Variation in Proficiency Testing Performance by Testing Site

Tina M. Stull, MD; Thomas L. Hearn, PhD; John S. Hancock; James H. Handsfield, MPH; Carlyn L. Collins, MD, MPH

Context.—Congress enacted the Clinical Laboratory Improvement Amendments of 1988 (CLIA) to promote uniform quality and standards among all testing sites in the United States. The performance indicators specified in the legislation are proficiency testing (PT) performance and periodic inspections.

Objective.—To evaluate variation in PT performance by type of testing facility during the first year of compulsory participation under CLIA.

Design.—All 1994 PT score data electronically reported to the Health Care Financing Administration as a component of compliance with the CLIA regulations were obtained. Over 1.2 million PT event scores from 17,058 unique testing sites were sorted into 2 groups based on the type of testing facility: hospitals and independent laboratories (HI) and all other testing sites (AOT).

Main Outcome Measures.—Satisfactory and unsatisfactory performance rates for HI and AOT for each analyte and/or test, according to the criteria specified by the CLIA regulations.

Results.—The aggregate rates of satisfactory event performance for all regulated analytes, tests, and specialties were 97% and 91% for the HI and AOT groups, respectively. The aggregate odds ratio for unsatisfactory PT event performance for the AOT group compared with the HI group was 2.89, with a range of 2.19 to 7.51 for the individual analytes.

Conclusion.—There was a consistent difference in PT performance during the first full year of compulsory PT under the CLIA regulations based on the type of testing facility performing the analysis. Traditional testing sites achieved higher rates of satisfactory performance than newly regulated, alternative testing sites.

The CLIA legislation established federal regulation of all sites performing testing on “materials derived from the human body for the purpose of providing information for the diagnosis, prevention, or treatment of any disease . . . [of] human beings.” Although some state regulation of laboratories existed before CLIA was enacted, only traditional testing sites such as hospitals and independent laboratories, which were engaged in interstate commerce or receiving Medicare payments, were subject to federal regulation. Therefore, the CLIA legislation imposed uniform, site-neutral laboratory practice standards for the first time on the other 90% of sites performing clinical laboratory testing. Inherent in the site-neutrality model on which the CLIA regulations are based is this notion: the quality of laboratory test results, as measured by accuracy and reliability, should be equivalent irrespective of the testing site. The CLIA model specifies minimum standards for quality control, quality assurance, and personnel requirements, which were designed to ensure site-neutral test comparability. Proficiency testing performance and inspection of testing sites are the 2 quality indicators specified in the CLIA model.

The Health Care Financing Administration (HCFA) indicates there are 157,002 sites currently performing clinical laboratory testing in the United States (HCFA Online Survey, Certification and Reporting database, unpublished).
Proficiency testing event scores for the 3 testing events of 1994 were obtained from HCFA for all testing sites participating in the following DHHS-approved PT programs: the American Association of Bioanalysts, College of American Pathologists, External Comparative Evaluation for Laboratories, American Academy of Pediatric Physicians, American Academy of Family Physicians, American Osteopathic Association, and Medical Laboratory Evaluation programs. Over 1.2 million individual test event scores were sorted into 2 groups: (1) hospitals and independent laboratories (HI) and (2) other testing sites (AOT). Scores were grouped by merging the PT data with the HCFA Online Survey, Certification and Reporting database, a comprehensive database used by HCFA to administer the CLIA program. Satisfactory and unsatisfactory performance rates were determined for the 2 laboratory groups for each of the analytes, tests, or specialties according to the criteria specified in the CLIA regulations where failure to attain an overall testing event score of at least 80% is unsatisfactory performance. χ² test statistics and logit odds ratios were calculated for each analyte, test, or specialty using SAS statistical software. An aggregate odds ratio, including all reported testing events for all analytes, tests, and specialties, was also calculated for both laboratory groups.

