Comparison of Laboratory Performance With Blind and Mail-Distributed Proficiency Testing Samples

LOUIS C. LaMOTTE, Jr, ScD, GORDON O. GUERRANT, PhD,
D. SUE LEWIS, BS, and CHARLES T. HALL, PhD

PROFICIENCY TESTING (PT), in which samples are distributed by mail, has served as a basis for evaluating the competence of clinical laboratories for some time. The proponents of such testing have recognized its value as well as its limitations (*1–5*). In recent years the Center for Disease Control (CDC), in administering the Clinical Laboratories Improvement Act of 1967, has received some disturbing reports of serious misuse of this testing system (*6, 7*). Some laboratories reportedly have regularly sent their proficiency test samples to more competent laboratories for evaluation, so that the results have not reflected the work of the laboratory being tested. Directors of competent laboratories and others knowledgeable in the field have expressed the opinion that in many instances mailed proficiency test samples only provide a measure of the most competent employee in the laboratory rather than of the employee who routinely tests patients' specimens. These commentators contend that the performance of laboratories is thereby made to appear better than it actually is, so that physicians are falsely reassured of the accuracy of laboratory work.

The studies reported in this paper were designed to determine whether laboratories serving drug treatment centers and associated hospitals performed better with recognized proficiency test samples than with identical samples that were not recognizable as being test samples. One study was conducted in 1973 and another in 1975.

Materials and Methods

Twenty-four institutions were selected for the 1973 study. All of them provided a testing service for abused drugs and were already taking part in CDC's regular mail-distributed proficiency testing program. The laboratories that served these institutions were not licensed but were enrolled in the program as

☐ *The authors are, or at the time of the study were, in the Licensure and Proficiency Testing Division of the Bureau of Laboratories, Center for Disease Control, Atlanta, Ga. Dr. LaMotte is the director of the division. Dr. Guerrant was formerly, and Ms. Lewis is currently, in the Chemistry Section, Proficiency Testing Branch. Dr. Hall was formerly chief of that branch. The National Institute on Drug Abuse and its predecessor, the Special Action Office for Drug Abuse Prevention, collaborated in the selection of hospitals and methadone centers for the study and arranged for the cooperation of their personnel.*

Tearsheet requests to Dr. Louis C. LaMotte, Jr., Bureau of Laboratories/CDC, Bldg. 6, Rm. 154, Atlanta, Ga. 30333.

nonlicensed participants. A special comparative test was designed and scheduled to coincide with one of the 1973 regular quarterly CDC shipments of 10 urine samples. Simultaneously with this regular shipment, an identical set of 10 urine samples was provided to those hospital administrators or staff physicians who, before the shipment, had agreed to assist with the study. They were each instructed to submit the samples to their laboratories for analysis as though the samples were ordinary urine specimens from patients who were to be tested for those drugs for which the institution ordinarily requested analysis. Every effort was made to assure that the mailed samples and the "blind" samples would be exactly the same; the only difference was that one set would be recognized as test samples and the other would be disguised as patients' specimens. To encourage the institutions to participate in the study, the Center for Disease Control agreed not to disclose their identities. The detailed results were provided to the sponsoring agencies for their use in bringing about needed improvement in performance.

The drugs in the 1973 urine samples and their concentrations were as follows:

Sample No.	*Drugs in sample*	*Concentration (ug per ml)*
1	None	...
2	d-amphetamine	5.0
	Secobarbital	2.0
	Morphine	0.75
3	Phenobarbital	4.0
	Methadone	1.5
4	d-amphetamine	2.5
	Morphine	2.0
5	d-amphetamine	2.5
	Methadone	2.0
6	Secobarbital	4.0
	Methadone	1.5
	Morphine	0.5
7	Phenobarbital	2.0
	Methadone	1.5
8	d-amphetamine	5.0
	Morphine	1.0
9	Secobarbital	2.0
	Methadone	2.0
10	Morphine	2.0

These five drugs were those for which most of the 24 institutions provided a laboratory testing service. The drug concentrations in the samples were at or above readily detectable levels. The samples each contained no drug, one drug, or more than one drug, to simulate addicts' urine specimens. The samples were sterilized by filtration and shipped as a liquid to the 24 institutions. A laboratory's performance was assessed on the basis of the number of correct drug detections, provided that the laboratory offered a detection service for each of the drugs in the samples.

