A Method Evaluation Decision Chart (MEDx Chart) for Judging Method Performance

JAMES O. WESTGARD

OBJECTIVE: To introduce a new graphical tool that improves the process of making decisions about method performance.

DATA SOURCES: Scientific literature, mathematical models, and the author's experience.

STUDY SELECTION: Not applicable.

DATA EXTRACTION: Not applicable.

DATA SYNTHESIS: Relationship to charts of operating specifications (OPSpecs charts) provides guidance for classification of performance as poor, marginal, good, or excellent.

CONCLUSION: Objective decisions on the performance of methods can be made quickly and easily with the aid of a Method Evaluation Decision Chart.

KEY WORDS: Method evaluation; Total error criteria; Analytical quality requirements; Quality control; Operating specifications.

ABBREVIATIONS USED: s = Standard deviation; CV = Coefficient of variation; TE = Total error; TE_a = Allowable total error; SE = Systematic error; PT = Proficiency testing; QC = Quality control; OPSpecs = Operating specifications; CLIA = Clinical Laboratory Improvement Ammendments; AQA = Analytical quality assurance; EP = Evaluation Protocol; NCCLS = National Committee for Clinical Laboratory Standards; PCA = Portable clinical analyzer; ED = Emergency department staff; CP = Clinical pathology; TV = Target value.

Clin Lab Sci 1995; 8:277-283.

Method evaluation studies are widely performed in laboratories today, but many clinical laboratory scientists still find it difficult to make objective judgments or decisions on the acceptability of method performance. The development of standard evaluation protocols, such as the Evaluation Protocol (EP) series from the National Committee for Clinical Laboratory Standards (NCCLS), has helped define the data to be collected, identified the statistics to be calculated, and provided some guidance for comparing a laboratory's estimates of performance against a manufacturer's claims. Unfortunately, the question of how to judge acceptability for clinical needs is not answered in these or similar documents, which leaves laboratorians with the responsibility to decide for themselves whether the observed performance characteristics satisfy the quality requirements of the test. This responsibility is appropriate, but the lack of guidance on how to make these decisions may compromise otherwise worthwhile evaluation studies.

Common practice today seems to be to accept a method if it meets the manufacturer's own claims for performance, which assumes that the manufacturer has carefully determined what performance is necessary to assure the quality required in the application of the test. Little evidence exists to support this assumption. No statement about clinical or analytical quality appears in the standard protocols accepted by industry and government. Instead, these protocols focus on the data needed to make claims for the performance observed under stable operation for certain characteristics, such as

James O. Westgard, Ph.D., is the guest editor for this Focus topic.

James O. Westgard, Ph.D., is a professor in the Department of Pathology and Laboratory Medicine, University of Wisconsin Medical School, Madison, WI 53792.

Address for reprints: James O. Westgard, Room D4/237, University of Wisconsin Hospital, 600 Highland Avenue, Madison, WI 53792.

Focus CE Credit: The other components of this Focus topic and the continuing education materials will be published in the November/December issue of *CLS*.

precision, accuracy, interference, recovery, detection limit, working range, and reference range. Such an approach facilitates the comparison of the performance between different systems and likely leads manufacturers to set product specifications based on what is needed to be competitive in the marketplace, rather than the actual clinical or analytical quality required for a test.

Another limitation is that manufacturers' claims represent the performance observed when the method is working properly, i.e., stable operation. For laboratory users, it would be of more interest to know the performance expected under routine operating conditions, which likely include some incidents of unstable operation. Unstable operation is supposed to be eliminated by quality control (QC), assuming these incidents can be detected by the QC procedures recommended by the manufacturers or those commonly used by laboratories. Again there is little evidence to support this assumption, since neither manufacturers or laboratories document that medically important errors can be detected by their QC procedures. Thus, the existing approach doesn't address the real issue of the performance required to assure quality under routine operating conditions. "Buyer beware" continues to be good advice in the laboratory marketplace!

Historically, decisions on method performance were based on statistical tests of significance (such as the *t*-test and the *F*-test) and the correlation coefficient.[1] In the 1970s, we studied the application of these statistics and found they had limited value for estimating the random, proportional, and systematic errors that are the focus of method evaluation studies.[2] We developed guidelines for obtaining more reliable estimates of errors and recommended that these estimates be compared to the amount of error allowable in the medical use of the test.[3] It was proposed that the allowable error be defined in the form of a total error (TE) that reflected the combined effects of imprecision and inaccuracy on the test result. We argued that it made little difference to the user (physician) and consumer (patient) whether a test result was inaccurate or imprecise; what was important was how far wrong it could be—the total error that includes both inaccuracy and imprecision.