RESULTS

A total of 17,058 unique testing sites were included in this study. These testing sites represent all the participants in the 7 PT programs included in the study, which designated HCFA as a recipient of their PT results as part of their compliance with the CLIA program in 1994. Forty-three percent of the testing sites were in the HI group; the remaining 57% were in the AOT group. The percentage of unsatisfactory test event scores for the 30 most commonly offered, regulated analytes, tests, or specialties in AOT ranged from 1.3% to 5.6% for the HI

<table>
<thead>
<tr>
<th>Analyte, Test, or Test Specialty</th>
<th>% Unsatisfactory Event Performance at All Other Testing Sites</th>
<th>% Unsatisfactory Event Performance at Hospitals and Independent Laboratories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>15.0</td>
<td>5.1</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>13.1</td>
<td>3.6</td>
</tr>
<tr>
<td>Hematocrit</td>
<td>13.1</td>
<td>4.1</td>
</tr>
<tr>
<td>Thyroxine (T₄)</td>
<td>11.9</td>
<td>4.5</td>
</tr>
<tr>
<td>High-density lipoprotein cholesterol</td>
<td>11.8</td>
<td>5.6</td>
</tr>
<tr>
<td>Potassium</td>
<td>11.5</td>
<td>1.7</td>
</tr>
<tr>
<td>Urea nitrogen</td>
<td>11.4</td>
<td>3.7</td>
</tr>
<tr>
<td>Sodium</td>
<td>9.7</td>
<td>3.6</td>
</tr>
<tr>
<td>Creatine phosphokinase</td>
<td>9.6</td>
<td>3.1</td>
</tr>
<tr>
<td>Total bilirubin</td>
<td>9.5</td>
<td>2.4</td>
</tr>
<tr>
<td>Triglycerides</td>
<td>9.4</td>
<td>2.5</td>
</tr>
<tr>
<td>Albumin</td>
<td>9.3</td>
<td>2.6</td>
</tr>
<tr>
<td>Amylase</td>
<td>9.2</td>
<td>3.3</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>9.1</td>
<td>2.1</td>
</tr>
<tr>
<td>Aspartate aminotransferase</td>
<td>8.9</td>
<td>2.7</td>
</tr>
<tr>
<td>Theophylline</td>
<td>8.5</td>
<td>1.7</td>
</tr>
<tr>
<td>Creatinine</td>
<td>8.3</td>
<td>2.3</td>
</tr>
<tr>
<td>Lactate dehydrogenase</td>
<td>8.3</td>
<td>3.0</td>
</tr>
<tr>
<td>Uric acid</td>
<td>8.3</td>
<td>2.1</td>
</tr>
<tr>
<td>Alanine aminotransferase</td>
<td>8.1</td>
<td>2.6</td>
</tr>
<tr>
<td>Calcium</td>
<td>8.1</td>
<td>2.4</td>
</tr>
<tr>
<td>Alkaline phosphatase</td>
<td>7.6</td>
<td>2.4</td>
</tr>
<tr>
<td>Bacteriology</td>
<td>7.2</td>
<td>3.4</td>
</tr>
<tr>
<td>Total protein</td>
<td>7.1</td>
<td>2.0</td>
</tr>
<tr>
<td>Platelet count</td>
<td>7.0</td>
<td>2.8</td>
</tr>
<tr>
<td>White blood cell differential</td>
<td>7.0</td>
<td>2.6</td>
</tr>
<tr>
<td>Prothrombin time</td>
<td>6.7</td>
<td>2.3</td>
</tr>
<tr>
<td>White blood cell count</td>
<td>6.7</td>
<td>2.6</td>
</tr>
<tr>
<td>Red blood cell count</td>
<td>6.6</td>
<td>3.0</td>
</tr>
<tr>
<td>Mononucleosis antibodies</td>
<td>3.6</td>
<td>1.3</td>
</tr>
</tbody>
</table>
group and from 3.6% to 15.0% the AOT group. The aggregate rate of satisfactory test event performance for all regulated analytes, tests, and specialties was 97% for the HI group and 91% for the AOT group. A complete listing of the percentage of unsatisfactory test event scores for the 30 most commonly offered, regulated analytes, tests, or specialties in AOT is shown by testing site group in Table 1. In each case, the percentage of unsatisfactory test event scores was higher for the AOT group than for the HI group. When the testing site is the unit of analysis instead of the analyte or test, 73% of HI and 68% of AOT passed all challenges undertaken in 1994. The overall site scores for all 1994 PT challenges are shown by testing group in the Figure.