The 1975 study was designed to provide confirmation of the 1973 observations. Ten laboratories were selected because they served collaborating methadone treatment centers. Nine of these laboratories served 10 or more methadone centers; one laboratory served only one center. Personnel of the methadone centers were contacted to enlist their participation and obtain their agreement not to inform the selected laboratories that a blind test was in progress. Ten centers agreed to participate, but only nine returned to CDC the results sent them by the laboratories. The centers were asked to provide fictitious patient identities with each sample, to request tests for the drugs just as they ordinarily did, and to forward the laboratory's original report on each sample to CDC. The personnel of the methadone centers did not know the content of the samples.

Participating laboratories and reference laboratories had previously tested each of the urine samples used in the 1975 blind study, since these samples had been part of an earlier proficiency testing shipment. On April 3, 1975, just before the blind samples were shipped, one reference laboratory again tested each of the samples and confirmed that no change had occurred since the earlier assay by the reference laboratories. On April 18, 1975, the 10 samples were then shipped to the collaborating methadone centers. The drugs and metabolites in these samples, along with their concentrations, were as follows:

Sample No.	*Drugs in sample*	*Concentration (ug per ml)*
1	Morphine [1]	1.7
2	Morphine [1]	3.0
3	Pentobarbital	2.0
	Methadone [1]	1.4
4	Secobarbital	1.5
	Methadone [1]	1.0
5	Methamphetamine	3.0
	Benzoylecgonine	5.0
6	Morphine [1]	1.7
7	Secobarbital	3.0
	Methadone [1]	1.0
8	Pentobarbital	2.0
	Methadone [1]	2.0
9	Morphine	3.0
	Methadone [1]	1.5
10	Morphine	4.0
	Methadone [1]	1.5

[1] Sample contained primary drug as well as principal metabolite in concentrations comparable to those that would be found in an addict's urine.

Results

The performance of the participating laboratories in 1973 with mail-distributed proficiency test samples and with blind proficiency test samples (as indicated by the number of drugs each laboratory detected) is compared in table 1. In every case for which comparative data were available except one, the laboratories detected a greater percentage of the drugs in the mailed samples than in the blind samples.

Table 2 shows the errors made by the participating laboratories in 1973. Although 24 hospitals had agreed to participate in the 1973 study, for a variety of reasons not every hospital provided its laboratory's result for each sample. In a few instances, particularly for the blind samples, hospitals did not return the results, the laboratory providing the service either lost one or more samples, or the laboratory claimed that it did not offer the requested detection service for one or more of the drugs in the test sample. Accordingly, table 2 shows the number of laboratories reporting each of the listed drugs correctly, the number not detecting the drugs (false negatives), and the number erroneously reporting the presence of drugs (false positives). The total number of laboratories that tested any one sample ranged from 20 to 23.

The greater frequency of false-negative errors with the blind samples as compared with the mailed is apparent. The drug d-amphetamine was present in samples 2, 4, 5, and 8. Its frequency in the mailed test samples offered the 22 participating laboratories 88 opportunities to detect it, and all 22 succeeded in doing so; that is, in 88 reports d-amphetamine was cited as being present. In contrast, although the identical samples provided the participants in the blind part of the study 82 opportunities to detect d-amphetamine, only 39 correct detections were made; the participants missed the drug 43 times. The 22 laboratories participating in the mailed proficiency tests had 110 opportunities to detect the barbiturates phenobarbital and secobarbital, and they made 110 correct detections. The blind samples also provided 110 opportunities for detection, but in contrast, only 79 correct detections were made; the drugs were missed 31 times. The mailed proficiency test samples afforded 110 opportunities for detecting morphine, and the drug was missed only once. In the blind study, there were 111 opportunities to detect morphine, and the drug was missed 41 times. The mailed proficiency test samples afforded 110 opportunities to detect methadone, and the drug was missed twice. In the blind study, there were 106 opportunities to detect it, and the drug was missed 21 times.