During the late 1970s, a series of papers were published in a precursor of this journal to help establish these concepts and practices of method evaluation more widely in clinical laboratories.[4] These ideas were also discussed in an extensive review of the practices of method evaluation that summarized the state of the art in the early 1980s.[5]

It took many years for the concept of total error to become accepted in U.S. laboratories. Now, with the implementation of CLIA proficiency testing (PT) criteria, allowable total errors (TEa) have been defined through regulations for approximately 80 laboratory tests. In other countries, specifications for laboratory performance may still be expressed in terms of allowable imprecision and allowable inaccuracy, as illustrated by the recent proposal by a European group.[6] Even in this case, Stöckl argues that the European specifications are actually a "simplification of the total error model, splitting it into its two extreme cases, absence of imprecision and absence of bias"[7], thus the concept of total error seems to be behind these recommendations as well.

In applying total error criteria to method evaluation studies, one issue is which criterion to adopt. The first recommendation[3] in 1974 was that total error be estimated as the inaccuracy (bias) plus 2 times the imprecision (s). In 1990, Ehrmeyer et al.[8] recommended a TE criterion of bias+3s and Westgard and Burnett[9] recommended bias+4s. The argument for using the 3s TE criterion is that multiple testing of up to 27 analytes for CLIA PT compliance requires this level of performance to avoid false rejections from PT. The argument for using the 4s TE criterion is that internal quality control procedures lack the sensitivity to detect medically important changes unless the stable performance of a method is very precise.

Another issue is how to best compare method performance to the TE criterion. Statistical expressions and calculations, though commonly used, are often difficult to understand and correctly interpret. A graphical representation of the bias and imprecision might provide a clearer picture of method performance relative to the quality required. The form of a graphical tool that could be applied for evaluating method performance is suggested by the chart of operating specifications (OPSpecs chart), which shows the precision and accuracy that are allowable and the QC procedure that is necessary to assure achievement of a defined quality requirement.[10] OPSpecs charts have been derived from quality requirements in the form of an analytical total error and a clinical decision interval, thus this approach permits method performance to be evaluated based on both analytical and clinical quality requirements. Furthermore, the OPSpecs chart allows a laboratory to simultaneously verify that method performance and QC performance are both adequate for the quality required for the test, which should improve the technical management of analytical quality.

My objectives here are to (a)introduce a simple Method Evaluation Decision chart (MEDx chart) that can be used to improve decision-making in method evaluation studies, (b)show the relationship between the criteria on the MEDx chart and the operating specifications that are desirable to adequately QC a laboratory testing process, and (c)illustrate the use of the chart for some simple applications.

METHOD EVALUATION DECISION CHART (MEDx CHART)

Figure 1 shows a MEDx chart for a TE requirement of 10%. The chart displays the allowable inaccuracy on the

FIGURE 1. A Method Evaluation Decision (MEDx) chart which displays the allowable inaccuracy on the y-axis vs the allowable imprecision on the x-axis. The different decision criteria from top to bottom correspond to TE criteria of bias+2s, bias+3s, and bias+4s. The regions bounded by these decision criteria classify performance as poor, marginal, good, and excellent.

