. Some female donors contributed panties worn for at least one day and not laundered prior to submission. Semen was counted for sperm density before spiking swabs and/or panties. Under the assumption of 2.5 pg DNA/cell, swabs and/or panties were spiked with semen sufficient to contain at least 1–2 μg sperm DNA, more than enough, in theory, for RFLP typing.

Agreements were made and signed with law enforcement agencies and laboratories (conduit labs) that participated as submitters in this project. In this project, blind proficiency test cases were introduced to DNA laboratories in one of two ways: through law enforcement agencies, and through other laboratories who did not do DNA testing themselves, and who regularly submitted evidence for DNA typing to the target lab. Thus, for this project, a "target" lab is a DNA typing lab that actually received a "case" for DNA typing. A "conduit" lab was another forensic laboratory that does not do DNA typing, but regularly sends cases or evidence items to the target lab for DNA typing. Some conduit labs were forensic labs that regularly submit evidence to private, independent labs for DNA typing. Others were part of state systems where DNA typing is restricted to a central location or to a few labs in the system. We made contact with appropriate conduit laboratories just as was done with law enforcement agencies. As noted above, an agreement was signed with the conduit labs that was very similar to that used with law enforcement agencies submitting "cases" for this project.

In Phase 1, the blood transfer case was one in which the laboratory was asked if a bloodstained item of cloth taken from a suspect's home was associated with the victim. One general scenario described a victim bludgeoned to death in her residence and a bloodstained shirt recovered from the suspect's home. The bloodstained shirt and bloodstain standards from both the victim and suspect were submitted to the laboratory for testing. The laboratory was asked to determine if the bloodstained cloth associated the victim with the suspect.

In Phase 2, 67 laboratories doing DNA analysis responded, and 30 agreed in writing to be potential participants. Emphasis in Phase 2, as noted earlier, was on "more challenging" blind tests. By "more challenging" was meant that the analysts would have to use some judgment in selecting specimens, or in selecting from among multiple stains, for analysis. Enough information about the case was provided to enable reasonable judgments. Another way of stating this goal of the Phase 2 testing is that it was designed to require criminalistics reasoning. The Phase 1 tests were quite straightforward, requiring mainly analytical ability. One of the oftenmentioned advantages of blind testing over declared testing is its ability, at least in theory, to test judgment and interpretation as well as analytical ability.

Phase 2 tests were designed and manufactured by the same contractor as was employed in Phase 1. It was decided to use an assault/sharp-force scenario for this series of tests, one that would result in multiple bloodstains on a victim's clothing but where both victim and perpetrator were injured and contributed bloodstains. The "victim" was the major contributor (i.e., most of the stains were victim). Another issue in proficiency test manufacture is replicability, i.e., are all the evidence specimens really identical, and in particular in this test, are all the target lab and reference lab evidence specimens identical.

This issue is important in proficiency testing, because it relates to whether reference laboratory values can be taken as the standard by which to judge the performance of target labs, and/or whether performance comparisons among target laboratories are justified. We decided to partly use the Phase 2 actual testing protocols to test our ability to manufacture replicate evidence specimens, particularly because the scenario required depositing two bloods on clothing to resemble a certain blood stain pattern. In addition to the case information, the pattern was designed to "lead" the analysts to look at certain stains while perhaps ignoring others. For these reasons, we decided to manufacture ten "replicate" tests. Five were used as blind test evidence items and submitted to laboratories. The other five were submitted to other "reference" laboratories. The latter were not given any more information about the "case" than the target laboratories, but they did know they were acting as "reference" laboratories for the project. The reference labs were told to type the "evidence" specimens the way they would do an actual case, but to type the reference (exemplar) specimens for every system they had up and running. The evidence manufacturing scheme is shown in Table 1.

All the Phase 2 cases were designed around the submission of three items: a pair of pants or sweat pants worn by the "victim," a "victim" exemplar and a "suspect" exemplar. For most of the cases, a female "victim" and a male "suspect" were used, but in one jurisdiction, it was necessary to have a male "victim." Therefore, three specimen donors, including two males and one female, contributed biological specimens for Phase 2. All the suspects and victims in all the blind test "cases" were assigned fictitious names, dates of birth, and race/ethnicity, always in consultation with the submitting police agency or conduit laboratory.

