Governmental Perspectives on Evaluating Laboratory Performance

D. Joe Boone

The quality of laboratory analytical performance required to support medical decision-making has been defined in four major ways: (a) by the analytical variance of the state of the practice; (b) by the total variance, including analytical and biological variability; (c) by the loss of diagnostic efficiency attributable to analytical error; and (d) by medical-usefulness criteria. From the federal government's perspective, the answer to the question "How good must a laboratory test result be to be medically relevant?" must take into account the clinical context of the test, with accompanying concerns about access, timeliness, and cost, as well as limits for precision and accuracy in the analytical process and the frequency and potential patient-care impact of error in the pre- and postanalytical steps of the total testing process. Therefore, medically relevant goals should encompass not only analytical precision and accuracy but also goals to provide access to clinically effective tests and to reduce errors in the total testing process that can lead to medically misleading information. Development of more appropriate regulatory requirements for laboratories, as well as any needed improvements in instrumentation and methodology, should focus on ensuring that goals for medically relevant results are met by appropriate design and management of the entire process of laboratory testing.

Indexing Terms: standards • proficiency testing • total testing process • patient care

With the passage of the Clinical Laboratory Improvement Amendments of 1988 (CLIA '88) (1), Congress has profoundly influenced laboratory testing practices in the USA. Media stories (2) leading to the passage of CLIA '88 raised the public's concern about inaccurate and imprecise results for certain laboratory tests. In contrast, many government officials, laboratory professionals, physicians, manufacturers, and others involved in health-care delivery consider most laboratory testing conducted today to be highly accurate and reliable. As a consequence, the new regulatory standards to implement CLIA '88 (3) have come under intense scrutiny. The goal is to create a regulatory framework that is not burdensome, incorporating equitable requirements and maintaining timely access to testing at a reasonable cost while ensuring that all laboratories have personnel with the technical ability and skills to provide quality results.

Despite the diligent efforts of professional organizations, industry, academia, and governmental agencies to provide the information needed to reassure the public that laboratory testing results are in general highly accurate and reliable, the perception of problems with laboratory testing remain. Thus far, one of the handicaps to changing this public perception has been our inability to demonstrate a clear linkage between traditional measures to assure laboratory test quality, such as proficiency testing (PT), quality control (QC), and quality assurance (QA), and patient outcome. Establishing medically relevant goals for the analytical accuracy and precision of test results as well as processes to eliminate or reduce the number of mistakes in the pre- and postanalytical phases of testing should enhance our ability to link PT, QC, and QA activities with patient outcomes and justify our reliance on these measures to ensure the accuracy and reliability of the complex process of laboratory testing.

Much of the emphasis in CLIA '88 rests on the recognition that the laboratory world now includes not only hospitals and referral laboratories, but many other testing environments, each of which requires assurance of the quality of patient testing. The successful application of QC to ensure that a laboratory's goals for test precision are met and of external QA and PT to ensure test accuracy goals are met forms the cornerstone of CLIA. In addition, QA, which is addressed by one of the shortest sections of the CLIA regulations, recognizes PT, QC, and other factors as important features in maintaining reliable patient testing.

The federal government, taking into account the constraints imposed by the law and >60 000 comments, has attempted to apply logic and science to the greatest extent possible in developing regulatory requirements. For example, in the area of PT, the Centers for Disease Control and Prevention (CDC) has for at least the last 15 years sought the advice of the medical and laboratory community on what should constitute the minimal content and grading criteria for a regulatory PT program. Features on which little disagreement has occurred include: (a) the need for medical-usefulness criteria; (b) the need for statistical power; (c) minimal program cost; (d) PT providers having the skill to offer an operationally simple program and having the ability to assist laboratories, when necessary; (e) providing PT programs that are easy for the participants and the public to understand; (f) providing programs that are applicable to all laboratories; and (g) minimizing pro-

US Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, Division of Laboratory Systems, Public Health Practice Program Office, Atlanta, GA (address for correspondence: Division of Laboratory Systems, Mailstop G-35, Centers for Disease Control and Prevention, Atlanta, GA 30333).