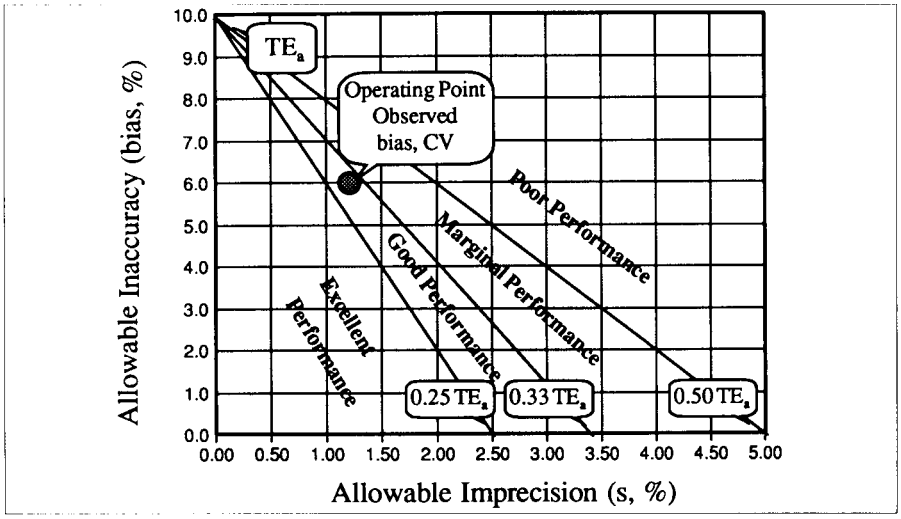
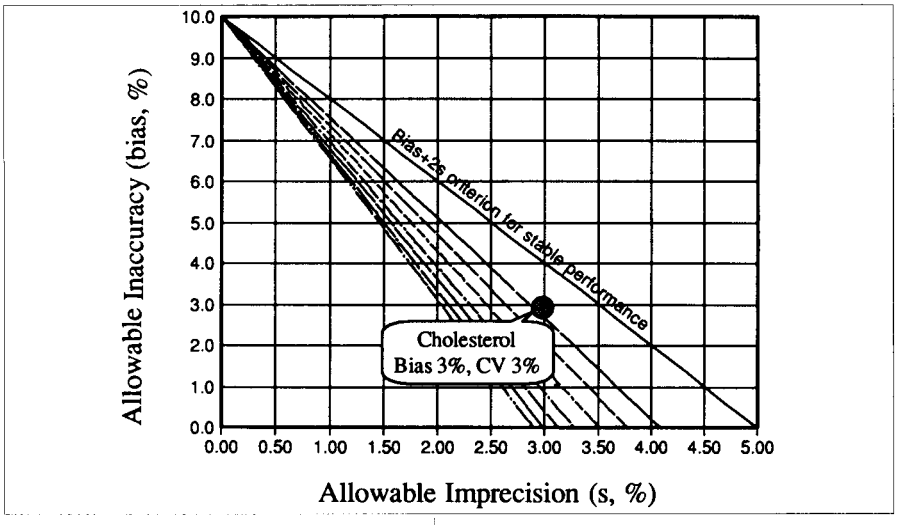


FIGURE 2. Comparison of the bias+2s TE criterion with the operating specifications needed to provide 25% analytical quality assurance (AQA). The other lines represent commonly used QC procedures (from top to bottom): $1_{3s}/2_{2s}/R_{4s}/4_{1s}$ with N=4, $1_{2.5s}$ N=4, 1_{2s} N=2, 1_{3s} N=4, $1_{2.5s}$ N=2, $1_{3s}/2_{2s}/R_{4s}$ N=2, and 1_{3s} N=2. The operating point represents the performance of a cholesterol method having a 3% CV and 3% bias.



y-axis and the allowable imprecision on the *x*-axis. Three different TE criteria are included on the chart.The highest line is for TE = bias+2s, the middle line for bias+3s, and the lowest line for bias+4s, where bias describes the size of systematic error (or inaccuracy) and s is the standard deviation that describes the size of random errors (or imprecision). If a method exhibits errors in the region above the top line, the method is classified as having poor performance; a method in the region between the middle and top lines provides marginal performance; a method in the region between the lowest and middle lines provides good performance, and a method in the region below the lowest line provides excellent performance.

To prepare a MEDx chart for a test of interest, define the analytical quality requirement in the form of a total allowable error (TEa). Express TEa as a percentage of the decision level concentration or target value (TV) of interest. Take a sheet of graph paper, label the y-axis allowable inaccuracy (bias, %) and scale from 0 to TEa; label the *x*-axis allowable imprecision (s, %) and scale from 0 to 0.5 TEa. Draw the line for the bias+2s TE criterion from TEa on the y-axis to 0.5 TEa on the *x*-axis; draw the bias+3s TE criterion from TEa on the y-axis to 0.33 TEa on the *x*-axis; draw the bias+4s TE criterion from TEa on the y-axis to 0.25 TEa on the *x*-axis. Label the regions poor, marginal, good, and excellent, as identified in Figure 1.

