Are Physicians' Office Laboratory Results of Comparable Quality to Those Produced in Other Laboratory Settings?

Jerry Hurst, MA; Karen Nickel, PhD; Lee H. Hilborne, MD, MPH

Context.—In 1995, California adopted a bill that brought laboratory laws in line with the 1988 Clinical Laboratory Improvement Amendments' standards for clinical laboratories and mandated a study comparing results in physicians' office laboratories (POLs) with other settings.

Objective.—To determine whether persons conducting tests in POLs produce accurate and reliable test results comparable to those produced by non-POLs.

Design.—Survey of clinical laboratories using proficiency testing data.

Setting.—All California clinical laboratories participating in the American Association of Bioanalysts proficiency testing program in 1996 (n=1110).

Main Outcome Measures.—“Unsatisfactory” (single testing event failure) and “unsuccesful” (repeated testing event failure) on proficiency testing samples.

Results.—The unsatisfactory failure rate for POLs was nearly 3 times (21.5% vs 8.1%) the rate for the non-POLs and about 1.5 times (21.5% vs 14.0%) for POLs that used laboratory professionals as testing or supervisory personnel (P<.001). The POL unsuccessful rate was more than 4 times (4.4% vs 0.9%) the rate for non-POLs and more than twice (4.4% vs 1.8%) the rate for the POLs using laboratory professionals (P<.001).

Conclusions.—Significant differences exist among POLs, POLs using licensed clinical laboratory scientists (medical technologists), and non-POLs. Testing personnel in many POLs might lack the necessary education, training, and oversight common to larger facilities. We must better understand the contributing factors that result in the poorer results of POLs relative to non-POLs. In the meantime, patients should be aware that preliminary findings suggest that differences in quality of laboratory tests based on testing site may exist. Laboratory directors at all testing sites must ensure that they understand laboratory practice sufficiently to minimize errors and maximize accuracy and reliability. Directors must understand their obligation when they elect to oversee those assigned testing responsibility. Legislators may wish to reconsider the wisdom of further easing restrictions on those to whom we entrust our laboratory specimens.

JAMA. 1998;279:468-471

THE IMPORTANCE of quality assurance in clinical laboratory testing has been recognized for many years, long before discussions of quality in other areas of medicine were considered. Intralaboratory and interlaboratory proficiency testing are the hallmark of laboratory quality assurance efforts; these programs make it possible for the public to be assured of accurate and precise results irrespective of where their tests are performed. Ten years ago the Wall Street Journal released findings suggesting that, at least for Papanicolaou smears, the quality of laboratory testing may not be as good as expected.12 Responding to these findings, Congress adopted the Clinical Laboratory Improvement Amendments of 1988 (CLIA 88) that mandated compliance with national quality standards as defined by the CLIA 88 regulations. CLIA 88 established minimum acceptable criteria for all facilities performing all classifications of clinical testing (ie, waived tests, provider-performed microscopy, and moderate-complexity and high-complexity tests).

See also pp 463 and 473.

Over the last 10 years, laboratory instrumentation has become much more sophisticated. Physicians who find it desirable to perform their patients' laboratory tests immediately in a physicians' office laboratory (POL) rather than sending patients' specimens to larger reference or hospital-based laboratories can now avail themselves of automated analyzers that perform a broad spectrum of tests. Devices proximal to the patient allow more rapid turnaround time with presumably the same level of quality one would expect of larger hospital and reference laboratories.

In 1995, California adopted Senate Bill 113, legislation that, among other things, brought California laboratory laws in line with CLIA standards. Although California historically has had stringent laboratory testing personnel standards, pressures existed to reduce some of these standards, particularly for POLs (in California, defined as 5 or fewer physicians performing tests on their own pa-
tients). The Senate Bill 113 revision permitted “any other person within a physician office laboratory” to perform testing when appropriately supervised by the patient’s physician.

Although the non-POL community expressed concern about the diminution of laboratory personnel qualifications in POLs, a compromise was reached. The legislature permitted “any other persons” in POLs to perform tests under the supervision of the physician responsible for the test while at the same time instructing the California Department of Health Services (DHS) to “conduct a study to determine whether persons conducting tests in physician office laboratories . . . produce accurate, reliable, and necessary test results comparable to those produced by other persons performing moderate or high complexity testing, or both.” The results of this mandated study constitute the basis for this article.