As briefly noted earlier, we used semen from two donors to construct cases that had the potential to be detected by CODIS. At the time, the national CODIS system was not fully operational, but laboratories were coming on line as our project was progressing. As a result, we were not able to predict whether the matches would be found. In the two potential "CODIS hit" pairs, the cases were sexual assaults and one member of the pair of blind tests went to the FBI Laboratory through its contractor. The other member of the pair went to a target state laboratory.

In both phases of the project, throughout all the actual testing trials, we made formal written agreements with both potential target labs/reference labs and with potential participating law enforcement agencies. Among other things, we gave all of them assurances of anonymity. They gave the researchers certain assurances as well, including one guaranteeing that any blind PT case profiles would be purged from databank and/or database files.

Results

Laboratory Survey Data and Characteristics of Participating Laboratories

A total of 151 laboratories were identified and surveyed in December 1996 during Phase 1. In Phase 2, we surveyed 91 laboratories, 67 of which were actually doing DNA analysis. A comparison of our Phase 1 laboratory survey results and those from the FBI CODIS lab survey is shown in Table 2. Signed agreements to potentially participate as target and/or reference laboratories were received from 38 laboratories in Phase 1 and 30 in Phase 2. The characteristics of those laboratories are shown in Tables 3 and 4.

Results with the Small-Scale Blind Proficiency Testing

As noted, ten tests were set up, manufactured, and submitted, in Phase 1. Two reference laboratories were chosen. The latter re-

TABLE 1—*Manufacturing scheme of the Phase 2 blind proficiency tests.*[*]

Blind Tests	Male Suspect (M1)	Male Victim (M2)	Female Victim (F1)	Exemplars (Reference)
11	Minor†	Major‡	. . .	M1 and M2
12	Minor	. . .	Major	M1 and F1
13	Minor	. . .	Major	M1 and F1
14	Minor	. . .	Major	M1 and F1
15	Minor	. . .	Major	M1 and F1
Reference Labs				
1	Minor	Major	. . .	All three
2	Minor	Major	. . .	All three
3	Minor	. . .	Major	All three
4	Minor	. . .	Major	All three
5	Minor	. . .	Major	All three

[*] The first ten blind tests were administered in Phase 1 and were numbered 1–10; Phase 2 blind tests were numbered 11–15.

† Minor: 1 drip stain + satellites (2–3 max).

‡ Major: approximately 6–7 drip stains + satellites (7–9 total); smear right pocket, or "right pocket" area.

TABLE 2—*Comparison of FBI CODIS laboratory data and survey data from Phase 1 of this study.*

	1995 FBI CODIS Survey	1996 UIC DNA Survey
Number of Labs	120	102
States represented	42	42
Number of federal agencies	2	1
Labs performing RFLP	58 labs in 32 states	11 labs in 11 states
Labs performing PCR	55 labs in 22 states	42 labs in 24 states
Labs performing RFLP and PCR	30 labs in 17 states	41 labs in 28 states
Labs performing STR	5 labs in 5 states	27 labs in 15 states

TABLE 3—*Characteristics of laboratories that agreed to participate in Phase 1 (N = 38).*

Number of Labs (% Total)	Type	Service Area	Number of DNA* Analysts (Avg.)	Cases Come Mainly From
17 (44.7)	State System	Entire State†	*N* = 17 (7.35)	Police/Other Labs
10 (26.3)	County	Entire County	*N* = 8 (2.75)	Police
5 (13.2)	City	Entire City	*N* = 5 (5.6)	Police
6 (15.8)	Private/Independent	No Limits	*N* = 6 (4.67)	Police, Attorneys, Other Labs

*N= number of labs that reported the number of DNA analysts.
†Not necessarily the sole provider of DNA typing services in the state.

TABLE 4—*Characteristics of laboratories that agreed to participate in Phase 2 (N = 30).*

Number of Labs (% Total)	Type	Service Area	Number of DNA Analysts (Avg.)*	Cases Come Mainly From
16 (53.3)	State System	Entire State†	*N* = 17 (7.35)	Police/Other Labs
7 (23.3)	County	Entire County	*N* = 8 (2.75)	Police
4 (13.3)	City	Entire City	*N* = 5 (5.6)	Police
3 (10.0)	Private/Independent	No Limits	*N* = 6 (4.67)	Police, Attorneys, Other Labs

*N = number of labs which reported the number of DNA analysts.
†Not necessarily the sole provider of DNA typing services in the state.

