Analytic Performance Goals Based on Direct Effect of Analytic Bias on Medical Classification Decisions

George G. Klee, M.D., Ph.D.
Department of Laboratory Medicine
Mayo Clinic and Mayo Foundation
Rochester, Minnesota

Abstract: A major clinical use of laboratory tests is for classifying patients into diagnostic and treatment categories. Multiple factors influence the decision limits used for these decisions; however, the analytic performance of the assays seldom is explicitly considered in these decisions, even though both imprecision and bias may significantly alter the decisions. Of these two factors, changes in analytic bias have the most dominant effect. Assay imprecision adds only indirectly to the overlap of the distributions of test values by increasing the variance. The analytic contribution to the total variance is indirect because it is combined with biologic variations, which serve to buffer any changes.

Analytic bias directly affects classification decisions by shifting the distribution of test values. This effect is greatest for values near the decision levels where the bias may alter classification. Thresholds for medical decisions generally are determined using data collected when the assays are initially calibrated; if the assays shift or are recalibrated to a different level, the number of patients exceeding the decision thresholds are directly increased or decreased. Unlike precision problems, repeat testing does not help to minimize these misclassifications. For example, decisions to pursue hyperparathyroidism often are triggered by calcium levels exceeding a defined threshold. Small analytic shifts upward markedly increase the number of patients investigated.

Analyzing variation in the percentages of patients exceeding selected thresholds during times when the assays are analytically unbiased can provide reference standards for establishing tolerance limits for analytic bias. All levels of analytic bias will directly affect the medical decisions, but assays in which the bias is held within these tolerance limits should not substantially alter the number of patients misclassified. Maintaining analytic bias within these tolerance limits then becomes the primary analytic performance goal. Secondary goals for assay precision can be defined in terms of the quality control systems required to maintain the bias goals.

Introduction

The two major analytic quality control parameters are imprecision and inaccuracy.[1] The EFCC expert panel defines imprecision as the "standard deviation or coefficient of variation of the results in a set of replicate measurements."[2] Inaccuracy is defined as the "numerical difference between the mean of a set of replicate measurements and the true value." Westgard has used the terms "random error" for imprecision and "systematic error" or bias for inaccuracy.[1]

For many clinical chemistry analytes, reference methods and reference standards do not exist. Therefore, it is not possible to define "true values." In this discussion the term “bias” is used to describe the systematic error representing the mean analytic difference between the current measurement system and the system that was used to

