Sharon S. Ehmeyer,1 Ronald H. Laessig,1,2 James E. Leinweber,2 and Jeffrey J. Oryall1

On March 14, 1990, the Centers for Disease Control and the Health Care Financing Administration published criteria for defining minimum performance in proficiency testing (PT). Using our previously described computer modeling technique, we determined the likelihood of passing PT under the new rules. The model relates combinations of intralaboratory CV and bias to PT performance criteria. For example, a laboratory with a bias of zero and an internal CV of 5% will pass a 10% fixed-limit PT criterion (i.e., the criterion for glucose analyses) 98% of the time when five samples are used. The model provides similar analyses for all PT criteria and all relevant combinations of CV and bias. The probability of passing PT decreases as the number of analytes tested increases, i.e., from 98% to 37% as the number of analytes increases from 1 to 20. A laboratory's internal CV has a greater effect on the outcome of PT than do the corresponding bias values. We conclude that a laboratory that operates with methods that have internal CVs ≤33% and biases ≤20% of the PT criteria will have a >99% chance of passing PT.

On August 5, 1988, the Centers for Disease Control (CDC) and the Health Care Financing Administration (HCFA) published proposed revised minimum standards for all U.S. proficiency testing (PT) programs that certify laboratory performance under Medicare and the Clinical Laboratory Improvement Act (CLIA) 1967 (1). Final rules for PT, published on March 14, 1990, will be effective January 1, 1991, for current Medicare-certified and CLIA-licensed clinical laboratories (2). Additional regulations required to implement CLIA-1988 amendments, mandating HCFA to promulgate regulations that contain PT for all other laboratories including physician's office laboratories, will follow (3).

The concept of using interlaboratory PT to assess the intralaboratory quality of clinical chemistry tests began with Belk and Sunderman in 1947 (4). Soon thereafter, the College of American Pathologists (CAP) adopted PT as a means of further improving intralaboratory quality. With the enactment of the CLIA-1967 regulations, successful participation in an approved PT (such as that of CAP) was federally mandated as a condition of licensure (5). In 1968, Medicare adopted a similar requirement for certification (6). Traditionally, PT programs, including those of CAP, the American Association of Bioanalysts, and various states, have evaluated the quality of intralaboratory test performance with four PT shipments per year of two samples, each covering a broad spectrum of different analytes. The newly enacted regulations require five samples per quarterly shipment (PT event) and the "outcome measurements," or PT results, have become the major focus of the regulatory certification/licensure requirements (2, 7). Numerous authors, including ourselves, have commented on the compromise of the originally intended purpose, i.e., education, of interlaboratory comparisons when PT is transformed into a regulatory exercise (8–15).

Materials and Methods
We have demonstrated previously that one can determine the minimum intralaboratory performance levels necessary to meet the requirements specified in the federal rules for PT programs (10, 15–17). Under the plausible and tractable assumption of gaussian imprecision (18), these relationships can be established by computer modeling with a Monte Carlo simulation approach, or by direct calculation based on statistical analysis. In both approaches a laboratory's internal performance characteristics, i.e., its unique imprecision [expressed as the standard deviation (SD) or coefficient of variation (CV)] and its bias, i.e., its offset from the target value, must be taken into account when determining the probability of "passing" one or a series of PT challenges. Other factors such as clerical errors, shipping problems, matrix effects, grading errors, etc., which can amount to 50% of the apparent causes of PT failures, should not be neglected, but a laboratory's analytical prowess is fundamental (19).

We computed the relationship between a laboratory's internal CV and (or) bias for a given analyte and the recently published federal interlaboratory PT criteria. Each analyte has its own PT criterion; for the routine chemistry subspecialty, the criteria are identified in Table 1. By standardizing bias and SD as a percentage of these criteria, in effect obtaining "Z scores," we reduce each analyte to one generic case for analysis. From given (standardized) CV and bias values, we induce a probability of producing a laboratory result outside the (standardized) PT criterion. This allows us to determine the probability of passing one PT event under the new format of five samples per shipment, modeled as a Bernoulli trial.

Required PT Performance

Single analyte. The new regulations specify that "passing performance" in a PT event (one quarterly shipment) requires that four of the five results for each analyte (80% of the results) must fall within the defined acceptable range (target value ± PT limit). The target value can be the mean of the results once outliers have been removed from a group of participants who use the same method or instrument, or can be established by Definitive or Reference Methods. Alternatively, the target value can be the mean of results from 80% of 10 or more referee laboratories (2). From Table 1, when "grading" PT results, the acceptable range for glucose is the target value ± the performance criterion (68 mg/dL or 10%) that yields the greater range. In the case of a PT specimen with a target value of 1000 mg/L (5.58

1 Department of Pathology and Laboratory Medicine, University of Wisconsin, Madison, WI 53706.
2 State Laboratory of Hygiene, Madison, WI.
3 Nonstandard abbreviations: CDC, Centers for Disease Control; HCFA, Health Care Financing Administration; CLIA, Clinical Laboratory Improvement Act; CAP, College of American Pathologists; and PT, proficiency testing.