TABLE 5—*Characteristics of the ten Phase 1 blind proficiency tests.*

Test	Target Lab Type	Submission Through[a]	Type of Case†	Type DNA Testing Done‡	Reported Findings	Turnaround Time (*m*)§
1	Private	CL	SA	PCR	Suspect included	0.38
2	Private	CL	SA	PCR, STR	Suspect included	1.63
3	State	CL	BT	RFLP	Blood on suspect clothing consistent with victim	5.06
4	State	LEA	SA	RFLP	Suspect included	6.90
5	Federal	LEA	SA	RFLP	Suspect included	16.53‖
6	State	LEA	SA	PCR	Suspect included	3.46
7	Municipal	LEA	SA	RFLP	Suspect included	1.00
8	State	LEA	BT	N/A	Not completed#	3.00
9	State	LEA	SA	RFLP	Suspect included	4.06
10	Federal	CL	SA	RFLP	Suspect included	6.36

*CL: conduit laboratory; LEA: law enforcement agency.
†SA: sexual assault; BT: blood transfer.
‡RFLP; PCR: HLA-DQA1 and/or PM and/or D1S80; STR; N/A: not applicable.
§Months, obtained by consistently dividing turnaround time in days by 30. The turnaround time is calculated from the date of submission of the last specimen in the case to the date of the laboratory's report.
‖Hairs and fibers unit completed its report with a turnaround time of 2.16 m.
#Lab request to police for additional specimen not communicated in a timely way to project team.

ceived whole blood anticoagulated in Na_2EDTA, corresponding dried bloodstains, semen-free vaginal swabs from the female donors, and semen-only containing swabs from the males, as well as semen-spiked vaginal swabs prepared contemporaneously with the "cases."

Overall features of and results from the ten Phase 1 tests are shown in Table 5. Two tests were submitted to private, for-profit labs, 5 tests to state labs or labs that were part of state systems, 1 to a municipal lab, and 2 to the FBI Lab through its contractor. The tests were manufactured and submitted to forensic DNA laboratories in two cycles, each consisting of five "cases." Cycle one submissions were completed around August 6, 1997, and cycle two submissions around October 20, 1997. Four tests went to DNA laboratories via conduit laboratories while six others were submitted through law enforcement agencies. One of the blind proficiency test cases was detected by the target laboratory because the criminalist noticed that the vaginal smears (slides) were not streaked in the manner typical to the jurisdiction. However, the laboratory manager let the evidence items move on to the DNA typing unit without the DNA examiners knowing it was a proficiency test until after the fact. In addition, one of the cases submitted to a state system lab was never completed because the police, who were asked by the lab for a specimen of victim's blood for comparison, did not communicate the request to the project team.

Of the two pairs of cases having potential case-to-case, cross jurisdiction CODIS matches built in, one was found. One member of this pair was submitted to the FBI Lab by a state lab for DNA typing while the other member of the pair was submitted to a state system DNA laboratory by a local police agency. The state lab completed its case first, and entered the profile into national

CODIS. It was then detected when the FBI Lab worked the case. The second potential CODIS match was not found because the state laboratory that received one member of the pair was not connected to CODIS at the time the case was completed, and it had been notified the "case" was a blind and purged the profile before the second lab ever worked the paired "case."

All the labs reported the correct results, in the sense of including and excluding a possible depositor of biological evidence. Some target labs did RFLP, some did the PCR-based loci HLA-DQA1, PM and/or D1S80, and one did several STR loci. Reference labs typed six or seven RFLP loci, HLA-DQA1, PM loci, D1S80, and one did a number of STR loci. There was complete agreement among all laboratories on the discrete genotypes. Further, the band sizes obtained by RFLP typing of independent specimens from the same source, or from duplicate typing of the same specimen, were closely clustered, both within and between laboratories. Private laboratories generally had low turnaround times compared with state and federal labs. The Phase 1 target municipal lab finished its case within a month.