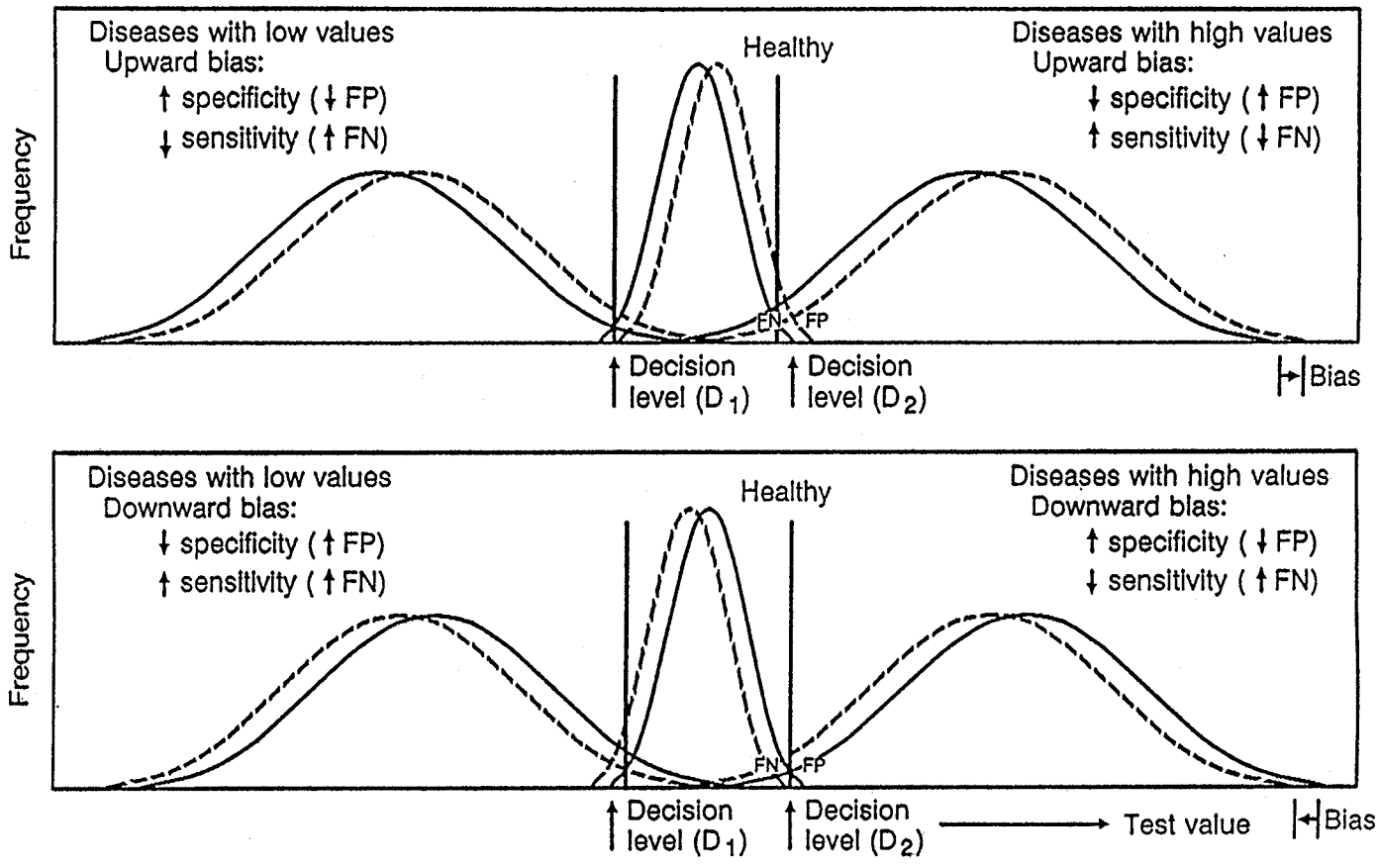

Figure 1. Effect of analytical bias on the classification of patients.

establish the clinical decision limits. Analytic bias shifts both the disease and reference populations relative to these predefined decision limits. The classification of patients is directly altered by this analytic bias.

Two terms used to describe the performance of analytic assays for classifying patients are “sensitivity” and “specificity”. Sensitivity is an assessment of "the percentage of patients with the disease who exceed the decision level."[3] Specificity is "the percentage of patients without the disease who are within the decision level." Sensitivity and specificity depend on both the distribution of the test values (in disease and reference population, respectively) and the decision levels. Changes in the decision levels either increase sensitivity and decrease specificity or vice versa. Changes in the decision levels cannot increase or decrease both sensitivity and specificity.

Analytic imprecision has minimal effect on classifying patients. Imprecision causes some values to be falsely high and others to be falsely low, thereby canceling out many of the classification errors. The analytic imprecision adds to the overall biologic scatter of the reference and disease populations. This broadening of the distribution functions causes a minor decrease in both sensitivity and specificity. This effect generally is minor because the analytic standard deviation (SD) adds to the population SD by the sum of squares:[4]

$$\text{SD total} = SD^2_{\text{analytic}} + SD^2_{\text{population}}$$

As long as the analytic SD is less than one-half of the population SD, the total SD will not increase by more than 12%:

$$\begin{aligned}\text{SD total} &= [(O.5y + I] \times Sd_{\text{population}} \\ &= l.25\ SD_{\text{population}} \\ &= 1.12\ SD_{\text{population}}\end{aligned}$$

Tonks' "allowable limit of error" set at one-fourth of the reference range can be directly tied to this concept if one estimates the reference range as 4 $SD_{\text{population}}$ (i.e., mean ± 2 SD).[5] Tonks' limits allow approximately twice the error described by the ½ SD formula because it corresponds to 1.0 $SD_{\text{population}}$ rather than 0.5 $SD_{\text{population}}$.