Received May 7, 1990; accepted July 19, 1990.

1736 CLINICAL CHEMISTRY, Vol. 36, No. 10, 1990
mmol/L), acceptable performance would be between 900 and 1100 mg/L (5.00–6.10 mmol/L). If the target value were 500 mg/L (2.78 mmol/L), the acceptable range would be 440 to 560 mg/L (2.44–3.12 mmol/L).

If a laboratory does not achieve at least 80% acceptable performance on any given analyte (four of five correct results) for a PT shipment, the laboratory is, in effect, put on probation for the entire subspecialty in which that particular analyte is listed. To actually “fail” PT and be subject to “adverse action” (the terms used in the regulations), the laboratory must again fail to achieve acceptable performance for the same analyte on one of the next two PT shipments. Hence, by failing to achieve acceptable performance for the same analyte for any two of three consecutive events, the laboratory may fail the entire subspecialty and thereby suspend testing in the subspecialty. A curious anomaly is that a laboratory could “fail” different analytes, e.g., glucose for one PT shipment, uric acid for the next, and CK for the third, etc., indefinitely without being suspended in the subspecialty of routine chemistry.

Multiple analytes. In addition to passing criteria for the individual analytes, the regulations require a laboratory to achieve an 80% correct response rate over all analytes in a particular subspecialty. To be subject to adverse action, a laboratory must have <80% of all results correct for any two of three consecutive PT shipments. Obviously, to fail this, a laboratory must also fail at least one analyte, i.e., have two or more incorrect out of five results.

Results and Discussion

Required Intralaboratory Performance

One analyte, zero bias. Figure 1 shows the probability of failure, i.e., <80% correct for a laboratory analyzing only one analyte in any one PT shipment. The x-axis is in units of internal CV, or SD as a percent of the PT limit. For example, if the PT limit is 10%, as for glucose (Table 1), 100 on the x-axis denotes a laboratory with an internal CV of 10%. Under these circumstances, this laboratory has a 51% probability (y-axis) of “failing” the analyte glucose in any one PT shipment, for which it analyzes five PT samples. Figure 1 is based on the assumption that the laboratory has zero bias, i.e., any deviation from the target value is caused only by the laboratory’s internal imprecision. Similarly, for analytes whose performance criteria (Table 1) are defined as multiples of the group SD, e.g., 3 group SD for alkaline phosphatase, the 100% point on the graph is equivalent to a laboratory whose internal SD is equal to the entire performance criteria, or 3 group SD.

Further, if the laboratory’s internal CV is 50% of the stated performance limit, the probability of “failing” a single five-sample PT shipment for one analyte drops to just <2%. With an internal CV of 33%, or 1/3 of the PT limit, the laboratory will, in essence, always pass PT. The presence of co-existing bias reduces the “tolerable” CV, as will be shown in subsequent Figures.

Multiple analytes, zero bias. Figure 2 shows the effect of analyzing multiple analytes (glucose, blood urea nitrogen, cholesterol, etc.) on the laboratory’s ability to pass PT. In general terms, for any given internal CV, the more analytes tested, the greater the probability that a laboratory will fail one or more analytes. For example, if a laboratory tests two analytes, with each internal CV equal to 100% of the PT performance criteria, the chance of failing at least one analyte increases from 51% to 76%. A laboratory doing 20 analytes, e.g., operating a large, multichannel instrument (SMAC™, Hitachi™, Hitachi™, etc.) with all the analytes’ CVs equal to 100% of the PT performance criteria, would virtually be assured of failing at least one analyte on every PT event. By reducing the internal CVs to 50% for all analytes, the probability of a failure for the same laboratory doing 20 tests is reduced to 32%. With all CVs below the 33% level, the chances of failure are nearly zero.
shows the likelihood of a laboratory failing a single PT event with co-existing bias (20% and 50% of the PT criterion). As in Figure 2, the family of curves represents, respectively (from the right to left), the probabilities of failure for one, two, five, 10, 20, and 27 analytes. The presence of bias increases the likelihood that a laboratory will fail a PT event. In the case of glucose, where the PT criterion is ±10%, a 20% bias is equivalent to a consistent 20 mg/L (0.11 mmol/L) error. Figures 2 and 3 show the effect of increasing bias on the likelihood of failing a PT event. For one analyte, with biases of 0%, 20%, and 50% of the PT criterion and a consistent co-existing internal CV of 50%, the probability of failure increases from 2% to 4% to 18%, respectively. Further, for 20 analytes, the probability of a failure for biases of 0%, 20%, and 50% increases from 32% to 51% to >98%, respectively. Although a laboratory does not have the same, or for that matter any, bias on every test, any significant bias seriously impairs a laboratory's ability to pass that analyte in a PT shipment. Large bias rather than large imprecision is a common reason for analytical failures (18). Bias from a pipetting error or a reconstitution problem may extend to all analytes and thereby impose a large chance of an 80% failure. Consequently, to pass PT, a laboratory should first minimize the amount of bias for each analyte and then reduce the internal CV, if possible.

