

Medical Need for Quality Specifications within Laboratory Medicine

Ronald H. Laessig

*State Laboratory of Hygiene, University of Wisconsin Center for Health Sciences,
Madison, Wisconsin, USA*

MOTTO: MANAGEMENT IS DOING THE THING RIGHT; LEADERSHIP IS DOING
THE RIGHT THING.

INTRODUCTION

The continuous and consistent improvement in the quality of laboratory results is well documented. Within the past decade, the quality of laboratory results, i.e. reduced intralaboratory coefficients of variation (CV) and bias values of essentially zero for many tests have exceeded the requirements of clinicians who utilize the laboratory results to treat patients. Heretofore, laboratorians have concentrated on improving quality of test results; in the future we must concentrate on focusing limited resources where they will be most cost effective.

As numerous studies have documented, clinicians and laboratorians have considerable difficulty when they attempt to translate laboratory attributes of quality, CV, Bias, Total Error, etc., into clinically meaningful criteria, i.e. changes in test parameters which will evoke a change in treatment regimens. The various studies of "medical usefulness" to date have resulted in a steadily improved mutual understanding of the problems but no truly definitive answers.

Laboratorians have recently resorted to speaking of clinicians as their "customers". This popular concept is based in the industrial quality assurance doctrines of Demming, Juran and others. The concept of the physician as a customer implies not only the requirement on the part of the supplier (laboratory) to satisfy the perceived needs of the shopper, but also the idea that the clinician is free to seek the most appropriate service from the most appropriate (or convenient) source.

It is most timely for the Board of NORDCHEM to sponsor a seminar and study to assess the need for Quality Specifications which relate to Medical Usefulness (i.e. clinical needs). We as laboratorians have, for over forty years, and with good reason, focused our efforts reducing the total error in the analysis processes. In view of the obvious benefits to be obtained by reducing error (bias) and improving precision (reducing CV) this was indeed "doing the thing right". The very existence of this Nordic Seminar on Quality Specifications indicates that it is time for laboratorians to assume the leadership role, that is "do the right thing". In today's health care environment, particularly with respect to diminished resources in all areas, including laboratories, doing the right thing will require us to allocate resources, including quality improvement and test development resources, to those areas of greatest potential benefit to patient and clinician.

As today's seminar clearly demonstrates, this is not a singular activity of laboratorians, or clinicians; it must be a joint effort of both. The effort must be initially channeled into finding the right thing to be done. This may include developing new tests; increasing accuracy and precision; improving test quality as characterized by turnaround time, sampling techniques or interpretive reporting or any other attributes of "total quality". This is leadership; but more basically, this is communications. Once new test parameters are defined, once error specifications, based in medical needs are established, laboratorians can manage the process. We can be counted up on to "do the thing right". This has always been our forte; it will continue to be.

SETTING QUALITY SPECIFICATIONS FOR LABORATORY TESTS

I would like to share some of our current research (with Sharon S. Ehrmeyer) with the group. We have, in the United States, inadvertently taken a new and interesting approach to setting quality specifications. This has come about because our federal government has determined that proficiency testing (PT) will become the major criteria by which laboratories are licensed by the federal government. PT is the process whereby an agency of

government sends unknown specimens to laboratories who analyze them and report back their results. If the laboratory gets the right answers, it passes, if not it fails. A laboratory must pass proficiency testing if it is to be reimbursed for tests it performs. The governmental regulations have been proposed in final form as of March 14, 1990. These regulations are very relevant to the theme of this NORDKEM conference in that they are in effect minimum intralaboratory performance standards for US laboratories. We suspect that these standards are not based on medical need, or laboratory capabilities but rather an intuitive approach to achievable levels of quality.

However, under the regulatory approach which is utilized, it is a logical assumption that a laboratory will seek to improve its performance to a level commensurate with consistently passing proficiency testing. These criteria then, in effect, have become US national minimum performance standards for clinical laboratories.

MATERIAL AND METHODS

We have demonstrated previously that it is possible to determine the minimum intralaboratory performance levels necessary to meet the requirements specified in the Federal rules for PT programs. Under the plausible and tractable assumption of Gaussian imprecision, these relationships can be established by computer modeling using a Monte Carlo simulation approach, or by direct calculation based on statistical analysis. In both approaches a laboratory's internal performance characteristics, that is, 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.

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 determines the probability of passing one PT event under the new five sample per shipment format, modeled as a Bernoulli trial.