Effect of Bias on Sensitivity and Specificity

Figure 1 shows that analytic bias directly affects the classification of patients. Unlike imprecision, analytic bias does not combine with the population scatter but directly shifts both the reference and the disease population. The effects on sensitivity and specificity depend on the direction of the bias and the relative position of the disease and reference population (Table 1).

The effects of analytic bias are equivalent to the effects of changing the decision limits. A reciprocal interchange exists between increased sensitivity and decreased specificity or decreased sensitivity and increased specificity with changes in decision limits.

In most clinical practices more non-disease patients (reference population) are found compared with the disease population. Also, the distributions of test values in the reference population generally is steeper (more leptokurtic) than the distribution in the disease population. This causes analytic bias to have much greater effects on changes in specificity than on changes in sensitivity.

Consider an example of serum calcium in classifying patients with parathyroid adenomas versus healthy reference subjects. If we use 10.2 mg/dL as the decision limit,

Position of Disease Population	Direction of Bias	Sensitivity	Specificity
Above reference population	Upward	↑	↓
Above reference population	Downward	↓	↑
Below reference population	Upward	↓	↑
Below reference population	Downward	↑	↓

Table 1. Effect of bias on sensitivity and specificity.

98% of the healthy patients have values below this limit (specificity), while 99% of the patients with surgically proven parathyroid adenomas have values equal or above this limit (sensitivity). If there is a 0.3 mg/dL upward bias, then only 92% of the healthy subjects have values above the decision limit (equivalent to 9.9 mg/dL, unbiased). This bias decreases the specificity by 6%, whereas the sensitivity increases by less than 1%. Also since there are about 300 patients tested for every one case of hyperparathyroidism identified, the net effect of the 0.3 mg/dL upward bias in calcium is that a large number of patients are subject to additional investigations unnecessarily.

Astute clinicians working in specialty medical clinics often can detect subtle analytic shifts before changes become apparent in routine laboratory quality control (QC) systems. In the example of the calcium shift, the specialty clinician probably would call the laboratory to inquire about possible analytic problems before subjecting numerous patients to further investigation. This clinician feedback is an important quality control parameter.

Proposed System to Define Medically Important Bias Limits

Specialty clinicians can detect analytic shifts by noting the increase in the prevalence of patients having test values exceeding their decision thresholds. Since there normally is a variation in the number of cases with values exceeding their action limits, the perception of a problem arises when this number exceeds the usual variation. It is proposed that this usual variation in the percentage of the patient population which exceed selected action limits be used to define medically important bias limits.

Statistically, the following procedure is proposed to calculate these bias limits: Test distributions are collected for 20 consecutive periods when the laboratory is operating without known bias. It is proposed that these test groups be composed of approximately 1000 test values each to provide reasonable estimates of the tails of the distributions. The mean and SD are calculated for the percentage of each distribution which exceeds selected decision thresholds. For analytes that do not have specific decision thresholds, the upper and lower normal value limits can be used as

decision thresholds. The composite frequency distribution for all 20,000 data points also is constructed and used to related cumulative frequency percentages to analyte concentrations.

Figure 2 shows a composite cumulative frequency curve for serum calcium. The average percentage of patients not exceeding the 10.2 mg/dL threshold for the consecutive distributions was 98.18% with an SD of 0.48 mg/dL. The mean plus and minus 2 SD range for the frequency distribution not exceeding the decision limit was 97.2 to 99. 1 %. This related to a calcium range of 10.09 to 10.39 mg/dL. Alternately, this tolerance range can be stated as ± 0. 15 mg/dL from the unbiased set-points.

If one can hold the analytic bias at less than one-half of variation seen across the patient population, the bias should not be perceptible by even the most astute clinicians. The proposed bias limits for the analytes therefore are set at the analyte ranges corresponding to ± 1 SD limits for the percentage of the consecutive population distributions exceeding the decision limit.

Proposed System to Define Precision Limits Based on QC Systems to Reliably Hold the Bias Limits

The statistical power of quality control systems to detect analytic bias depend on the algorithm used and the size of the bias relative to the precision of the assay.[6] Therefore, for a given QC algorithm (such as the Westgard multi-rule function), the analytic bias limits can be used to define the precision goals. The more precise an assay, the smaller the bias that can be reliably detected.