Predicted Failure Rates as a Function of CV and Bias Combinations

Figure 4 shows the percent probability of failing a PT event for one analyte as a function of both internal CV and bias. The x- and y-axes depict intralaboratory CV and bias, respectively, as a percentage of the PT limit. The curves have a negative slope, because the presence of bias reduces the "tolerable" internal CV consistent with a given percent probability of a laboratory failing a PT event. The "1" denotes a 1% probability of failing one PT event. As indicated by the continuous line, all combinations of CV and bias falling on the line will yield a 1% chance of failure. Those below (to the left of the line) have less chance of failure. Likewise, curves to the right denote the probabilities of failure of 3%, 5%, etc., for increasing values of CV and bias. Typically, in laboratories in which calibration is performed with reasonable care, biases are small; in laboratories in which biases are >20% and CVs >30% of the performance limit, both need to be reduced.
Minimizing the Contribution of Bias to the Probability of PT Failure

Figure 5 shows the probability of failing a PT event for one analyte in a laboratory with various internal CVs and all possible bias values. Bias does not affect the likelihood of failure to the same extent as do CVs of equivalent size. For example, if a laboratory reduces its CV to 33% of the value in Table 1 for any analyte, a co-existing bias of <40% is tolerable. If the bias can be reduced to <20%, its contribution to the probability of failure to pass PT is almost negligible. Most authors do not deal with the concepts of co-existing bias and imprecision, but rather set bias equal to zero (20). Based on Figure 5, this is a justifiable assumption, at least when dealing with PT data. In fact, the traditional function of PT programs has been to reduce bias because, even with five samples, PT is ineffective in measuring intralaboratory imprecision (13).

A laboratory can readily predict the probability of passing PT based on its internal imprecision (CV) and bias. For the proposed two-of-five (or 80%) PT rules, a laboratory with small (<20%) bias can reasonably ensure its likelihood of passing PT by reducing the imprecision of each analyte to one-third of the federally mandated performance limit. Thus, particularly for multiple analytes, a laboratory can adopt a strategy for passing PT.

A laboratory should reduce bias to a minimum by careful instrument calibration and sample reconstitution and pipetting. Once bias is minimized, the laboratory should then concentrate, in order of priority, on those analytes with the largest CVs by the CDC/HCFA PT criteria. The imprecision of a method is a measurement of the random error. Historically, laboratories have had little control over reducing random error (21). Consequently, the imprecision of a method should be taken into account during method selection and evaluation. Any method that has large random error (internal CV ≥50% of the PT criteria) should not be considered. Shrewd laboratory managers, who must assess PT challenges to stay in the "laboratory business," logically should select new instruments and methods in terms of the PT criteria.

As implied in Figure 2, a laboratory who has successfully applied the rule of "one-third" to 19 of 20 analytes would, for those 19 analytes, be successful in terms of PT; the 20th analyte, as depicted for the right-most curve, would determine the probability of the laboratory failing PT, little influenced by the excellent performance on the other analytes. Obviously, a laboratory in which imprecision is marginal for several analytes compounds its probability of failure. Only two analytes, each with CVs of 60% of the PT criteria, increase the probability of failure from 8% to 14% for zero bias and from 10% to 19% with 20% bias.

The new CDC/HCFA rules also incorporate an 80% across-analyte rule, which means that a laboratory must have at least 80% of all PT results correct in a subspecialty. Obviously, a laboratory cannot fail this 80% rule without first having two or more incorrect results for at least one analyte. This rule will not affect laboratories with relatively small (less than one-third of the PT criterion) internal CVs and biases. Instead, problems with PT will arise from the existence of "statistically" introduced failures in individual analytes. For most laboratories, the probability of an 80% failure (across analytes) occurring on two of three consecutive PT shipments is entirely negligible because of the relatively large number of analytes that would score five of five results correctly. For the marginal laboratory, i.e., one in which several analytes have CVs in the 50-100% category, single-analyte failures will be common and the probability of 80% across-analyte failures will gradually increase.

Ed. note: The choice of molar or other units for criteria was made by the regulators, not us. Our Journal. We have, however, expressed results as concentration per liter instead of per deciliter, for clarity, and report values for SI units. Bilirubin: 1 mg/L = 1.710 μmol/L; calcium: 1 mg/L = 24.96 nmol/L; creatinine: 1 mg/L = 8.840 μmol/L; glucose: 1 mg/L = 5.561 μmol/L; blood urea nitrogen: 1 mg/L = 35.70 μmol/L (urea).

References
13. Ehrmeyer SS, Laessig RH, Schell K. Use of alternative rules
(other than the 1<sub>sd</sub>) for evaluating interlaboratory performance data. Clin Chem 1988;34:250-6.