Odds ratios indicating the estimated relative risk for an unsuccessful PT challenge among the AOT group compared with the HI group were calculated for the 30 most commonly offered, regulated analytes, tests, and specialties among AOT. The odds ratios for these 30 procedures ranged from a high of 7.51 for potassium to a low of 2.19 for bacteriology; all 30 odds ratios were significant at a 95% confidence level. The aggregate odds ratio in favor of unsatisfactory PT performance for all regulated analytes, tests, and specialties for the AOT group when compared with the traditional laboratories of the HI group was 2.89; for an observed unsatisfactory event score, we estimated the odds were 2.89 times more likely that the score was received from an AOT than from an HI. A complete list of the logit odds ratios in favor of unsatisfactory PT performance for the 30 most commonly offered, regulated analytes, tests, and specialties among AOT compared with HI is shown in Table 2.

### COMMENT

The results of this analysis indicate disparate PT performance between traditional laboratories and alternative testing sites. Having established a difference in the PT performance of these 2 groups, the implications of this finding must be considered. However, to adequately evaluate the findings of this report, it is essential to establish the context in which the observations were made. Therefore, a brief review of the history, attributes, limitations, and functions of PT is presented.

Organized PT in the United States originated in 1946 with a small regional program in Philadelphia, Pa. Soon after, the College of American Pathologists conducted the first 2 nationwide PT surveys in 1947 and 1948, and the national Sunderland Proficiency Testing Service was established in 1949. These early surveys demonstrated a disturbing degree of interlaboratory variation in test results among the participating laboratories. The professional laboratory community responded quickly by voluntarily espousing PT as an educational tool and an external quality control tool to help identify testing problems in the clinical laboratory.

The participation in, and growth of, PT programs paralleled the growth of clinical laboratory testing in general. This growth was reinforced by the collective experience of laboratory professionals and PT providers who found PT to be serving the important function of bringing to light otherwise unnoticed or undetected problems with laboratory testing as evidenced by this compelling finding: PT performance improved over time. This finding appeared to indicate improved interlaboratory quality, and presumably intralaboratory quality, due at least in part to participation in PT programs.

Proficiency testing assumed a new role as a federal regulatory tool with the introduction of CLIA in 1967. The impact of this new role was mitigated by 2 factors: the law applied only to laboratories engaged in interstate commerce, and many of these laboratories were already voluntarily participating in PT. Enactment of the CLIA legislation in 1988, however, broadened the role of PT further by creating a uniform set of standards for all sites performing clinical laboratory testing.

The 1988 CLIA legislation specifies that the standards applied to testing sites will “assure consistent performance by laboratories . . . of valid and reliable laboratory examinations . . .” and further stipulates the standards will take into consideration issues related to the complexity of the testing procedures performed by testing sites. Therefore, CLIA standards are based on the parallel principles of site neutrality and the complexity model of testing. The CLIA regulatory model stipulated in the legislation specifies the comparability of laboratory performance, which is inherent in the concept of site neutrality, is to be monitored by PT performance and periodic on-site inspections of testing facilities.