To apply the MEDx chart, express the errors observed for the different method performance characteristics as a percent of the decision concentration or target value of interest. Plot separate estimates of imprecision (such as from within-run and within-day studies) on the *x*-axis, assuming a y-coordinate of zero; plot separate estimates of the inaccuracy (such as the errors due to recovery and interference) on the y-axis, assuming an *x*-coordinate of zero. Combine estimates of imprecision and inaccuracy by plotting imprecision as the *x*-coordinate and inaccuracy as the y-coordinate to locate an expected operating point of your method. The most critical application is the operating point that represents the total imprecision estimated from a day-to-day replication experiment and the bias between methods estimated from a method comparison experiment; however, you can also plot an estimate of short-term imprecision and a specific bias component to provide a preliminary operating point that will predict the performance of your method.

RELATIONSHIP BETWEEN MEDx CRITERIA AND OPERATING SPECIFICATIONS

The judgments of poor, marginal, good, and excellent performance are based on the relationships between the various TE criteria and the operating specifications necessary to assure the desired quality is achieved .[10] The objectives in selecting statistical QC

FIGURE 3. Comparison of the bias+3s TE criterion with the operating specifications needed to provide 50% AQA. The operating point represents a cholesterol method having a 2.5% CV and 2.5% bias. (See Figure 2 for the list of QC procedure represented.)

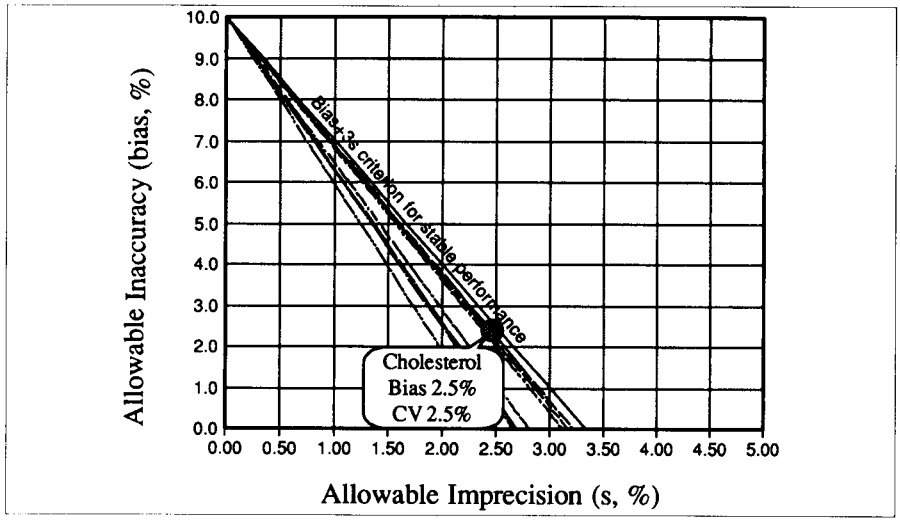
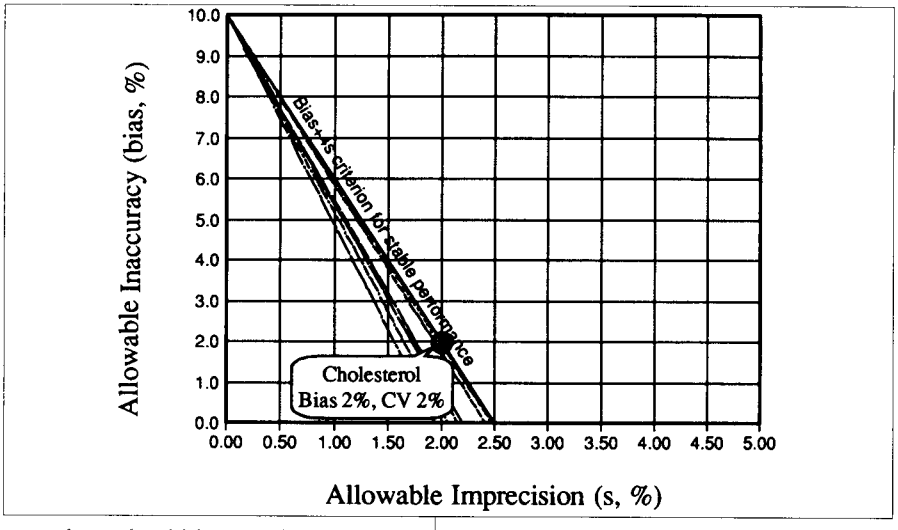


FIGURE 4. Comparison of the bias+4s TE criterion with the operating specifications needed to provide 90% AQA. The operating point represents a cholesterol method having a 2.0% CV and 2.0% bias. (See Figure 2 for the list of QC procedures represented).



procedures should be to achieve 90% detection of medically important errors while maintaining a false rejection rate as low as possible: generally in the 1–5% range. Lower error detection, such as 50% to 25%, may be acceptable for very stable measurement procedures that seldom have problems. It is generally best to aim for 90% detection so you don't have to worry about the stability of the process.