The DHS sought input from the Clinical Laboratory Technology Advisory Committee, a multidisciplinary committee constituted to advise the DHS on laboratory practice issues. Proficiency testing data were recommended to measure accuracy and reliability of testing because results could be evaluated for closeness to expected values. All laboratories performing moderate- and high-complexity testing are required to participate in proficiency testing (PT) and report their results to the state. The Clinical Laboratory Technology Advisory Committee advised the DHS to consider inspection data as a surrogate for “reliability” because laboratories that perform poorly as determined by on-site inspections would be more likely to produce unreliable results. However, 40% of California laboratories performing moderate- and high-complexity testing, as defined by the Centers for Disease Control and Prevention (CDC), were accredited by other agencies in 1996. Consequently, their inspection data were unavailable to the state, limiting the value of these data for evaluation.

This study compares the quality of laboratory testing, measured by PT scores, in 3 groups of laboratories: licensed California clinical laboratories (non-POLs); POLs that retain the services of licensed clinical laboratory scientists (medical technologists) (CLS/MTs) either as testing personnel, supervisory personnel, or laboratory consultants; and POLs that do not employ CLS/MTs to perform laboratory testing.

METHODS

The study was limited to the review of 1996 PT performance data for 11 analytes that are commonly performed in both POLs and non-POLs (Table 1). Analytes were chosen because they are clinically important and widely ordered by physicians. They are used for both preliminary patient screening and monitoring for common clinical conditions.

PT Provider Selection

Although all laboratories must enroll in an approved PT program for each of the 11 analytes they elect to perform, each laboratory can select from a list of PT providers approved by California. Three California-approved PT providers (ie, American Association of Bioanalysts [AAB], American Proficiency Institute, and Medical Laboratory Evaluation) enrolled the majority of POLs within California in 1996 and, therefore, were considered for this study. However, the largest numbers of POLs were enrolled in the AAB program; additionally, the relative numbers of POLs and non-POLs in this program were about evenly divided and the total number of laboratories was sufficient to yield statistically significant data. Consideration was given also to the availability of PT scores; of the 3 candidate organizations, only AAB had supplied complete 1996 data at the time the study was initiated. Based on these factors, it was decided to use the data provided by AAB.

Selection of POLs and Non-POLs

Strict adherence to the California definitions of POLs and non-POLs was the only criterion used to differentiate these 2 groups. Although CLIA 88 does not define a POL, the CLIA application form requires a laboratory to designate itself as a POL or 1 of 29 other “facility types”; this application information was used in the initial sorting process. Subsequent sorting was done to ensure that the facilities in each group met the appropriate California requirements. Other factors, such as the annual number of tests performed by the laboratory, types of analytes tested, and physical location of the testing facility, were not used to categorize laboratories because PT enrollment and participation requirements are independent of these factors.

Laboratories that did not perform moderate- or high-complexity testing were deleted from the study population. Similarly, laboratories that did not test for 1 or more of the 11 chosen analytes were excluded. Finally, the resulting lists were checked against California records to ensure that each laboratory was active in 1996 and any inactive laboratory was deleted from the study.

Of the remaining 291 POLs, 288 were reached by telephone and asked to provide information regarding their laboratory testing personnel in 1996. Those that used licensed CLS/MTs as testing persons, supervisors, or consultants were placed into a distinct cohort. The final distribution of the 288 eligible laboratories is shown in Table 2.

Of the 1110 laboratories enrolled in the AAB PT program, the 725 that were included were ultimately divided into the 3 groups (see Table 2 for final distribution).

PT Performance Evaluation

We evaluated each of the 3 laboratory groups for PT performance in each of the following categories: overall rates of “unsatisfactory” performance, overall rates of “unsuccessful” performance, and rates of “unsatisfactory” performance by testing event for each group of hematometry and chemistry analytes combined.