The false-positive errors (table 2) were less of a problem than false-negative errors. Even so, with the mailed proficiency test samples, eight of the lab-

Table 1. Number and percentage of drugs detected by participants in 1973 study

Laboratory No.	Number of drugs to be detected [1]	Number of drugs detected		Percentage of drugs detected	
		Mailed samples	Blind samples	Mailed samples	Blind samples
1	19	19	10	100	53
2	19	18	19	95	100
3	10	10	4	100	40
4	19	19	18	100	95
5	19	19	12	100	63
6	15	15	13	100	87
7	19	19	13	100	68
8	19	19	11	100	58
9	9	9	4	100	44
10	19	19	7	100	37
11	19	19	9	100	47
12	19	19	18	100	95
13	19	19	17	100	89
14	19	19	12	100	63
15	17	17	11	100	65
16	19	19	2	100	11
17	17	17	8	100	47
18	19	19	14	100	74
19	18	18	13	100	72
20	19	...	13	...	68
21	19	19	16	100	84
22	19	19	...	100	...
23	19	...	9	...	47
24	19	19	15	100	79

[1] Each laboratory received identical samples, but each laboratory was held responsible for detecting only those drugs for which it offered a service.

NOTE: Leaders (...) indicate laboratory reported no result.

Table 2. Errors made with mailed and blind samples by 24 laboratories in 1973 comparative study

Sample No.	Drugs in sample	Mailed samples			Blind samples		
		Laboratories reporting listed drugs correctly	Laboratories reporting erroneously		Laboratories reporting listed drugs correctly	Laboratories reporting erroneously	
			False negative	False positive [1]		False negative	False positive [1]
1	None	21	...	1(1)	21	...	1(3)
2	d-amphetamine	22	...	...	13	8	...
	Secobarbital	22	...	...	15	8	...
	Morphine	22	...	...	14	9	...
	..... [2]	...	...	...	...	...	1(1)
3	Phenobarbital	22	...	...	16	6	...
	Methadone	21	1	...	14	8	...
4	d-amphetamine	22	...	...	10	11	...
	Morphine	22	...	...	18	5	...
	..... [2]	...	...	3(3)	...	...	2(2)
5	d-amphetamine	22	...	...	8	12	...
	Methadone	22	...	...	19	2	...
	..... [2]	...	...	...	...	...	1(2)
6	Secobarbital	22	...	...	16	5	...
	Methadone	22	...	...	14	7	...
	Morphine	21	1	...	9	13	...
	..... [2]	...	...	1(1)	...	...	1(1)
7	Phenobarbital	22	...	...	16	6	...
	Methadone	22	...	...	19	2	...
	..... [2]	...	...	1(1)	...	...	1(2)
8	d-amphetamine	22	...	...	8	12	...
	Morphine	22	...	...	13	9	...
	..... [2]	...	...	1(1)	...	...	1(1)
9	Secobarbital	22	...	...	16	6	...
	Methadone	21	1	...	19	2	...
	..... [2]	...	...	...	...	...	2(3)
10	Morphine	22	...	...	16	5	...
	..... [2]	...	...	1(1)	...	...	1(2)

[1] Numbers in parentheses indicate erroneous false-positive reports. [2] Various drugs were falsely reported.

oratories in the 1973 study issued eight false-positive reports. In contrast, with the blind proficiency tests, 11 laboratories issued 17 false-positive reports. The following table shows the number of times that drugs were falsely reported as being present in the 1973 and 1975 urine samples:

Reported drug	1973 samples		1975 samples	
	Mailed	Blind	Mailed	Blind
Morphine	0	4	0	0
Barbiturate	2	4	0	0
Amphetamine	0	4	1	0
Methadone	4	5	0	0
Cocaine or metabolites	2	0	1	1

Methadone was the drug most frequently cited as having been detected. Morphine was falsely reported on four occasions and benzoylecgonine on two.

In the following table the ability of the participants in the 1975 study to detect drugs in mailed and blind samples is compared.