Serum calcium will be used as a specific example for calculating analytic precision goals. If we use Westgard's multi-rule control procedure with n = 6, 90% power exists for detecting a bias drift equal to twice the analytic precision expressed as an SD. Therefore, to achieve the 90% power of detection bias shifts equal to the previously defined goal, the analytic precision goal is set at one-half of the bias goal:

Calcium bias goal = 0.08 mg/dI.
Calcium precision goal (SD) = 0.04 mg/dL
Calcium precision goal (CV) =- 0.4%

Discussion

Most of the previous investigators have proposed using biologic variation to establish goals for analytic precision. The early work of Tonks related precision to the reference range, which is directly tied to across-person biologic variability.[5] A more statistical approach relating assay precision to both inter- and intra-individual biologic variation was later conducted by Cotlove and Harris.[7,8] The linkage of analytic precision to biologic variation also was a major emphasis of the Aspen Conference.[9] Fraser and Stockl further expanded these concepts,[10,11] but the major focus of these papers is analytic precision.

Recently, Petersen explained the importance of analytic bias in analytic decisions.[12] He also proposed linking bias goals to their effects on medical decision processes. His emphasis on analytic bias is the same as advocated in this paper, but his approach to goal setting is different.

The concept described here for defining bias and precision can be applied to most all quantitative analytic laboratory measurements including chemistry and hematology. Preliminary studies indicate some of these goals are more stringent than our current systems provide (such as for serum calcium), but others should be easily