Figure 2 shows the bias+2s TE criterion imposed on an OPSpecs chart for a method such as cholesterol that has TEa of 10% (based on the CLIA PT criterion) and for commonly used control rules having from 2 to 4 control measurements per run (*N*). The operating limits for imprecision and inaccuracy on this chart provide only 25% analytical quality assurance (AQA) that medically important systematic errors (SE) will be detected by the QC procedures. The operating point represents a method whose precision and accuracy satisfies the current National Cholesterol Education Program (NCEP) guidelines of 3% CV and 3% bias. Although this method would satisfy the bias+2s TE criterion, it would not be reliable for daily operation because of the low level of assurance that could be provided using current laboratory QC procedures. Therefore, the performance of this method is classified as marginal. When method performance is stable, i.e., the method is working right, the necessary quality will be achieved. In cases of unstable operation, commonly used QC procedures may not be able to detect medically important errors; thus there is no guarantee that the desired quality will be achieved in routine operation.

Figure 3 compares the bias+3s TE criterion with the operating specifications for 50% AQA(SE). When a method satisfies this TE criterion, such as a cholesterol method with a 2.5% CV and 2.5% bias, there is greater assurance (nearly 50%) that test results will meet the defined quality requirement. This level of assurance may be satisfactory for stable processes that have few errors; thus this method is classified as good performance.

Figure 4 compares the bias+4s TE criterion with the operating specifications for 90% AQA(SE). When a method satisfies this TE criterion, such as a cholesterol method with a 2.0% CV and 2.0% bias, it should be possible to provide nearly 90% assurance that test results will meet the defined quality requirement; therefore, this method is considered to provide excellent performance.

For cholesterol, the MEDx chart classifies performance as poor when the CV and bias both exceed approximately 3.5%, marginal when both values are 3.0%, good for 2.5%, and excellent for 2.0%. A laboratory might consider a method with a 2.5% CV and 2.5% bias to be acceptable if it were performed by an automated system that demonstrates good stability (low frequency of problems, low susceptibility to problems); a system with a 3.0% CV and 3.0% would have to be nearly perfectly stable to be considered acceptable. On the other hand, a method with 2.0% CV and 2.0% bias would generally be judged acceptable because, even with frequent problems, medically important errors would be

detected by commonly used laboratory QC procedures, thus assuring that the desired quality is achieved in routine testing.

EXAMPLE APPLICATIONS

A portable clinical analyzer (PCA) was evaluated to determine its suitability for use in an emergency department.[11] The authors performed an extensive study and provided sufficient data to illustrate how a MEDx chart can be applied. They point out that this analyzer is classified as moderately complex by CLIA regulations and would be subject to regular proficiency testing. Therefore, we can consider the CLIA PT criteria for acceptable performance as the analytical total error criteria that need to be satisfied.

GLUCOSE

According to CLIA, the PT criterion for glucose is ±6 mg/dL (0.333 mmol/L) or 10% of target value, whichever is greater. At a target value of 50 mg/dL (2.775 mmol/L), the 6 mg/dL figure is applicable and is equivalent to 12%. At a target value of 200 mg/dL (11.1 mmol/L), the 10% figure should be used for TEa. This example would require at least two MEDx charts for a complete assessment, but, for convenience, let's focus on the lower decision level.

Replication and comparison of methods experiments were performed by both the laboratory staff (clinical pathology, CP) and the emergency department staff (ED). For CP staff, a CV of 3.78% was observed based on 44 determinations on a stable control material (see level 3 control in Table 1).[11] Comparison of 162 patient samples with a Beckman CX3 analyzer gave a regression line of $y=5.36$ mg/dL$+0.952x$ ($r^2=0.966$, see Table 3).[11] At 50 mg/dL, the PCA would be expected to give a value of 52.96 mg/dL, which would be in error by 2.96 mg/dL, or 5.9%. For ED staff, the CV was 4.14% based on 125 determinations and the comparison study of 527 patients gave a regression line of $y=-2.19$ mg/dL$+1.051x$ ($r^2=0.970$). At 50 mg/dL, the PCA would be expected to give a value of 50.36 mg/dL, which would be in error by 0.36 mg/dL or 0.7%.