The scope of PT is limited to the analytic portion of the total testing process. Proficiency testing cannot and does not directly address some important pre- and postanalytic steps such as specimen collection, processing and storage, and customary reporting procedures, which contribute to the overall accuracy and reliability of test results reported by a testing site. The relative importance of the preanalytic and postanalytic processes in the overall accuracy and reliability of final test results reported by a testing site should not be underestimated. In a review of incidents captured by one hospital laboratory’s quality assurance program, Ross and Boone reported more than 90% of the detected problems occurred in either the pre- or postanalytic phase of the total testing process. A study of transfusion medicine monitoring practices by Boone et al demonstrated a similar pattern, with 95% of the reported defects occurring in either the pre- or postanalytic phase of testing. In a recent investigation of laboratory problems detected in office-based, primary care practices, Nutter et al reported a comparable pattern, with 93% of detected problems attributed to either pre- or postanalytic processes.

Also to be considered when evaluating PT performance is the testing event itself. A PT event is a nonrandom sample of the work performed at a given testing site and is therefore subject to all the limitations and biases inherent in such a process. The issue of how representative PT performance is of routine laboratory performance has been addressed by several investigators. In a survey conducted by Cembrowski and Vanderlinde, most respondents reported using extraordinary testing pro-

---

**Results**

<table>
<thead>
<tr>
<th>Analyte, Test, or Test Specialty</th>
<th>Odds Ratio (95% Confidence Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Potassium</strong></td>
<td>7.51 (6.59-8.54)</td>
</tr>
<tr>
<td><strong>Theophylline</strong></td>
<td>5.55 (4.54-6.64)</td>
</tr>
<tr>
<td><strong>Hemoglobin</strong></td>
<td>4.56 (4.06-5.09)</td>
</tr>
<tr>
<td><strong>Uric acid</strong></td>
<td>4.32 (3.78-4.94)</td>
</tr>
<tr>
<td><strong>Total bilirubin</strong></td>
<td>4.28 (3.78-4.85)</td>
</tr>
<tr>
<td><strong>Cholesterol</strong></td>
<td>4.04 (3.65-4.46)</td>
</tr>
<tr>
<td><strong>Triglycerides</strong></td>
<td>4.02 (3.55-4.55)</td>
</tr>
<tr>
<td><strong>Total protein</strong></td>
<td>3.85 (3.33-4.45)</td>
</tr>
<tr>
<td><strong>Albumin</strong></td>
<td>3.61 (3.34-3.95)</td>
</tr>
<tr>
<td><strong>Creatinine</strong></td>
<td>3.76 (3.33-4.24)</td>
</tr>
<tr>
<td><strong>Calcium</strong></td>
<td>5.38 (3.13-10)</td>
</tr>
<tr>