In Phase 2, five additional separate blind proficiency tests were manufactured and submitted to labs. As noted above, ten replicate three-item "cases" were manufactured simultaneously. Five were submitted to forensic science laboratories as blind proficiency tests; the other five were submitted to reference laboratories. Major characteristics of the five blind test target laboratories are shown in Table 6. The scenario underlying the manufacture of the Phase 2 "cases" involved assault and/or attempted sexual assault where both victim and perpetrator received sharp-force injuries from a knife that was not recovered. Bloodstains from dripping during a struggle, resulting in a low velocity pattern, were deposited on the "victim's" pants or sweat pants. In four of the cases, the "victim" was female; in one, the "victim" was male.

Target labs were told in the brief case description accompanying the submissions of the summary facts just noted. The "perpetrator" bloodstains were fewer than the "victim" stains in all the "evidence." There was always a blood smear near the right hand pocket, or a location on sweat pants in that general area, of "victim's" blood. The reason behind the placement of that stain was that the victim (wearer of the pants) could have wiped a bloody hand on the pants in that location creating the smear. Although the case scenario was not unduly complicated, it required criminalistics judgment in evaluating the blood stain pattern and in selecting stains for typing. Some labs did more stains than others. But every target and reference lab found at least one stain matching the "suspect" in addition to several that matched the "victim." Three target laboratories did PCR-based loci HLA-DQA1, PM (and sometimes D1S80), and two did STR loci, reflecting the transition from HLA-DQA1 and PM to STR loci taking place in the period when these tests were conducted. Turnaround times in the Phase 2 tests varied from a little over two to a little over 11 months, probably due to the individual lab situations such as number of DNA cases in the jurisdiction, case backlogs, seriousness of the cases, etc.

At one target lab site, our initial police contact person revealed the plans for the blind test to laboratory management. Subsequent discussions with higher ranking police officers in the same department enabled us to submit the test anyway. It turned out to be a true blind test in the end, because lab management thought we had decided to cancel the test after the initial revelation.

Of potential interest, though tangential to proficiency testing, is variation in DNA laboratory reports. All the target laboratories stated in some way that one or more stains could have originated from the "suspect," or that the "suspect" could not be excluded as a possible source, as well as some added language stating that the "victim" could be excluded. Every laboratory calculated probabilities of chance match for White, Black, and Hispanic U.S. populations for the "suspect" bloodstain profile. One laboratory calculated these values for the "victim" stain profile. Most of the reports were brief, and did not provide any primary data (i.e., actual types). Some listed the loci typed, but not the results obtained. One laboratory provided a summary of its results, but followed the summary with very detailed data on each item and stain examined, including controls it had run.

Discussion

Of the four possible models for blind proficiency testing developed in the project and described in the previous paper (1), we concentrated on the blind/LE and blind/CL modalities in our design of actual small-scale tests. The primary reason for this was our belief that these two modalities would be the most difficult and costly to set up and administer. There is no difference between the Blind/LE and the Blind/CL models in terms of the level of blindness to the DNA analyst. As expected, we found in this study that the manufacturing and submission of blind proficiency tests through conduit laboratories was a simpler process than submission through law enforcement agencies. A primary reason is that conduit labs are the first receivers of the evidence, and typically "work" the cases to some extent before sending selected biological-evidence items away to a larger or more specialized lab for DNA typing. Manufacturing a blind proficiency test "case" of this kind, therefore, requires fewer specimens and less paperwork. The target laboratories are not only removed from the case facts but also less likely to become suspicious. On the other hand, cases submitted by law enforcement agencies to DNA laboratories need to be more complete,

TABLE 6—*Characteristics of five Phase 2 blind proficiency tests.*

Test	Target Lab Type	Submission Through*	Type of Case†	Type DNA Testing Done‡	Reported Findings	Turnaround Time (*m*)§
11	State	LEA	BT	PCR	Suspect included	4.5
12	State	LEA	BT	STR	Suspect included	11.4
13	Municipal	LEA	BT	STR	Suspect included	3.1
14	Municipal	LEA	BT	PCR	Suspect included	9.5
15	State	LEA	BT	PCR	Suspect included	2.1

*LEA: law enforcement agency.
†BT: blood transfer.
‡PCR: HLA-DQA1, PM and sometimes D1S80; STR, various combinations of loci.
§Months, obtained by consistently dividing turnaround time in days by 30.

consisting of items actually seized at scenes, during investigations, or from suspects or victims. Specific paperwork, proper forms as well as evidence labels must accompany the evidence to the laboratories. As we learned in the test that was detected by the first examiner criminalist, a slight deviation from expectation or from the norm for a jurisdiction can raise suspicions.