1 Nonstandard abbreviations: CLIA '88, Clinical Laboratory Improvement Amendments of 1988; PT, proficiency testing; QC, quality control; QA, quality assurance; CDC, Centers for Disease Control and Prevention; and CAP, College of American Pathologists.

Received November 25, 1992; accepted February 25, 1993.
gram-to-program variations to assure consistency in assessments of laboratory performance. One of the most controversial areas has been the mode of PT delivery: whether on-site or blind testing assessment was necessary to circumvent the ability of a laboratory to cheat on its PT performance, and whether in some cases individual workers rather than laboratories should be tested.

Estimating the Reliability of Laboratory Performance

Both Jenny and Jackson (4) and Rej and Norton (5) have demonstrated that PT performance can predict the accuracy of patients' test results. In an evaluation of 200 laboratories enrolled in the New York State PT program for theophylline, Jenny and Jackson (4) found that the imprecision of test results obtained during observed on-site testing and during unobserved mail-distributed testing was equivalent. In addition, they found that whereas some substandard performance escaped detection, every laboratory with substandard PT performance was shown to provide substandard results for patients.

A laboratory's PT performance has been used as a proxy for assessing the total analytical error within the laboratory. Several different approaches have been used to establish goals for total analytical error. They can be defined in four major ways: (a) by the analytical variance of the state of the practice; (b) by the total variance, including analytical and biological variability; (c) by the loss of diagnostic efficiency; and (d) by medical-usefulness criteria (6, 7).

The state of the practice is usually defined in terms of arbitrary statistical assessments of laboratory PT performance, which assumes that the results of 10 or more laboratories using identical or similar methods will obey the principles of gaussian statistics. For practical purposes this assumption seems valid, if outliers are excluded and decimal points are appropriately taken into account (8). Under CLIA '88 PT rules (3) the evaluation of quantitative results is based on the acceptance of results within either fixed-limit criteria or the central 99% of all results, defined as within ±3 SD limits of the target value.

Defining acceptable performance in statistical or fixed-limit terms has several advantages, including: (a) allowing for greater variability of results at extremes of sample concentration; (b) permitting comparisons between precision estimates for different methods (so-called peer groups); and (c) allowing the application of Westgard-type QC rules (9) to determine the probability of failing PT with various amounts of inherent laboratory imprecision and inaccuracy (10). A disadvantage of this approach is that it defines acceptable performance in analytical rather than medical-usefulness terms. Therefore, results could be considered acceptable according to statistical criteria and yet fail to meet medical needs. The use of statistical or fixed-limit methods to evaluate laboratory performance establishes laboratory practice measures that can stimulate laboratories to produce consistent results and manufacturers to improve the precision and accuracy of their methods.

These techniques, however, fail to address the key question of the medical utility of the results regardless of whether a laboratory's results are inside or outside arbitrary limits, and thus they define patient risk in terms of analytical performance. On the other hand, improvements in analytical performance can lead to extremely narrow distributions of results, with limits that provide unacceptable performance ratings for medically useful data.

A second approach to evaluating analytical quality requirements has been to examine the relationship between analytical and biological variation. Tonks (11) proposed that analytical variation not exceed one-fourth of the normal range, whereas Cotlove et al. (12) indicated that analytical variation should not exceed one-half of the relevant biological variation. During the 1976 College of American Pathologists (CAP) Conference on Analytical Goals in Clinical Chemistry (13), participants developed a consensus that related analytical goals for performance to analytical and biological variance for both group screening and individual single-point testing. Although Gilbert (8) addressed accuracy of laboratory results during the conference, the primary focus was on precision of test results.

A third approach to evaluating laboratory performance was pioneered by Barnett (14), who attempted to take into account the medical use of the test results. Skendzel et al. (15, 16) and Elion-Gerritzen (17) extended this work to include queries of physicians to determine how much analytical variance can be tolerated in typical patient-care settings. Differences in opinion among physicians of medically tolerable variance in results were apparent. Given that the diagnostic truth usually is not known, the lack of agreement among physicians about the medical decision value of test results is not surprising. The essential findings of these studies, however, were that the analytical variation in laboratory test results is often better than the empirical estimates of physicians' needs. Assuming that allowable analytical variation of laboratory test results need not exceed that required for current medical decision-making, this approach to establishing performance criteria for laboratories has merit. A major difficulty with this approach, however, is that the criteria developed would depend on the clinical context of the test, and agreement might not be reached among clinicians about allowable variability for any specific patient's condition, let alone the vast number of potential patients' conditions for each test.