<td><strong>Aspartate aminotransferase</strong></td>
<td>3.57 (3.15-4.04)</td>
</tr>
<tr>
<td><strong>Alanine aminotransferase</strong></td>
<td>3.52 (3.09-3.99)</td>
</tr>
<tr>
<td><strong>Alkaline phosphatase</strong></td>
<td>3.38 (2.95-3.86)</td>
</tr>
<tr>
<td><strong>Urea nitrogen</strong></td>
<td>3.38 (3.06-3.74)</td>
</tr>
<tr>
<td><strong>Alanine aminotransferase</strong></td>
<td>3.31 (2.91-3.76)</td>
</tr>
<tr>
<td><strong>Creatine phosphokinase</strong></td>
<td>3.28 (2.86-3.75)</td>
</tr>
<tr>
<td><strong>Glucose</strong></td>
<td>3.28 (3.02-3.56)</td>
</tr>
<tr>
<td><strong>Prothrombin time</strong></td>
<td>3.12 (2.70-3.59)</td>
</tr>
<tr>
<td><strong>Lactate dehydrogenase</strong></td>
<td>2.96 (2.50-3.39)</td>
</tr>
<tr>
<td><strong>Amylase</strong></td>
<td>2.95 (2.50-3.45)</td>
</tr>
<tr>
<td><strong>Sodium</strong></td>
<td>2.91 (2.61-3.24)</td>
</tr>
<tr>
<td><strong>Mononucleosis antibodies</strong></td>
<td>2.60 (2.34-2.63)</td>
</tr>
<tr>
<td><strong>Thyroxine (T4)</strong></td>
<td>2.34 (2.51-3.21)</td>
</tr>
<tr>
<td><strong>White blood cell differential</strong></td>
<td>2.62 (2.50-3.18)</td>
</tr>
<tr>
<td><strong>White blood cell count</strong></td>
<td>2.71 (2.42-3.03)</td>
</tr>
<tr>
<td><strong>Platelet count</strong></td>
<td>2.61 (2.34-2.91)</td>
</tr>
<tr>
<td><strong>Red blood cell count</strong></td>
<td>2.31 (2.06-2.58)</td>
</tr>
<tr>
<td><strong>High-density lipoprotein</strong></td>
<td>2.28 (2.05-2.53)</td>
</tr>
<tr>
<td><strong>Cholesterol</strong></td>
<td>2.19 (1.97-2.43)</td>
</tr>
</tbody>
</table>
tocols, ie, above and beyond those used by the laboratory for routine patient specimens when testing PT samples. Studies conducted in 1977 and 1982 by the Centers for Disease Control and Prevention found participant laboratories’ performance on mailed PT samples was superior to that observed when blind PT samples were tested by the same laboratories.13,14 These findings suggest PT performance represents the best analytical work a laboratory is capable of producing. A study by Jenny and Jackson, however, found PT performance to be a reliable predictor of routine patient testing for theophylline. Another recent study by the Model Performance Evaluation Program at the Centers for Disease Control and Prevention found comparable error rates for blind and open samples tested for human immuno-deficiency virus antibodies. One plausible explanation for these seemingly contradictory findings is the impact of the overall accuracy of testing for a given analyte on the predictive value of PT, ie, the impact of the “prevalence” of inaccurate tests on the sensitivity and specificity of PT as an instrument to detect the inaccuracies.