As noted, the blind test that was detected by the first examiner was allowed by lab administration to be sent on for DNA analysis without revealing to the DNA analysts that it was a blind test. In this case, the first examiner and supervisor acted as a "conduit" to the DNA unit. If the test had been planned this way, we would characterize it as following the "blind analyst" model.

Another potential pitfall in blind/LE cases that we experienced first-hand is clandestine revelation of the test to the lab by the cooperating law enforcement personnel. In this instance, we learned about it and were able to submit a blind test anyway without the lab's knowledge. It is obviously possible, however, that the existence of the blind test could be revealed to the lab without anyone ever finding out. An added potential problem that surfaced in the trials is that lapses in monitoring of the active, submitted cases by program administrators can result in a case not being completed. In our trials, a request from the lab to the police to obtain and submit a further exemplar specimen was not conveyed to us. As a result, the case languished in the backlog of a busy lab.

We predicted, and believe that we have learned through the process, that successful planning, manufacture, and execution of a blind proficiency test involves detailed discussions with the submitter, whether that is a law enforcement agency or a conduit lab. The more "local" a laboratory is, the more difficult this problem becomes. By "local," we mean close relationships and familiarity between lab and users. There are generally personal relationships among user agencies and lab personnel, and, over time, high levels of mutual trust are developed. These factors, which are widely (and properly) touted as desirable become potential obstacles in blind proficiency testing programs. It may be difficult to convince people who like and trust one another, and who work together, to engage in the deception of the other party in the service of an intangible quality-assurance goal.

Another lesson from the small-scale testing project is the importance of the manufacturing laboratory. Its personnel are intimately involved in the planning of blind-test case submissions, obtaining materials and potential donors, actually preparing the evidence items, and transmitting everything in a pre-planned way to the submitting entity. Working with two or three manufacturing laboratory personnel at a time, we found that about ten "cases" could be manufactured at one time (i.e., simultaneously) while maintaining careful control over the process and minimizing errors. This number could probably be increased to some extent, but there are definite limits. It must be kept in mind that errors at the test manufacturing level could potentially be responsible for unsatisfactory responses from target laboratories. This problem is not limited to blind tests; it can happen with any proficiency testing program. A lab could later attribute its assessed "unsatisfactory" performance to errors in manufacturing, storage or transmittal of test materials rather than to any defects in its own protocols or personnel. QA is thus of first-order importance in a proficiency-test manufacturing facility. We were fortunate in this project in finding and using a very careful, reliable, and experienced manufacturing laboratory.

The blind/LE and blind/CL modalities of blind proficiency testing are fully "external," because no one in the target laboratory has to be involved in the design, manufacture or submission of the test "case." In the blind analyst model, others in the lab are in on the plans for the test by definition. And in random reanalysis, people within the same lab must be involved to some extent, although the selection of the case or items for audit/reanalysis could be done "externally," by another person in the lab system (if relevant) or by an outsider. This issue of "external" is raised because it was emphasized in the testimony before Congress leading up to the drafting and passage of the DNA Act. There are related issues here about whether a laboratory should be allowed to be solely responsible for its own QA program (including any blind proficiency testing) or whether there should be a requirement for the involvement of an outside ("external") entity. The thinking is that there is a greater chance of analysts secretly learning about a blind test if a lab constructs its own internal blind tests. Perhaps this logic applies even to another member lab in a system. But perhaps of more real importance in the end is the credibility factor: blind proficiency tests constructed and submitted to labs by an independent, external entity, and successfully completed, provide more public confidence in lab operations than if the tests and testing were internally administered, even if everyone in the lab and/or system is scrupulously honest.