Fraser (18) has advocated defining acceptable PT performance by assuming that the goal for inaccuracy is zero bias and that the goal for imprecision should be equal to or less than one-half of the average intraindividual (within-subject) biological variation. Thus, individual biological variability defines the goal of analytical test variability. This term is irreducible, so it defines the ultimate goal for allowable variability of each test. If the current state of practice indicates that this level of analytical performance cannot be achieved, state-of-practice goals should be used.
How close do laboratories come to achieving Fraser's goals of zero bias and imprecision equal to or less than one-half of the average within-subject biological variation? Using data presented by Fraser (18) for the within-subject biological variability of chemistry tests as the analytical goal and the allowable error for PT as stated in the February 28, 1992, regulations (3), we can compare the analytical goal with PT performance criteria (Table 1). Because allowable limits for PT include among-laboratory variability, limits are reduced by 60% to determine the expected within-laboratory PT allowable error. This factor corresponds to the usual difference between within-laboratory and among-laboratory imprecision noted in CAP Quality Assurance Survey data (19). Because the allowable PT performance is based on the state of practice, the ratio of within-laboratory allowable error to the analytical goal indicates how much improvement must occur before laboratory testing can reach the ultimate analytical goal. Ratios of 1.5 or less probably indicate that we are close to the goal. Similar assessments are made for therapeutic drug monitoring data (20) in Table 2 and for hematology data in Table 3, with use of biological variability data for hematology presented by Statland and Winkel (21). These findings indicate that the allowed variation in PT, which includes analytical bias and imprecision, is not usually grossly out of line with this estimation of the ultimate analytical goals.

Lott et al. (22) have described three sets of limits to define goals for medically acceptable error based on the clinical context of the test. For screening populations, fixed criteria were recommended, with these limits reduced by one-half for outpatients (long-term assessments) and by one-fourth for inpatients (short-term assessments). Hytlof Petersen and Hørder (23) have investigated specific diagnostic situations involving decisions when two well-defined groups exist, one with the disease and the other without disease (bimodal data), and the other where the probability of disease increases or decreases with concentration (unimodal data).

One way to encourage continued improvement of laboratory technology while preserving medical-usefulness criteria would be to assess not only the analytical performance of a laboratory but also the implied clinical utility of the test results. The CDC has used this approach to evaluate laboratory performance for neonatal screening tests for hypothyroidism and phenylketonuria in blood spots (24). Not only can inaccuracy and imprecision of quantitative data be detected, but also errors or differences in interpretations being provided to physicians for patient care. If reference intervals are inappropriate, clinically misleading information can be transmitted and possibly acted upon, even if the quantitative result is accurate and precise.

When is the magnitude of error in a test result sufficient to lead to a potentially inappropriate clinical decision? If a laboratory's glucose test results indicate hypoglycemia, while the actual glucose concentration is within the reference interval, the laboratory has provided misleading information that could affect patient care. The CDC has used the medical-decision points defined by Statland (25) to evaluate PT results for glucose (26) and phenytoin (27). Even if typical laboratory imprecision is taken into account, more laboratories than predicted (based on typical among-laboratory variability) reported results outside of medical-decision points for three hypoglycemic samples, for five of seven samples within the reference interval, and for seven of nine samples with abnormally high glucose concentrations (26). Although this approach requires refinement, medical-decision points could be developed to assist