TABLE 1

Analyte or Test (routine chemistry)	Criteria for acceptable performance (target value \pm)	
	Old units	SI units
<u>Enzymes</u>		
ALT/SGPT	20%	
ALP	3 SD	
Amylase	3 SD	
AST/SGOT	20%	
CK	3 SD	
CK isoenzymes	3 SD or MB elevated (+ or -)	
LDH	20%	
LDH isoenzymes	3 SD or LDH1/LDH2 (+ or -)	
<u>Blood gas</u>		
pO ₂	3 SD	
pCO ₂	5 mm Hg or 8%	
pH	0.04	
<u>General</u>		
Albumin	10%	
Bilirubin, total	0.3 mg/dL or 20%	5 μ mol/L
Calcium, total	1.0 mg/dL	0.25 mmol/L
Cholesterol, total	15%	
Cholesterol, HDL	3 SD	
Creatinine	0.3 mg/dL or 15%	23 μ mol/L
Glucose	6 mg/dL or 10%	0.33 mmol/L
Iron, total	20%	
Total Protein	10%	
Triglycerides	3 SD	
BUN	2 mg/dL or 9%	
Urea	43 mg/L or 9%	0.71 mmol/L
Uric acid	17%	
<u>Electrolytes</u>		
Chloride	5%	
Magnesium	25%	
Potassium	0.5 mmol/L	
Sodium	4 mmol/L	

The new regulations specify that "passing performance" in a PT event (one quarterly shipment) requires that four out of the five results for each analyte (80% of the results) must fall within the defined acceptable range (target value \pm PT limit). The target value can be the mean of the results from a group of participants using the same method or instrument once outliers have been removed, 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. From Table 1, when "grading" PT results, the acceptable range for glucose is the target value \pm the performance criterion [6 mg/dL (0.33 mmol/L) or 10%] yielding the greater range. In the case of a PT specimen with a target value of 100 mg/dL (5.56 mmol/L), acceptable performance would be between 90 (5.0) and 110 mg/dL (6.11 mmol/L). If the target value were 50 mg/dL (2.78 mmol/L), the acceptable range would be 44 (2.44) to 56 mg/dL (3.11 mmol/L).

FAILURE TO PASS PT - SINGLE ANALYTE

If a laboratory does not achieve at least 80% acceptable performance on any given analyte (4 or 5 correct results) for a PT event, 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 term used in the regulations), the laboratory must again fail to achieve acceptable performance for the same analyte on one of the next two PT events. 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 event, uric acid for the next, and CK for the third, etc., indefinitely without being suspended in the subspecialty of routine chemistry.

FAILURE TO PASS PT - MULTIPLE ANALYTES

In addition to passing 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 less than 80% of all results correct for any two of three consecutive PT events. It is obvious that to fail this, a laboratory must also fail at least one analyte, i.e., have 2 or more incorrect out of 5 results.

Intralaboratory Performance required to pass one PT event for one analyte, zero bias.

The right curve in figure 1 shows the probability of failure, i.e., achieving less than 80% correct for a laboratory analyzing only one analyte in any one PT event. 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 event in which it analyzes 5 PT samples. The right curve in figure 1 is based on the assumption that the laboratory has zero bias, i.e., any deviation from the target value is due only to the laboratory's internal imprecision. Similarly, for analytes whose performance criteria (Table 1) are defined as multiples of the group SD, i.e., 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 event for one analyte drops just below 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.

% chance of some PT failures, related to CV
one 2 of 5 event, bias=0% of PT limit