Laboratory No.	Number of drugs to be detected [1]	Number of drugs detected		Percentage of drugs detected	
		Mailed samples	Blind samples	Mailed samples	Blind samples
1	17	13	9	76	53
2	17	16	9	94	53
3	17	17	13	100	76
4	17	16	16	94	94
5	17	17	9	100	53
6	16	14	14	88	88
7	17	17	6	100	35
8	17	17	16	100	94
9	16	16	12	100	75

[1] Each laboratory received identical samples, but each laboratory was held responsible for detecting only those drugs for which it offered a service.

With two exceptions the participating laboratories

performed better with mailed samples than with those used in the blind proficiency testing.

Results of an analysis of the errors made by the laboratories participating in the 1975 study are shown in table 3. As in the 1973 study, the differences between the laboratories' performance with the mailed and the blind proficiency test samples is readily apparent in respect to most drugs. The presence of morphine and its metabolite in samples 1, 2, 6, 9, and 10 provided the 9 participating laboratories 45 opportunities for detection in the mailed samples; the drug was detected 42 times and missed 3 times. In the blind study with the identical samples, the drug was detected only 31 times and missed 24. The barbiturates present in samples 3, 4, 7, and 8 represented 36 opportunities for detection; in the mailed samples, the barbiturates were detected 36 times, but in the blind samples they were missed 10 times. Methadone and its metabolite were present in 6 samples, representing 54 opportunities for detection; in the mailed samples this drug was missed 1 time, and in the blind samples it was detected 54 times. An amphetamine was present in only one sample. There were nine opportunities to detect it in the mailed samples, and it was missed twice, but in the blind samples it was missed six times. The cocaine derivative benzoylecgonine was present in one sample, and only seven laboratories offered a service for this drug. It was missed two times in the mailed sample and seven times in the blind sample.

The frequency of false-positive errors in the 1975 study is shown in table 3; for drugs that were falsely reported, see the first text table on page 557. There were only three false-positive reports, two in the mailed samples and one in the blind samples. Benzoylecgonine was falsely reported twice, once in the mailed samples and once in the blind samples.

Discussion

The purpose of the 1973 and 1975 studies was to compare the performance of a laboratory's staff when they knew they were working with proficiency test drug samples and when they did not know. The results suggest that in general laboratories perform much better when the staff members know they are being tested. Thus, drug samples for proficiency test-

Table 3. Errors made with mailed and blind samples by 24 laboratories in 1975 comparative study

Sample No.	Drugs in sample	Mailed samples			Blind samples		
		Laboratories reporting listed drugs correctly	Laboratories reporting erroneously		Laboratories reporting listed drugs correctly	Laboratories reporting erroneously	
			False negative	False positive [1]		False negative	False positive [1]
1	Morphine	9	...	...	3	6	...
2	Morphine	8	1	...	4	5	...
	..... [2]	...	...	1(1)	...	...	...
3	Pentobarbital	9	...	...	6	3	...
	Methadone	8	1	...	9	...	...
	..... [2]	...	...	1(1)	...	...	...
4	Secobarbital	9	...	...	7	2	...
	Methadone	9	...	...	9	...	...
5	Methamphetamine	7	2	...	3	6	...
	Benzoylecgonine	5	2	...	0	7	...
6	Morphine	8	1	...	3	6	...
7	Secobarbital	9	...	...	7	2	...
	Methadone	9	...	...	9	...	...
	..... [2]	...	...	...	...	...	(1)
8	Pentobarbital	9	...	...	6	3	...
	Methadone	9	...	...	9	...	...
9	Morphine	8	1	...	5	4	...
	Methadone	9	...	...	9	...	...
10	Morphine	9	...	...	6	3	...
	Methadone	9	...	...	9	...	...

[1] Numbers in parentheses indicate erroneous false-positive reports. [2] Various drugs were falsely reported.

ing that are distributed by mail in such a way as to be readily recognizable as test samples would be expected to result in a more optimistic appraisal of a laboratory's routine proficiency than that afforded by a blind proficiency test. This view appears to be compatible with one expressed by Gottheil and coworkers (*8*) in a recent paper on the same subject.