### Table 1. Comparison of Allowable Error in PT with Analytical Goals Defined by Within-Subject Biological Variability

<table>
<thead>
<tr>
<th>Analyte/test</th>
<th>Analytical goal* (AG) CV, %</th>
<th>Allowable error* (AE), %</th>
<th>(0.6 × AE/AG)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine aminotransferase</td>
<td>13.6</td>
<td>20.0</td>
<td>0.9</td>
</tr>
<tr>
<td>Bilirubin, total</td>
<td>11.3</td>
<td>20.0</td>
<td>1.1</td>
</tr>
<tr>
<td>Calcium, total</td>
<td>0.9</td>
<td>10.5</td>
<td>7.0</td>
</tr>
<tr>
<td>Cholesterol, total</td>
<td>2.7</td>
<td>10.0</td>
<td>2.2</td>
</tr>
<tr>
<td>Glucose</td>
<td>2.2</td>
<td>10.0</td>
<td>2.7</td>
</tr>
<tr>
<td>Potassium</td>
<td>2.4</td>
<td>10.0</td>
<td>2.5</td>
</tr>
<tr>
<td>Sodium</td>
<td>0.3</td>
<td>2.8</td>
<td>5.6</td>
</tr>
<tr>
<td>Thyroxine</td>
<td>3.4</td>
<td>20.0</td>
<td>3.5</td>
</tr>
</tbody>
</table>

* Goals from Fraser (18).

### Table 2. Comparison of Allowable Error in PT with Analytical Goals Defined by Within-Subject Biological Variability

<table>
<thead>
<tr>
<th>Analyte/test</th>
<th>Analytical goal* (AG) CV, %</th>
<th>Allowable error* (AE), %</th>
<th>(0.6 × AE/AG)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbamazepine</td>
<td>3.2</td>
<td>25.0</td>
<td>4.7</td>
</tr>
<tr>
<td>Digoxin</td>
<td>2.7</td>
<td>20.0</td>
<td>4.4</td>
</tr>
<tr>
<td>Lithium</td>
<td>1.8</td>
<td>20.0</td>
<td>6.7</td>
</tr>
<tr>
<td>Phenobarbital</td>
<td>1.1</td>
<td>20.0</td>
<td>11.0</td>
</tr>
<tr>
<td>Theophylline</td>
<td>7.0</td>
<td>25.0</td>
<td>2.1</td>
</tr>
<tr>
<td>Valproic Acid</td>
<td>8.3</td>
<td>25.0</td>
<td>1.8</td>
</tr>
</tbody>
</table>

* Goals from Fraser (20).

### Table 3. Comparison of Allowable Error in PT with Analytical Goals Defined by Within-Subject Biological Variability

<table>
<thead>
<tr>
<th>Hematology analyte/test</th>
<th>Analytical goal* (AG) CV, %</th>
<th>Allowable error* (AE), %</th>
<th>(0.6 x AE/AG)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hematocrit</td>
<td>2.2</td>
<td>6.0</td>
<td>1.6</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>2.2</td>
<td>7.0</td>
<td>1.9</td>
</tr>
<tr>
<td>Leukocyte count</td>
<td>7.8</td>
<td>15.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Platelet count</td>
<td>3.3</td>
<td>25.0</td>
<td>4.5</td>
</tr>
</tbody>
</table>

* Goals from Statland and Winkel (21).

* Allowable error reduced by 60% to obtain estimate of within-laboratory allowable error.
regulatory agencies in identifying those laboratories whose results are most likely to be clinically misleading. A possible enhancement of the evaluation process would be to incorporate the suggestions of Ehrmeyer and Laessig (28) to provide a cumulative percentile ranking of laboratory performance on the basis of algebraic and absolute mean error, which offers the advantage of separating random and systematic sources of error in the analytical portion of testing.

Lawson et al. (29) describe another way to enhance the value of information being collected by PT and QC programs. They recommend that PT samples be used as samples in shared pools for laboratory QC. This linkage of activities would provide laboratories with a better accuracy base for QC and improved assessments of how their internal imprecision compares with that of other users of the same methods.

Evaluating the Total Testing Process

Use of voluntary and regulatory PT programs as well as laboratory internal QC activities has permitted us to track significant improvements in the analytical performance capability of laboratories in terms of reduced inaccuracy and imprecision over the last 40 years. While it would be useful to define the amount of inaccuracy and imprecision that would limit the medical relevance of test results, the error inherent in the analytical phase of the total testing process will probably be considered acceptable for most clinical purposes. Such assessments will require periodic revisits to confirm continued appropriateness of the clinical criteria.