The DNA Act (2) defined a "blind external proficiency test" as one ". . . that is presented to a forensic laboratory through a second agency and appears to the analysts to involve routine evidence." Congress did not make clear what it meant by "second agency," and in reality, this issue is much more complicated than it first seems. In the case of a municipal laboratory, for instance, is the "East Acme" Police Department a "second agency" with respect to the "East Acme Police Department Forensic Science Laboratory?" In a multi-laboratory system, is one of the labs a "second agency" with respect to the others? By definition, the success of blind proficiency testing depends on the test administrator's ability to plan, manufacture and then submit cases exactly in accordance with the laboratory's normal procedures, practices and expectations. The normal expectations include receiving cases from the usual submitters. And, in some cases, laboratories can have virtually a single agency as their client.

As a result of the project, we have attempted to define an external blind proficiency test as one of: (a) A test presented to a target lab through a law enforcement agency or a conduit lab in which the "case" or "evidence" was externally manufactured, and no one in the target lab has any advance information about the test; or (b) A test presented to the DNA analysis unit in which the "case" or "evidence" was externally manufactured, and in which the fewest possible personnel outside the DNA unit are informed about the test in advance; or (c) A test by "random reanalysis" in which auditors/analysts from outside the laboratory (and outside the laboratory system if the lab is part of a system) select the case for reanalysis, audit and review all the work done in the case, and reanalyze the biological evidence. This definition is more cumbersome, but more reflective of the circumstances of real world forensic science laboratories.

We did not purposely design tests in this project using the Random Reanalysis model. We thought it better to use the limited resources attempting to execute tests under the most difficult models. However, it is obviously possible because a great number of laboratories routinely use it as a part of their ongoing QA program.

In Phase 2 of the project, we focused more attention on whether evidence items representative of more challenging cases could be replicated with sufficient reliability to insure uniform results from competent laboratories. The findings showed that, at least for purposes of blind proficiency tests, manufacturing replicate evidence is possible, although it is labor intensive. By "more challenging," we meant that a modest level of criminalistics judgment, i.e., the selection of stains for DNA typing based on case information, was

part of the test. Obviously, the ability to perform the criminalistics judgment task affects the overall outcome in that failure to test the proper stains could result in an uninformative (victim's blood on victim's clothing) outcome. Blind proficiency test cases that were even more challenging and complicated could easily be imagined. However, it must be kept in mind that greater complexity in the number or nature of case materials (such as having more items, or having more complicated stain patterns) makes replicate manufacturing more difficult. At some level of difficulty, it is likely to be rendered impossible from a practical point of view. We did not choose biological-specimen mixtures for our Phase 2 tests in part because of the difficulties of replicate manufacturing, and in part because the NIST has conducted a national trial using mixtures (6).

The issue of proficiency-test difficulty is more complicated than it might first appear. First, difficulty is a relative term. And second, choice of test "difficulty" is related to the purpose of doing the proficiency testing in the first place. It doesn't matter whether proficiency testing is declared or blind in terms of this point. The purpose of proficiency testing has to be decided in advance of any program, to let participants know why they are being tested, how their performance will be evaluated, and thus to secure their enrollment. If the tests are to be constructed so far to the margin that most labs will perform unsatisfactorily, voluntary participation would suffer. Labs would see such testing as a set up, designed to trip them up on purpose, for possible impeachment later in court, but not for any articulable QA goal. However, one of the often discussed "virtues" of blind over declared testing is the ability to "test the whole system." In the context of forensic DNA laboratories, that means testing the criminalistics judgments and interpretation of results in terms of the case facts along with the analytical capabilities.

As noted in the previous paper (1), the project advisory board did not in the end recommend plans for implementation of a large scale, national blind proficiency testing program for forensic DNA laboratories. The project did, however, yield some information relevant to consideration of such a large-scale program. The only way that the program would serve the purpose of giving a national picture of performance levels would require a national level review of results. Another issue is whether to have a national coordinating entity for the tests, which is probably essential when implementing the blind/LE, blind/CL, or blind analyst proficiency testing models. Coordination of the program would involve planning the tests, setting up appropriate contacts, setting out manufacturing specifications for blind tests, making decisions on the number and types of tests per lab per year, and developing guidelines for acceptable responses. In the prior paper, estimates of the costs of a national program were given. There is wide variation, depending on how conservative one's cost and overhead estimates are, but a national program would be costly by anyone's standards. A national blind proficiency testing relying solely on random audit/reanalysis is much simpler, because there is no manufacturing involved.