The AAB PT participants received a set of 5 unknown samples for each analyte for each testing event they were enrolled in during 1996. “Unsatisfactory” performance is defined for the analytes used in this study as a score of less than 80% for any given analyte during any single testing challenge (ie, less than 4 acceptable results for each set of 5 unknown specimens). “Unsuccessful” performance means 2 or more consecutive unsatisfactory scores or 2 unsatisfactory scores of any 3 consecutive testing events for each analyte. The failure rates

Table 1.—Analytes Used for Comparison

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>No. of Facilities</th>
<th>Total Analyte Challenges</th>
<th>Average No. of Analytes per Facility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physicians' office laboratories</td>
<td>159</td>
<td>656</td>
<td>4.1</td>
</tr>
<tr>
<td>Physicians' office laboratories using clinical laboratory scientists (medical technologists)</td>
<td>129</td>
<td>662</td>
<td>5.1</td>
</tr>
<tr>
<td>Non–physicians’ office laboratories</td>
<td>437</td>
<td>2991</td>
<td>6.8</td>
</tr>
</tbody>
</table>

©1998 American Medical Association. All rights reserved.
for each of these 5 categories were calculated as percentages of the applicable totals and compared for each of the 3 groups. The statistical significance of each failure rate comparison was evaluated for goodness of fit applying the χ² statistic using computerized statistical software (ProStat, Poly Software International Inc, Salt Lake City, Utah).

Overall failure rates for each group were calculated as percentages using the total number of analytes tested in 1996 as the sample. To calculate the overall unsatisfactory performance, we used the total number of analytes with 1 or more scores of less than 80% within each group. For determining the overall unsuccessful performance, we used the total number of analytes that had 2 or more unsuccessful scores.

When calculating failure rates by test event for groups of analytes (eg, chemistry and hematology), the unit of observation is a single testing challenge (ie, a laboratory may have up to 3 observations in this analysis for any 1 analyte during 1996). The numbers varied slightly for each testing event within each cohort due to variability in individual laboratory enrollment and participation. The unsatisfactory rate is the percentage of testing challenge failures within the analyte group.

RESULTS

Table 3 shows the overall unsatisfactory and unsuccessful performance rates for each group. The unsatisfactory performance failure rate for POLs was nearly 3 times as great as for the non-POLs and about 1.5 times that of the POLs that used CLS/MTs as either testing or supervisory personnel (P < .001). Although this latter group showed a significant improvement over POLs not having input from CLS/MTs (P < .001), they still had nearly twice the failure rate compared to non-POLs. The unsuccessful performance failure rates demonstrated similar findings. The POL failure rate was over 4 times that of the non-POLs and more than twice that of the POLs using CLS/MTs (P < .001). Although POLs using CLS/MTs performed statistically better than POLs without this professional input, they still showed twice the failure rate compared to non-POLs (P = .02).

Table 4 shows the unsatisfactory failure rates of each cohort for each testing event when chemistry and hematology analytes were combined into their respective CLIA specialty. For chemistry analytes, POLs had failure rates 2 to 3 times that of the non-POL group (P < .001). Physicians’ office laboratories using CLS/MTs, when compared to non-POLs, also had significantly higher failure rates (P < .001) for testing events 1 and 3. Similar findings were observed for hematology analytes. Physicians’ office laboratory failure rates were 4 to 5 times greater than non-POLs for test events 1 and 3 (P < .001) and, although not statistically significant, twice as high for testing event 2. Physicians’ office laboratories using CLS/MTs, compared to the non-POLs, had failure rates significantly higher for the third testing event (P < .001).

The overall rates of unsatisfactory performance for each PT testing event showed similar findings for the 3 laboratory groups. The POLs had approximately 3 times the failure rate for each of the 3 testing challenges in 1996 compared to non-POLs (8.5% vs 2.5%, 8.5% vs 2.3%, and 10.8% vs 3.8%; P < .001). The POLs using CLS/MTs had unsatisfactory rates approximately twice that of the non-POLs for the first and third testing challenges (5.6% vs 2.5%, 7.2% vs 3.8%; P < .001); the differences for the second challenge were not statistically significant (4.1% vs 3.3%, P = .29).

Differences at the chemistry and hematology group levels are similar to those discussed above (Table 4). Although not explicitly presented, individual analytes showed generally similar findings.

COMMENT

This study examined PT performance in 3 distinct groups: POLs, POLs that use CLS/MTs to perform or oversee laboratory testing, and non-POLs. Although 95.6%, 98.2%, and 99.1% of POLs, POLs using CLS/MTs, and non-POLs, respectively, were “successful” in their PT efforts, findings demonstrate statistically significant differences among these 3 groups with respect to PT performance. The “unsuccessful” failure rates were of particular interest because they represent significant repeated failures on the part of the laboratory and indicate noncompliance with state and federal law.