Of greater concern is the location of errors within the total testing process, the frequency of their occurrence, and the impact of these errors on patient care. Ross and Boone (30), investigating the mistakes occurring in laboratory testing in a 570-bed general and acute care hospital, found the following distribution of laboratory testing mistakes: 46% preanalytical, 7% analytical, and 47% postanalytical. Nonlaboratory personnel accounted for 28.6% of these mistakes. Of 336 patients involved in 304 incidents in which mistakes were made, no effect on patient care was found for 233 patients; 78 patients were not harmed but were subjected to an unnecessary procedure not associated with increased patient risk, and 25 patients were not harmed but were subjected to an additional risk of inappropriate care. Today, mistakes in the preanalytical and postanalytical phases of testing seem more likely to affect the usefulness of laboratory results in patient care than does variability in the analytical phase of testing. Therefore, consideration of sources of test inaccuracy must be expanded to include detecting and correcting mistakes that could affect patient care in the pre- and postanalytical phases of testing. Imprecision should be redefined in terms of the ability of a testing facility to reliably conduct test ordering, specimen collection, transport, storage, analysis, and result reporting in an accurate, timely, and cost-effective manner. Developing a facility-wide QA process to detect and correct mistakes in the entire process of laboratory testing is a fundamental require-

ment under the CLIA regulations (3). Participation in the CAP Q-Probe or similar program (31), which allows laboratories to determine how their internal operations compare with those of similar facilities, seems essential to ensure that standards of practice in nonanalytical phases of the testing process are being appropriately met. In a recent CAP Q-Probe study (32) of blood-bank QA practices, in which the testing process was divided into 24 discrete steps, with each step representing an area for which a testing mistake could occur, 96 of 1328 facilities had an active system to monitor all 24 steps. Of ~64 000 testing process mistakes that were detected and corrected by these facilities, 52% occurred in preanalytical steps, 43% in postanalytical steps, and 5% during the analytical steps. The consequences of these mistakes on patient care were not determined.

Conclusions

Are we ready to set medically relevant goals for the precision and accuracy of laboratory tests in the US? Are such goals necessary? Can they be kept consistent with the current standards of medical practice? Most importantly, will the availability of agreed-upon goals improve patient care?

On the basis of the nation’s experience with the National Cholesterol Education Program, we can answer yes to all of these questions for cholesterol testing. Perhaps our accomplishments with cholesterol can be extended to include other laboratory tests.

Unfortunately, successful definition of medically relevant goals for a laboratory test is not sufficient to ensure its effective use for patient care. Reducing the risks to patients of inaccurate or imprecise results will require emphasizing improvements in the entire process of laboratory testing, not just the analytical phase of the process. Therefore, although medically relevant goals for analytical accuracy and precision would be helpful in setting more-realistic requirements for PT and QC activities, more than that is needed to capture and correct the majority of mistakes in the laboratory testing process. To capture and correct these mistakes in a timely and cost-effective manner before they affect patient care requires a focus on obtaining and maintaining accuracy and reliability throughout the testing process.

How can we achieve the broader goal of ensuring that laboratory data are medically relevant throughout the testing process? Certainly, medically relevant goals can and should be defined for the analytical phase of testing to permit realistic assessments of PT and QC data. Evaluation of both the quantitative result and its accompanying clinical interpretation could help identify laboratories with inaccurate and imprecise results and reference intervals. In addition, norms should be established within each institution based on the frequency of nonanalytical mistakes in the total testing process over a specific time interval. Such internal norms might be compared with those of other institutions with similar complexity of operations and clinical contexts of testing. These norms would permit a facility to identify areas most in need of improvement and allow the steps taken
to improve testing performance to be measured. They would also foster the integration of the concept of total quality management by all personnel involved in the testing process, including physicians, nurses, physicians' assistants, phlebotomists, medical technologists, pathologists, clinical scientists, and others.

Accomplishing accurate and precise laboratory testing will require the cooperation of laboratories, manufacturers, professional societies, health-care providers, and governmental agencies. For institutions that I have studied, the incidence of mistakes in the laboratory testing process is very low, with most occurring in the nonanalytical portions of the testing process. Even when mistakes occurred, the likelihood of adverse patient consequences was lessened by the QA practices of the laboratory and clinical staff. Health-care providers, however, need not wait for another CLIA '88 for motivation to further reduce the chance that these rare events will harm patients. The CAP and the Joint Commission on Accreditation of Hospital Organizations have developed tools to assist laboratories and health-care providers in reducing mistakes and documenting continuing improvement in the testing process. Many laboratories and institutions already have processes in place to detect and correct mistakes in the total testing process before these mistakes can adversely affect patient care. With the cooperation and leadership of everyone involved in the laboratory testing process, beginning with test ordering and ending with the use of test results for patient care, we can assure that patients have access to accurate, timely, cost-effective, and reliable laboratory testing.