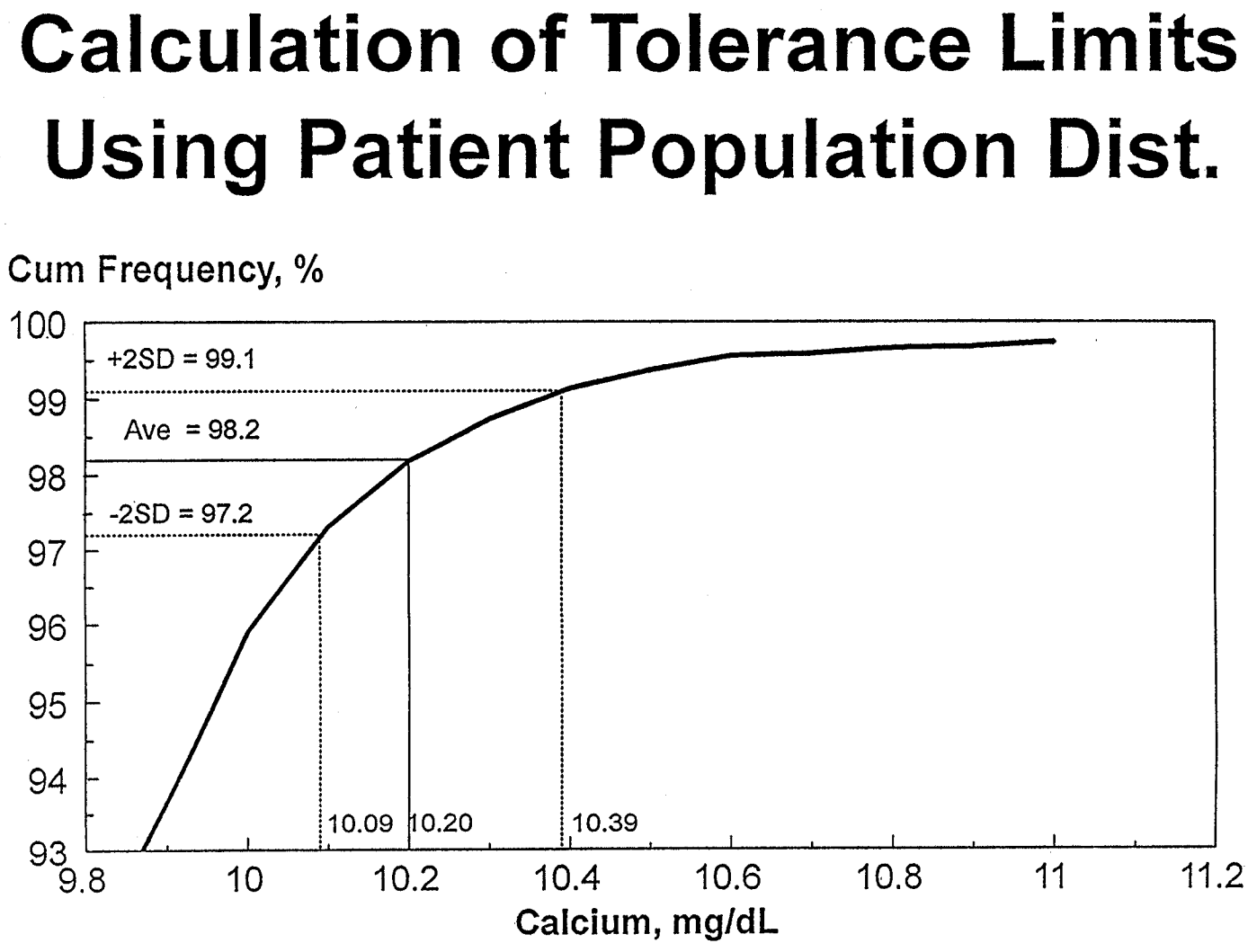

Figure 2. Calculation of tolerance limits using patient population distribution.

met with less control monitoring than currently used.

Conclusion

Analytic bias has a major effect on the diagnostic, prognostic, and therapeutic classification of patients. The analytic bias generally has a more pronounced effect on clinical specificity than on sensitivity. These observations were used to develop an approach to goal setting for analytic bias, which is based on variation of normal reference population test distributions. Analytic precision goals then are calculated based on the statistical power of the quality control systems needed to maintain the bias goals.

References

1. Westgard JO. Precision and accuracy: concepts and assessment by method evaluation testing. *CRC Crit Rev Clin Lab Sci*. 1981;13:283-330.

2. Buttner J, Borth R, Boutwell JH, Broughton PMG. International Federation of Clinical Chemistry provisional recommendation of quality control in clinical chemistry. 1. General principles and terminology. *Clin Chim Acta*. 1975; 63:525.

3. Galen RS, Gambino SR. Beyond normality: the predictive value and efficiency of medical diagnoses. John Wiley & Sons, New York, 1975.

4 Westgard JO**,** Klee GG. Quality management. In: Tietz Textbook of Clinical Chemistry, Second Edition. W. B. Saunders, Philadelphia, 1994: 548-592.

5. Tonks DB. A quality control program for quantitative clinical chemistry estimations. *Can J Med Technol.* 1968;30:38-54.

6. Westgard JO, Barry PL, Hunt MR. A multi-rule Shewhart chart for quality control in clinical chemistry. *Clin Chem*. 1981;27:493-501.

7. Cotlove E, Harris EK, Williams GZ. Components of variation in long term studies of serum constituents in normal subjects. III. Physiological and medical implications. *Clin Chem*. 1970;16:1028-32.

8. Harris EK. Statistical principles underlying analytical goal-setting in clinical chemistry. *Am J Clin Pathol*. 1979;72:374-82.

9. Elevitch FR, ed. College of American Pathologists Conference Report Conference on analytical goals in clinical chemistry, at Aspen, Colorado. Skokie, IL. CAP, 1977.

10. Fraser CG, Hyltoft Petersen P. Desirable standards for laboratory tests if they are to fulfill medical needs. *Clin Chem*. 1993;39:1447-55.

11. Stockl D, Baadenhuijsen H, Fraser CO, Libeer J-C, Hyltoft Petersen P, Ricos C. Desirable routine analytical goals for quantities assayed in serum. *Eur J Clin Chem Clin Biochem*. 1995;33:157-69.

12. Hyltoft Petersen P, de Verdier C-H,

Groth T, Fraser CG, Blaabjerg 0, Horder M. On the influence of analytic bias on the outcome of clinical strategies. In press, *Clin Chim Acta*. 1995.