Only a few studies have examined the relationship between testing personnel and laboratory quality. Lunz et al studied PT data and showed that those laboratories that employed American Society of Clinical Pathologists Board of Registry–certified medical technologists performed better than laboratories that did not employ individuals with this certification. The Board of Registry established stringent education and training standards for laboratory personnel seeking certification. The findings, although they received little attention as CLA was being drafted, suggested that properly educated and trained personnel may be important for high-quality laboratory results.

Several years later, Menmemeyer et al compared patient outcomes based on testing volume and type of laboratory, specifically for prothrombin times. They found that the risk of stroke and myocardial infarction was significantly increased in patients whose laboratory results were determined in low-volume laboratories. These disease outcomes were used because they represent adverse outcomes of inappropriate anticoagulation therapy in specific patient populations. Winkelmann et al then reported additional findings for digoxin-related death and hospitalization. There

Table 3.—Overall Rates of Unsatisfactory and Unsuccessful Performance

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>Total Analyte Challenges</th>
<th>Unsatisfactory, % (n)</th>
<th>Unsuccessful, % (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physicians’ office laboratories</td>
<td>656</td>
<td>21.5 (141)</td>
<td>4.4 (29)</td>
</tr>
<tr>
<td>Physicians’ office laboratories using clinical laboratory scientists (medical technologists)</td>
<td>662</td>
<td>14.0 (93)</td>
<td>1.8 (12)</td>
</tr>
<tr>
<td>Non–physicians’ office laboratories</td>
<td>2991</td>
<td>8.1 (243)</td>
<td>0.9 (26)</td>
</tr>
</tbody>
</table>

*P < .001. †P < .02.

Table 4.—Unsatisfactory Scores for Chemistry* and Hematology† by Testing Challenge

<table>
<thead>
<tr>
<th>Testing‡</th>
<th>Analyte Group</th>
<th>Non–Physicians’ Office Laboratories, % (n)</th>
<th>Physicians’ Office Laboratories Using Clinical Medical Scientists (Medical Technologists), % (n)</th>
<th>Physicians’ Office Laboratories, % (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event 1</td>
<td>Chemistry</td>
<td>2.9 (1628)</td>
<td>6.1 (375)</td>
<td>9.0 (363)</td>
</tr>
<tr>
<td></td>
<td>Hematology</td>
<td>1.9 (883)</td>
<td>4.3 (188)</td>
<td>8.5 (213)</td>
</tr>
<tr>
<td>Event 2</td>
<td>Chemistry</td>
<td>4.2 (1624)</td>
<td>4.2 (381)</td>
<td>12.6 (366)</td>
</tr>
<tr>
<td></td>
<td>Hematology</td>
<td>2.3 (1002)</td>
<td>3.4 (207)</td>
<td>4.8 (228)</td>
</tr>
<tr>
<td>Event 3</td>
<td>Chemistry</td>
<td>4.7 (1652)</td>
<td>7.9 (378)</td>
<td>10.5 (373)</td>
</tr>
<tr>
<td></td>
<td>Hematology</td>
<td>2.4 (953)</td>
<td>6.6 (211)</td>
<td>11.9 (227)</td>
</tr>
</tbody>
</table>

*Chemistry is defined as cholesterol, triglycerides, glucose, potassium, thyroid-stimulating hormone, and digoxin. †Hematology is defined as erythrocyte counts, leucocyte counts, and prothrombin times. ‡A total number of records for each testing event varies depending on enrollment as shown by the number of events.
were increased numbers (14% vs 12%) of patients with adverse events when their digoxin levels were determined in POLs. In 1996, the CDC published initial findings that compared PT outcomes of CLIA-regulated laboratories. They presented data for 17,058 laboratories enrolled in the 7 largest federally approved PT programs. The CDC study evaluated PT performance for 10 analytes commonly used in PTs and found that PT failure rates ranged from 1.2% to 5.3% for hospital and independent laboratories, 4.1% to 15.9% for POLs, and 2.1% to 11.6% for other testing sites. An expansion of these findings is presented as a companion paper to this article.