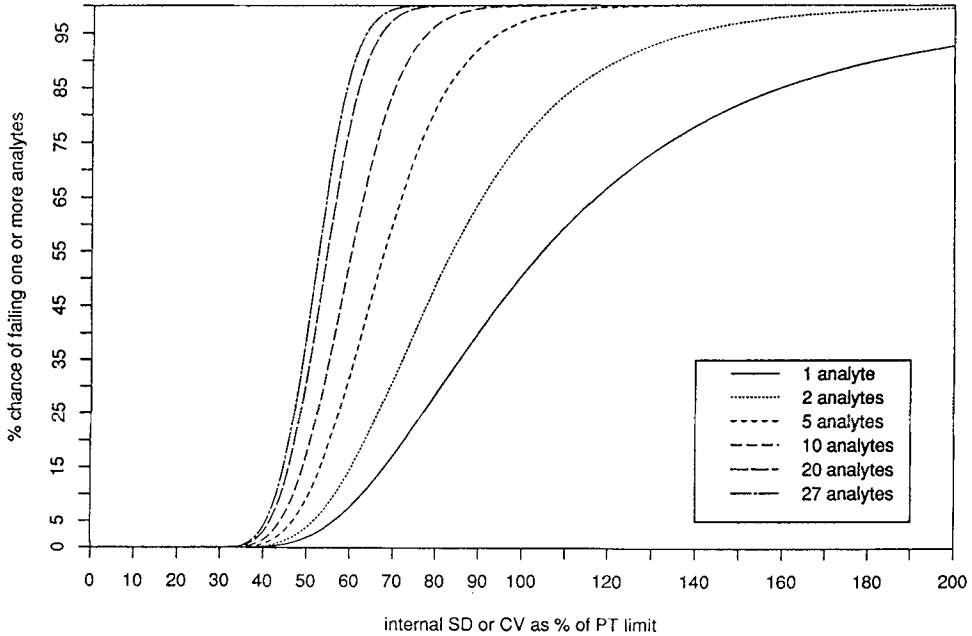


Figure 1.

Intralaboratory performance required to pass one PT event for multiple analytes, zero bias.

Figure 1 with the rest of the curves shows the effect of analyzing multiple analytes (i.e., glucose, BUN, 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, and its internal CV is 100% of the PT performance criteria for both, the chance of failing at least one analyte increases from 51 to 76%. A laboratory doing 20 analytes, i.e., operating a large, multi-channel instrument (DuPont aca,TM Hitachi,TM SMAC,TM etc.) with all the analytes' CVs equal to 100% of the acceptable performance criteria, would virtually be assured of failing at least one analyte on every PT event. By reducing all of these internal CVs to 50%, the probability of a failure for the same 20 test laboratory is reduced to 32%. With all CVs below the 33% level, the chances of failure are nearly zero. Obviously, a laboratory's CVs are not consistent across all parameters; some may be

considerably less than 33%, and these tests would cause no problems in PT. However, even 2 analytes near 50% of the PT criteria would cause a 4% probability of a failure, and 2 analytes at the 100% level would portend an 76% chance of a failure. The obvious conclusion is that a laboratory should strive to reduce all its tests' internal CVs to less than 33% of the Table 1 performance criteria; but in particular to concentrate on reducing the imprecision of any tests whose intralaboratory CVs approach 100% of the PT criteria.

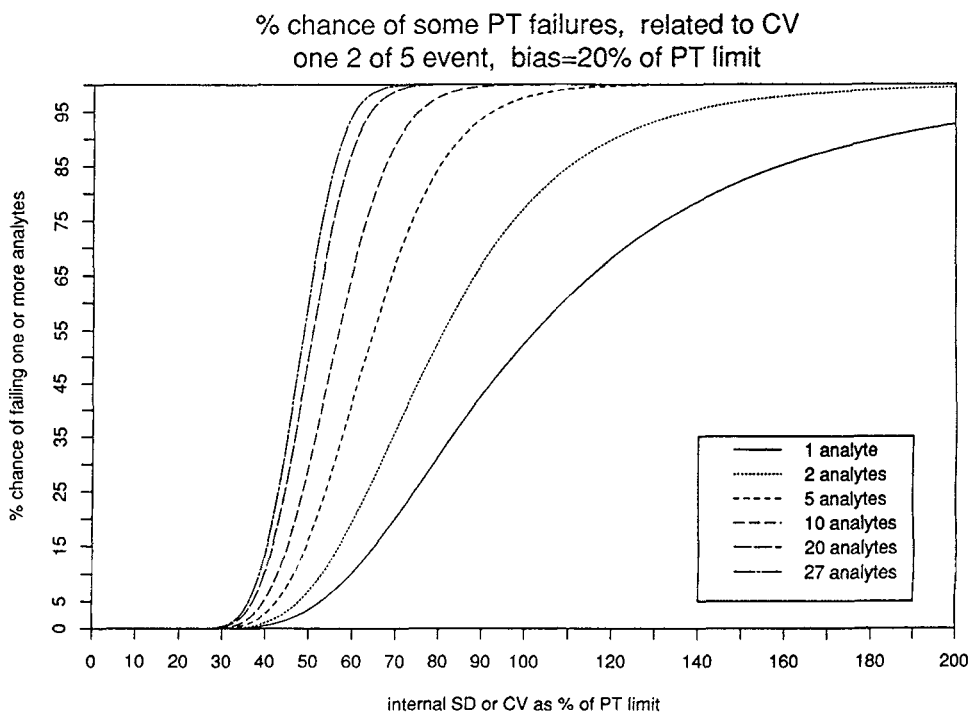


Figure 2.