The kinds of errors discerned in our studies are of interest. Most errors in the blind study were of the false-negative type; that is, laboratories did not report drugs or key metabolites that the test samples contained. These same laboratories, however, correctly detected, identified, and reported drugs present in the identical samples they received in the mail. Thus, the failure to detect drugs in the blind samples was not related to a lack of technical ability. Although there were relatively few false-positive errors in the results for either the blind or the mailed samples, the false-positive reports of morphine and other drugs in these samples should be of concern to drug treatment centers. False-positive errors made in respect to specimens from patients or from suspected addicts could result in needless surveillance and an erroneous record of drug use. False-negative errors could result in a failure to provide needed surveillance and treatment for an addicted person, with adverse consequences for both the patient and society.

There appears to be little technical reason for either kind of error. The difference in the attention given the two types of samples seems to account for the difference between the performance with recognized proficiency test samples and with blind samples.

A sociological factor, it has been suggested, may account for the apparently poor performance of laboratorians with drug samples they believe to be patients' specimens. Some laboratorians have stated informally that they feel a degree of reticence about reporting the presence of drugs except in the most unequivocal cases. The laboratorian recognizes that the finding of drugs in a urine sample may represent a failure of the drug treatment center's efforts with a drug-prone person. Such evidence of treatment failure could be disconcerting to the staff of the center, which is often the contractor for the laboratory's services. Also, most assuredly a positive report is disturbing to the person suspected of drug use and to that person's family. Laboratorians may therefore subconsciously take a conservative position that appears to be expedient and refrain from reporting some of the positive results they find rather than incur the displeasure of the suspected drug user, his family, and others who have an emotional investment in the person.

Another explanation of the observed difference is that routine patient specimens and blind samples assumed to be patient specimens are processed en masse without strict quality controls. Judging from statements of laboratory directors, we assume that the special attention given to the recognized proficiency testing sample is not given to the routine patient specimen, so that errors on routine patient specimens are much more likely to go undetected.

It would be inappropriate to infer from this study with drugs that an identical disparity exists in other disciplines between the performance with proficiency test samples and specimens from patients. It is likely, however, that more attention is given to all categories of mail-distributed proficiency testing samples than to routine patient specimens. Excellent laboratories have no difficulty with recognized test samples, blind samples, or patient specimens because they use sensitive quality control procedures with all their tests. However, borderline laboratories that give less care to patient specimens than to mailed proficiency test samples are less likely to exhibit the same accuracy with both. Evaluation agencies cannot conclude that excellence with mailed proficiency test samples indicates a laboratory's performance with patient specimens.

Our studies do not negate the value of mail-distributed proficiency test samples. The single most effective way to improve laboratory performance is to motivate each laboratorian to participate in an objective assessment of his or her own competence. Mail-distributed proficiency test samples provide an effective and economical means for such an objective self-assessment and can serve as a stimulus to remedial action when the staff perceives a weakness. Also, the mail-distributed samples permit an evaluating agency to detect incompetence when it is of so glaring a degree that the persons being tested, even though they recognize that they are dealing with proficiency test samples and give special care to their analysis, are unable to perform adequately. In such cases, performance with mail-distributed proficiency test samples may be used as a basis for an adverse action against a laboratory. Many evaluation agencies, however, hesitate to use inadequate performance on mailed samples as a basis for adverse action unless there is evidence of sustained poor performance. Laboratories generally are given the opportunity to improve their performance after they receive an unfavorable report.

Obviously, a more sensitive procedure for evaluating performance is needed to complement the classic system of mailed samples. Blind samples, al-

though ideal for determining the quality of the testing given to patient specimens, are difficult to arrange routinely. For the present, the monitored onsite performance evaluation is a practical compromise. In such an evaluation system, test samples are hand-carried into each participating laboratory at least annually. The performance of the person responsible for testing a particular type of specimen is then determined in a monitored test. Samples characteristic of each discipline are similarly tested by the employees who routinely test such specimens. The intent is to determine the competence of those technologists or technicians upon whom the laboratory regularly depends for the testing of patient specimens rather than to determine the competence of the most knowledgeable employee in the laboratory. Such onsite testing has been used successfully in the excellent New York City evaluation program (*9*), in Pennsylvania (personal communication from Dr. Vern Pilcoe, director, Bureau of Laboratories, Pennsylvania Department of Health), and in Connecticut (personal communication from Dr. William F. Vincent, assistant director, Laboratory Division, Connecticut State Department of Health). The Center for Disease Control is currently field-testing an onsite performance evaluation system that may prove useful and practical for State health departments. If a practical onsite program can be developed and is put into use by other evaluation agencies, proficiency testing will become a more reliable means of evaluating clinical laboratories.