We used PT data to assess both “accuracy” and “reliability” of testing because PT is required of all laboratories performing moderate- or high-complexity tests. PT performance is not a perfect surrogate for actual laboratory quality; it is useful to identify analytical performance concerns and has been shown to reflect the quality of actual patient specimen testing. Proficiency testing providers evaluate the test results, calculate the scores, and notify the participating laboratory and appropriate regulatory agencies. Therefore, accuracy of testing is reflected in the scores on each testing challenge, and the reliability of testing is reflected in the scores from 1 testing event to the next throughout the year.

We evaluated only 1996 data because, although CLIA standards began in 1992, California’s Senate Bill 113 was not implemented until January 1996. Data from earlier years might have put previously unregulated POLs at a disadvantage in that they may not have had an adequate opportunity to improve their PT performance following the CLIA implementation. However, by 1996, most California laboratories had been subjected to CLIA compliance, including PT, for 4 years and had experienced 1 or 2 on-site inspections. The 11 analytes selected for this study were chosen because they are commonly used laboratory tests representing multiple specialties and are routinely performed in many POLs and non-POLs. Limiting this study to a single PT provider should not compromise the study’s validity because all PT programs are uniformly administered and must comply with the same federal and state requirements. A laboratory that performed poorly in one program would be expected to perform similarly in others.

The grouping of POLs, non-POLs, and POLs using CLS/MTs was completed without prior knowledge of PT scores. Although the Senate Bill 113 mandate was to conduct a study to compare POLs that might use “any other person” for laboratory testing with non-POLs, the third group was identified separately to determine if the presence of licensed laboratory professionals (ie, CLS/MTs) in the POL affected PT performance. California-licensed technologists must possess a baccalaureate degree, have 1 year of approved laboratory training in all test areas, and pass a state examination. This study shows that the unsatisfactory and unsuccessful failure rates were significantly lower in POLs that included licensed laboratory professionals as part of their team.

Although instrumentation in POLs often differs from that in larger hospital and reference laboratories, all PT services, including AAB, establish performance expectations and score laboratories based on similar instrumentation. Therefore, laboratories are scored only with other facilities using similar equipment and reagents.

It might be postulated that some PT samples require reconstitution, thereby introducing an additional factor into the testing process unique to these samples as compared to patient specimens. However, we found similar differences in failure rates for nearly all analytes, including some, such as hemoglobin and hematocrit, that do not require any unique sample preparation in advance of the testing process.

The CDC study examined a larger number of PT challenges but were able to evaluate only 2 comparison groups, the larger laboratories and all other laboratories. Our California study has gone an additional step by subdividing laboratories into 3 discrete groups, comparing POLs that involve laboratory professionals in the testing process with those that do not. Statistically significant differences exist among all 3 comparison groups. Despite study design differences, the California and CDC results are strikingly consistent.

These initial findings suggest that important differences in the quality of laboratory testing exist that may be dependent on the type of testing personnel. Does this mean that every clinical laboratory, irrespective of its size, must involve laboratory professionals in the testing process? Probably not; however, the results suggest that testing personnel in many POLs might lack the necessary education, training, and oversight to perform the tests properly. Many physicians who operate office laboratories may not fully understand the importance of ensuring the integrity of the total testing process (ie, preanalytical, analytical, and postanalytical) and may be unaware of the inaccurate test values that result from improper test performance by their staff.

These data suggest the need to better understand the contributing factors that result in the poorer results of POLs relative to non-POLs. At present, however, patients should be aware that these preliminary findings suggest that there may be a difference in quality of laboratory tests based on where those tests are performed. As we celebrate the 10th anniversary of the Wall Street Journal that highlighted specific problems with certain laboratory tests, we may wish to reconsider the wisdom, particularly in this volatile health care environment, of further easing restrictions on those to whom we entrust our laboratory specimens. At the very least, laboratory directors at all testing sites must ensure that they command a sufficient understanding of laboratory practice to minimize errors and maximize accuracy and reliability.

This study was supported by the California Department of Health Services.

The authors express their sincere thanks to Emily Lee, MS, Ricky Chang, MS, Judy Schlosser, Alice Brydon, MA, and Paul Kimsey, PhD, for their work on the study. The authors and the Department of Health Services are indebted to the Clinical Laboratory Technology Advisory Committee, under the guidance of Fred Struve, Jr, for their dedication to this project and to the assurance of quality laboratory care for all Californians.

References

4. Calif Bus and Prof Code §1224.5.

©1998 American Medical Association. All rights reserved.