References

Discussion

Larry Miller: Mistakes are made, and our patients can reasonably demand some limits on accuracy and precision. Certainly, our regulators are demanding this, as is evident in a thorough reading of the responsibilities of the director in the CLIA '88 rules. Accuracy and precision limits need to be set at realistic values now, so that they can be refined over time. Limits also need to be established in other disciplines, such as coagulation, hematology, and immunology. In my opinion, the limit of error needs to be defined at the decision limit of the analyte. One of our major challenges is people doing laboratory testing who don't recognize when mistakes are made. We need to provide education to correct this and help to define limits of error, and have mechanisms of follow-up.

Bradley Copeland: Dr. Boone, you mentioned that CDC had detected progress in the number of errors over the last 4 years. Could you give us the latest figure on that? Second, is my calculation correct that 99.97% of errors were corrected before medical decisions were made in the study you were reporting? I would like to know how many mistakes were made per hospital in this study?

Joe Boone: In terms of the data over the 4 years, I was just talking about trends. I don't have any particular information about the actual prevalence of errors. My general assumption all along has been that we are talking about rare events. In fact, when trying to put together our CLIA studies, we recognized that we think we will be looking for very rare occurrences, and it is going to be very difficult for us to find anything.

In terms of your second point, it depends on whether in the testing process that the mistakes occurred and whether they were actually corrected before medical management. In some cases, it was difficult to discern whether they had been corrected first. I'm not sure it is as high as 99.97%; it is probably 90%. However, I don't know whether these results are representative. Of course, one problem with an error detection system is that you don't know how many errors you missed; some you simply will never find. In the Q-Probe study, all mistakes were detected and corrected. There is a big difference between detecting and correcting mistakes and detecting but not correcting mistakes.

Herbert Fritsche: We all spend a great deal of time with quality assurance committees and protocols within our hospitals as required by the JCAHO. Do you have any indication that any additional quality-assurance measures are necessary?

Joe Boone: Beyond those that are already in place? No, I don't.

Diana Trundle: One item not included by Dr. Ross is another consequence of an error: a request for consultation by a specialist. This becomes costly and occurs when the error is subtle, i.e., when the difference from the "true" value is small and not really inconsistent with the clinical presentation. A large difference is instantly recognized. If the consultant evaluates the patient, and orders more tests for the next day, the consequences are compounded and expensive.

Eileen Gorman: We've been spending a lot of energy on quality management systems. To quote a prominent saying: "If you do what you have always done, you are going to get what you have always gotten." If the majority of laboratory mistakes are not caused by analytical error but are caused by the rest of the system, then trying to correct those mistakes by focusing on proficiency testing will not get us the result we really want: an improvement in the quality of the patient's result and patient testing.

Wiveka Elion-Gerritzen: I have heard many times here that physicians are unaware of analytical, preanalytical, and biological variability. People have been saying and writing this for many years, and yet I know of no proof for these statements. In my study, I interviewed clinicians and, for example, for the lower limit of normal for potassium and the upper limit of normal for calcium I registered zero differences between the limit of normal and the action value. In these cases, physicians took action exactly at the limit of normal or even before the limit of normal was reached—which did not mean they were unaware of variability of laboratory results. A physician when making a decision will consider the seriousness of the illness, being, for example, much more keen on liver function tests when liver metastases are suspected than when alcohol abuse is at stake. Medical action will be indicated by the availability of therapy, the chance of success, and the risk taken when the physician does not do anything. The first action is very often to repeat the test. I think physicians are very well aware of analytical, preanalytical, and biological variability—and also of our blunder rate!