Intralaboratory performance required to pass one PT event for non-zero bias.

Figure 2 shows the effect of coexisting bias (20% of the PT criterion) for relevant levels of imprecision on the likelihood of a laboratory failing a single PT event. Like Figure 1, the family of curves represents respectively (from the right to left) the probabilities of failure for 1, 2, 5, 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 $\pm 10\%$, a 20% bias is equivalent to a consistent 2 mg/dL (0,11 mmol/L) error. Comparing figures 1 through 3 indicates the effect of increasing bias on the likelihood of failing a PT event. For one analyte, with biases of 0%, 20% and 50% and a consistent coexisting internal CV of 50% of the PT criterion, the probability of failure increases from 2% to 4% to 18%. Further, for 20 analytes, the probability of a failure for biases of 0, 20 and 50%, increases from 32% to 51% to over 98%. While a laboratory does not have the same, or for that matter, any bias on every test, it is obvious that the presence of any significant bias seriously impairs the laboratory's ability to pass that analyte in a PT event. A common reason for analytical failures is the introduction of a large bias rather than large imprecision. Due to a pipetting error or a reconstitution problem bias may extend to all analytes and impose a large chance of an 80% failure. Consequently, to pass PT, a laboratory first needs to minimize the amount of bias, and then reduce its internal CV if possible.

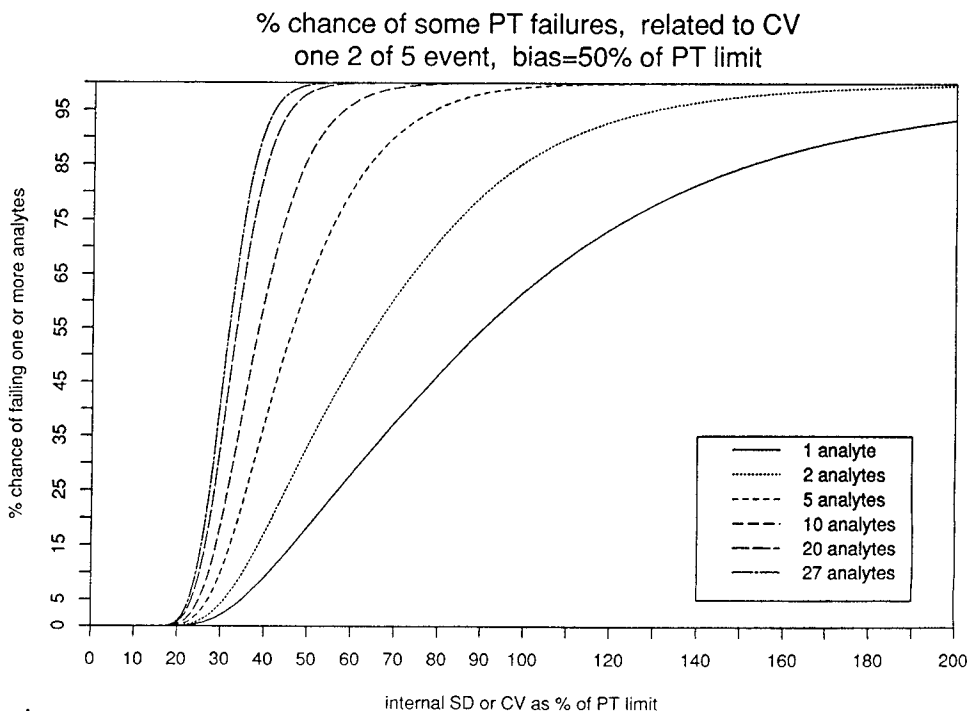


Figure 3.

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, since 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 (left) of the line have a lesser chance of failure. Likewise, further to the right, subsequent curves denote the probabilities of failure of 3, 5, etc., percent for increasing values of the CV and bias. Typically laboratories using reasonable care in calibration have small biases. Those laboratories whose bias exceeds 20% of the performance limit and whose CVs are in the range of 30 to 100% of the performance limit need to reduce both.