References

1. Sunderman, F. W.: The origin of proficiency testing for clinical laboratories in the United States. *In* Proceedings of 2nd National Conference on Proficiency Testing held in Washington, D.C., Oct. 6–7, 1975. National Council on Health Laboratory Services, Public Information Services, Inc., Bethesda, Md., undated, pp. 6–7.
2. National Communicable Disease Center: A guide for establishing an intrastate proficiency testing program. Public Health Service, Atlanta, Ga., August 1967.
3. National Communicable Disease Center: Satisfactory achievement in proficiency testing in clinical chemistry, hematology, and immunohematology. Public Health Service, Atlanta, Ga., Aug. 23, 1968.
4. Boutwell, J. H.: The role of proficiency testing from the perspective of the program providers—The perspective of the Center for Disease Control. *In* Proceedings of 2nd National Conference on Proficiency Testing, held in Washington, D.C., Oct. 6–7, 1975. National Council on Health Laboratory Services, Public Information Services, Inc., Bethesda, Md., undated, p. 17.
5. Prier, J. E., Sideman, L., and Yankevitch, I. J.: Clinical laboratory proficiency testing. Health Lab Sci 5: 12–18, January 1968.
6. LaMotte, L. C.: The role of proficiency testing from the perspective of the regulatory agencies—The perspective of the Center for Disease Control. *In* Proceedings of 2nd National Conference on Proficiency Testing, held in Washington, D.C., Oct. 6–7, 1975. National Council on Health Laboratory Services, Public Information Services, Inc., Bethesda, Md., undated, pp. 33–35.
7. Kachenmesiter, L. A.: The role of proficiency testing from the perspective of the participants—The perspective of a reference library. *In* Proceedings of 2nd National Conference on Proficiency Testing, held in Washington, D.C., Oct. 6–7, 1975. National Council on Health Laboratory Services, Public Information Services, Inc., Bethesda, Md., undated, pp. 21–22.
8. Gottheil, E., Caddy, G. R., and Austin, D. L.: Fallibility of urine drug screens in monitoring methadone programs. JAMA 236: 1035–1038, Aug. 30, 1976.
9. Blatt, S.: The role of proficiency testing from the perspective of the regulatory agencies—The perspective of New York City Department of Health. *In* Proceedings of 2nd National Conference on Proficiency Testing, held in Washington, D.C., Oct. 6–7, 1975. National Council on Health Laboratory Services, Inc., Bethesda, Md., undated, pp. 41–42.

SYNOPSIS

LaMOTTE, LOUIS C., Jr. (Center for Disease Control), GUERRANT, GORDON O., LEWIS, D. SUE, and HALL, CHARLES T.: *Comparison of laboratory performance with blind and mail-distributed proficiency testing samples. Public Health Reports, Vol. 92, November–December 1977, pp. 554–560.*

Simulated addict urine samples containing drugs were sent to collaborating hospital administrators and officials of methadone centers, who then forwarded the samples to their supporting laboratories as though they were ordinary specimens from patients. The laboratories, which were already participating in the proficiency testing program of the Center for Disease Control, received the identical test samples in the mail as part of a regular Center for Disease Control proficiency testing program. Most of the laboratories performed acceptably with the mail-distributed samples, but many performed poorly when the identical samples were sent to them as if they were specimens from patients. Because of the limitations of proficiency testing involving mail-distribution samples and the impracticality of extensive testing with blind samples on a national level, the Center for Disease Control proposes to compliment its regular proficiency testing program with a monitored, onsite program of performance evaluation.