Some of the drawbacks to a random audit/reanalysis model as against a blind/LE model that were explored in Phase 2 of this project include: (a) evidence from cases selected for audit must still be available; (b) there must be sufficient biological evidence in cases selected for random audit to enable reanalysis universally, or if it is judged to be warranted; and (c) the same type of national, inter-laboratory comparisons that can be made with manufactured cases and evidence are not possible with random audit/reanalysis because of individual differences in cases. In our Phase 2 lab surveys, about 75% of the DNA laboratories retained cuttings of biological stains or preserved DNA from cases after analysis. And over 80% of the surveyed labs returned case items to the submitting agency for preservation and storage or disposal. Most discussions of a random audit/reanalysis program as a QA tool, usually in lieu of a blind manufactured-case submission program, involve randomly selecting worked cases from the universe of already-adjudicated ones. This practice would avoid complicating a legal case with a QA measure. But for a complete audit/reanalysis, it would require that all the items the lab examined be available to the auditor and that representative examples of biological evidence stains be available for actual or potential reanalysis. A program in which cases selected for audit/reanalysis were selected from those already returned by the target laboratories to submitting agencies would be logistically complex. Procedures are different in almost every jurisdiction. Someone might have to actually go to the property or evidence room and look at the items. And there could be issues of improper handling and/or contamination by property room personnel if a reanalysis result materially differed from an original result.

Both turnaround time and CODIS present additional problems for a large-scale program. Most blind proficiency test "cases" must be low-profile to avoid raising examiner suspicions. And, in backlog situations, low-profile becomes low-priority. In our project, only two of fourteen laboratories completed cases within a month. It is true that laboratories have taken major steps to eliminate their backlogs, but it is also true that a lot more data banking and casework have been added to laboratory workloads by legislatures and other policy makers. The backlog problem is not going to disappear quickly. The CODIS system is a useful, very significant and positive development in the use of DNA profiles in law enforcement and criminal investigations. But as more labs become on-line participants and the databases and data banks grow, it becomes more of an obstacle to a blind proficiency test program involving manufactured cases and evidence. We showed in our small-scale trials that a case pair submitted to different CODIS-participating labs resulted in a cold "hit." Thus, in any large-scale program, a significant number of biological-evidence donors would be required to avoid most of the tests being revealed by CODIS matches.

It is finally noted that the problem of protection of human subjects from research risks cannot be minimized in any discussion of a large-scale program involving manufactured evidence. The DHHS regulations governing these activities have recently become more comprehensive, and the procedures for compliance more onerous. In addition, the regulations are now enforced government-wide, so that any agency actually engaged in such work must comply. Further, grantees or contractors of governmental agencies must demonstrate compliance as well. Research universities and other research institutions have most of this infrastructure in place, but most other types of organizations do not.

We have shown that external blind proficiency testing in forensic DNA laboratories is possible, and that somewhat complicated cases involving bloodstain patterns could be replicated and manufactured. Our tests were conducted in small numbers as proof of principle. To scale up to a national program involving 100–200 laboratories in one or two tests per year would be significantly costly. In addition, a number of questions would have to be decided by policy makers in consultation with the forensic-science community to define the shape of a viable, comprehensive, national program.

References

1. Peterson JL, Lin G, Ho M, Chen Y, Gaensslen RE. The feasibility of external blind DNA proficiency testing. I. Background and findings. J Forensic Sci 2003;48(1):1–11.
2. 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.

3. Technical Working Group on DNA Analysis Methods (TWGDAM). Guidelines for quality assurance program for DNA analysis. Crime Laboratory Digest 1995; 22(2):21–43.
4. US Congress. House and Senate Committees on the Judiciary. Subcommittee on Civil and Constitutional Rights and the Subcommittee on the Constitution. Forensic DNA Analysis. Hearings, June 13, 1991, No. 30, 83–315.
5. Technical Working Group on DNA Analysis Methods (TWGDAM). Guidelines for DNA proficiency test manufacturing and reporting. Crime Laboratory Digest 1994;21(2):27–32.
6. Duewer DL, Kline MC, Redman JW, Newall PJ, Reeder DJ. Interlaboratory comparison of DNA quantification practice and short tandem repeat multiplex performance with multiple-source samples. J Forensic Sci 2001;46(5):1199–210.

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