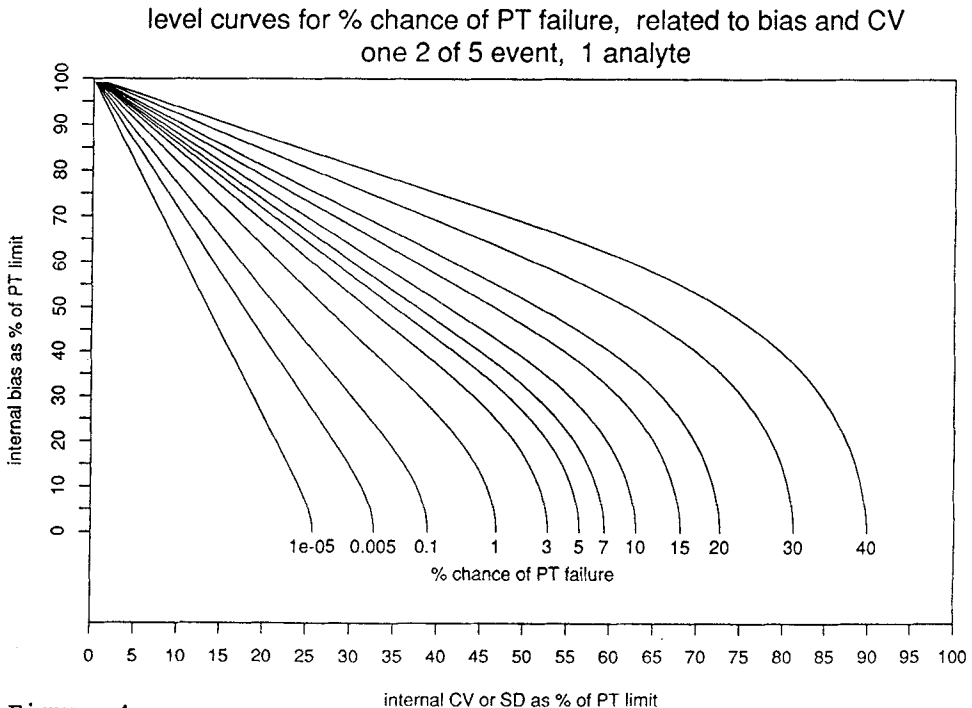


Figure 4.

Minimizing the bias contribution to the probability of failure in proficiency testing.

Bias does not effect the likelihood of failure to nearly the same extent as CVs of equivalent size. For example if a laboratory reduces its CV to 33% of the value in Table 1, for any given analyte, a coexisting bias of up to 40 percent is tolerable. If the bias can be reduced to any value below 20%, it's contribution to the probability of failure to pass PT is almost negligible. Most authors do not deal with the concepts of coexisting bias and imprecision, but rather set bias equal to zero. This is a justifiable assumption at least when dealing with PT data. In point of fact, the traditional function of PT programs has been to reduce bias since even with five samples, PT is ineffective in measuring intralaboratory imprecision.

CONCLUSION

A laboratory readily can predict the probability of passing PT based on a knowledge of its internal imprecision (CV) and bias. Our overwhelming conclusion for the proposed 2 of 5 (or 80%) PT rules is that a laboratory with small (<20%) bias can reasonably assure its likelihood of passing PT if, for each analyte, it reduces the imprecision of each analyte to one-third of the Federally mandated limit.

This "rule of 1/3" when applied to the federal criteria in Table 1, represents performance limits for US laboratories. Since the ability to receive revenues for tests performed depends on passing proficiency testing, these limits, without respect to medical usefulness or significance, will become the goal of laboratory's quality assurance efforts.

While one might dispute the logic of setting performance limits in this manner; there is a valuable lesson to be learned at the same time. External (or interlaboratory) quality assurance programs which closely mimic PT programs in concept and format, can translate analytical goals into performance limits. For example, if a decision could be reached that the total error requirement for Glucose tests should be 3.3 mg/dl (0.18 mmol/L); a scheme of interlaboratory quality assurance testing could be computer modeled to determine the criteria (in the case of our

example $\pm 10\%$ as in Table 1) to be used to assure successful participants that they met the minimum internal performance requirement. The regulatory approach represents only one use of proficiency testing. Properly managed within laboratories or by concerned professional groups, it can be a powerful quality assurance tool.

As the US standards outlined in this report are implemented beginning January 1, 1991, we will have an excellent opportunity as scientists to evaluate one of the tools which can be used for managing (or leading) the process of searching for meaningful performance goals for our clinical laboratories.

Correspondence:

Professor Ronald H. Laessig,
State Laboratory of Hygiene,
University of Wisconsin Center for Health Sciences,
Madison, Wisconsin, USA.