Mario Werner: I think the issue here is not what physicians do but what physicians should do. The difference between the two is well documented. For instance, consider the variability among physicians in assessing prognosis in a given situation. In one study asking physicians about 10-year survival after heart valve implantation, prognostic estimates ranged from 5% to 95%. Clearly the real probability of survival is a single finite number, even if we allow for sampling variability about that number. Thus, a realistic prognostic estimate can be made, even though many physicians would not make it. The same considerations apply to therapeutic decisions. For instance, the frequency of hysterectomy varies threefold between cities, but no explanation for these differences has been found. Therefore, optimal medical strategies cannot be derived from a general analysis of physician's behavior. Rather, the scientific basis that should underlie medical decisions must be analyzed. This is indeed the concept that underlies recent federal legislation for a centralized effort to develop practice parameters for outcome assess-
ment and their implementation through guidelines for medical actions.

Joe Boone: There are also specific areas that need to be focused on within institutions. For example, Dr. Ross’s hospital had just incorporated a new trauma center and that is where a majority of the arterial blood gas mistakes occurred. Mistakes in this area overshadowed practically everything else that was going on. So, you do have areas of focus within each institution, for which improvements really would pay off.

Katherine Erickson: Dr. Boone, your talk was titled “Government Perspectives on Evaluating Laboratory Performance,” and a good portion of your talk dealt with the total system, rather than just the analytical aspect. Should we assume that this is the direction the government is going to be taking: a view of the evaluation of our total systems rather than just the analytical component?

Joe Boone: Well, believe it or not, that is what I thought CLIA did. We tried to build flexibility and responsibility into CLIA, making it clearly the responsibility of the director of the laboratory to assure that certain practices are being performed in that facility. The quality assurance part of the regulation is just one page long but, in my opinion, is the most important aspect of the whole regulation. I don’t know where that philosophy extends beyond me; the government is a big operation, and we have a lot of different people with a lot of different views.

John Batsakis: We had a Medicare inspection several months ago on the heels of a JCAHO inspection. The Joint Commission looked at the total picture, but the Medicare inspection was a throwback to the 1960s: They asked to see signatures on the manuals and things like that; they did just the opposite of what you propose.

My second point is that, if you say that the CDC error detection is almost by chance, then how do you define the error?

Joe Boone: When I was speaking of detecting errors by chance, I think it was in the context of the blood bank data. We tried to separate passive error detection systems from active systems for detection of errors. In the passive system, you are not looking for errors, they are detected by chance. In an active system, you are really looking to see if mistakes are occurring.

William Hamlin: One of the largest pitfalls we can get into is to talk about error rate without specifically and precisely defining error. At least for some people, labeling failures are not a laboratory error. For other people, in the context of the laboratory being responsible for controlling all the activity surrounding its analytical responsibilities, including pre- and postanalytical activities, incorrect labels might be defined as an error. And some people would define error only in the context of an adverse clinical outcome: If there is no adverse clinical outcome, there is no error. Although many people might argue with that definition, I think there is some justification for it.

I have grave reservations about where the responsibility for so-called total quality management should be assigned in the clinical laboratory context. The average clinical laboratory is subject to a variety of vagaries. One of those is the nature of the people who request the work. They are uncontrollable to a large degree, and they receive no stimulus toward improvement of any kind. The classic example is the requirement that certain kinds of information be submitted on request slips for cytology examinations in this country. The average laboratory that does cytology screening simply cannot control that activity. Even telephone requests, which take up enormous amounts of time and energy, are met with either no response or an unsatisfactory response.

If, in fact, we are going to set analytical goals in the broadest sense, that is, including the variables within the spectrum surrounding the analytical process, then somebody other than those of us in this room (laboratorians) is going to have to be involved, because we can’t control most of the pre- and postanalytical problems.

Patricia Garrett: Dr. Boone, one of your slides showed error rates for various disciplines within the laboratory. The error rates were all quite low—98.8 per 100 000 was the highest and 37.1 per 100 000 was the lowest. Furthermore, the rates seemed to drop quite consistently with the volume of tests in the various laboratory sections. Would you please comment on this?

Joe Boone: With respect to the influence of test volume in proficiency testing, particularly in the microbiology area, we have observed that lower-volume facilities perform worse than laboratories that have higher volumes. There certainly seems to be volume correlation with certain PT performance.