Several additional variables unique to the PT process, and often unrelated to the testing site’s routine performance level, can contribute to unsatisfactory PT performance. Examples of these include, but are not limited to, the presence of PT report forms, use of differing units of measure, and matrix effects wherein the PT sample medium has an effect on analyte measurement.17-20 Several investigators have noted the relative contributions of these variables on the outcome of a PT event.21-24

Although there are limits to the sensitivity and specificity of PT in detecting substandard clinical laboratory performance, PT is clearly an important tool in achieving this end. However, the relationship of participation in a PT program to improved laboratory performance relies entirely on the response initiated in the testing site when a potential problem is identified by unsatisfactory PT performance. Throughout the 50-year history of PT in the United States, the relevant literature has been peppered with reports demonstrating improved PT performance over time.20-22 These studies are testimony to the successful commitment of the laboratory community to implement PT as a tool to improve the overall quality of clinical laboratory testing. Some have argued that improved PT performance does not reflect improved clinical laboratory performance, but rather the process of laboratories becoming acclimated to the PT process and its intrinsic subtleties. Several empirical studies, however, indicate this is not the case. Hoeltge and Duckworth22 in 1987 reviewed the PT performance of laboratories accredited by the College of American Pathologists over a 2-year period. Analytes were targeted by the investigators for a review by the respective laboratory directors on the basis of repeated unacceptable results in the College of American Pathologists Survey Program. The study reported 50% of the investigations revealed a problem in instrumentation, methods, or other technical aspects of testing; only 4.4 of the target cases were attributed to the survey format or materials. The investigations spurred by unacceptable PT performance in this study led to improved performance with 88% of the targeted analytes on subsequent PT evaluations.21 In 1994, Ehrmeyer et al23 with the Wisconsin State Laboratory of Hygiene demonstrated similar findings with unsatisfactory PT performance, bringing attention to correctable problems in the laboratory testing process, and a pattern of improved PT performance after a prior unsatisfactory evaluation. A review of PT performance and experience with the College of American Pathologists’ programs demonstrated “consistent and statistically significant improvement in performance for the first 3 to 4 years of proficiency testing.”22 The report notes PT error rates continue to decline over time, suggesting the improved PT performance represents true improvement in laboratory performance and not acclimation to the PT process. Additional studies support the notion of PT participation contributing to an overall improvement in laboratory performance by demonstrating a positive association between participation in PT and decreasing interlaboratory SDs and coefficients of variation with PT samples.24,25 In evaluating the findings of our study, it appears there is room for improvement, particularly among the alternative testing sites. The unsatisfactory test event performance rates for the 3 most commonly offered and regulated tests and specialties among the AOT are particularly notable: glucose, 15.0%; hemoglobin, 9.1%; and bacteriology, 7.2%. The less than optimal PT performance of the alternative testing sites, most of which were not previously regulated, is not unexpected; other researchers have reported similar findings with newly regulated laboratories.22,24 Some of the observed difference in the performance of the alternative testing sites vs the hospital and independent laboratories is probably attributable to previously discussed factors, which are not directly related to the quality of daily performance. However, given the consistently superior performance of the traditional laboratories across all analytes, and the magnitude of the observed odds ratios, it is likely the disparity also represents a difference in the actual quality of daily laboratory performance between the 2 groups. It is important to consider why this performance disparity exists. Like CLIA, prior state and federal clinical laboratory regulations have stressed the importance of personnel standards, quality control and quality assurance programs, and participation in PT. Therefore, the staffs of previously regulated laboratories generally include laboratory professionals who have had the benefit of specific training in each of these areas. Previously unregulated, alternative testing sites, however, may not have this advantage. HCFA indicates the staffs of alternative testing sites are less likely to include a laboratory professional than hospital and independent laboratories (HCFA Online Survey, Certification and Reporting database, unpublished data, January 1997). The absence of a laboratory professional in a testing site may leave the site in the undesirable position of having the best of intentions, but a lack of expertise to carry the intentions to fruition. Similarly, many alternative testing sites are directed by physicians who may not have been exposed to quality laboratory practice principles during their training. A recent study by Ferris et al25 showed clinical laboratory training was available in only 15% to 60% of primary care residency programs. Other economic, managerial, and technical factors, such as collective differences in the testing methods used by different types of testing facilities, may also contribute to the performance differential observed between the traditional laboratories and the alternative testing sites.

Like most investigations, this study presents new challenges. The common ground between the enactment of the CLIA legislation and the health care professionals affected by the CLIA standards is a fundamental commitment to deliver the best possible care to patients. Results of this study demonstrate the PT performance among the alternative testing sites is less satisfactory than that of traditional hospital and independent laboratories. Given the body of evidence in the peer-reviewed literature, it is prudent to theorize PT performance does at least partially reflect daily clinical laboratory performance. If all testing sites were to provide comparably accurate and reliable test results, there is clearly work to be done. The laboratory and health care community at large must work together to assure all individuals involved...
in the performance of clinical laboratory testing have the requisite knowledge and experience to provide optimally accurate and reliable test results. A “positive” PT result, i.e., one indicating unsatisfactory performance, provides information that may not have been known otherwise. Just as a positive laboratory test result, in the hands of an adequately trained and knowledgeable clinician, can be an opportunity for positive intervention in a patient’s condition, unsatisfactory PT can be an opportunity for the adequately trained and knowledgeable clinical laboratory testing site staff to improve the accuracy and reliability of patient test results. Both individual and specific groups of testing sites can benefit from continued monitoring of PT performance to help to assess the success or failure of intervention efforts to improve the overall quality of clinical laboratory testing.

The authors appreciate the assistance of Steven J. Steindel, PhD, as an expert reviewer of the manuscript and its contents.

References