FIGURE 5. MEDx Chart for a glucose example where the CLIA PT criterion is 12%. Operating points represent a portable clinical analyzer having a CV of 3.8% with a bias of 5.9%, which is classified as poor performance, and a CV of 4.1% with a bias of 0.7%, which is classified as good performance.

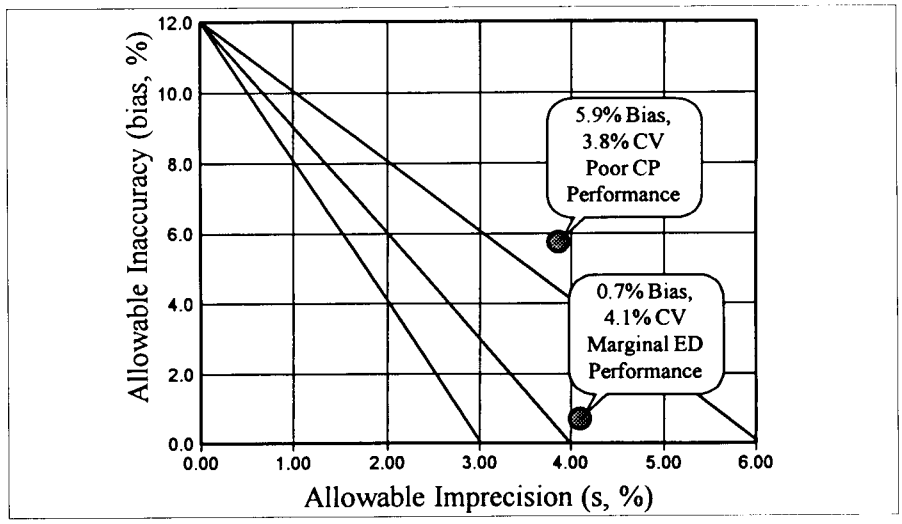


FIGURE 6. MEDx chart for a sodium example where the CLIA PT criterion is 3.0%. Operating points represents a portable clinical analyzer having a CV of 0.49% with a bias of 0.87%, which is classified as excellent performance, and a CV of 0.76% with a bias of 0%, which is classified as good performance.

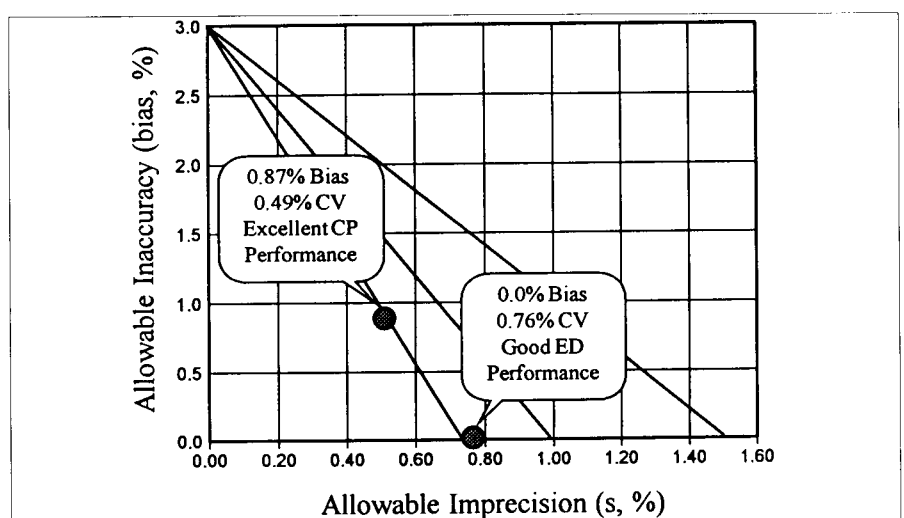


Figure 5 shows a MEDx chart for a 12% TE requirement and the operating points for the studies by CP (y-coordinate=5.9%, x-coordinate=3.8%) and ED (y=0.7%, x=4.1%). CP performance would be classified as poor. Discrepancies of nearly 5.9% may not be tolerable if patient results are also routinely obtained from another method not subject to this problem. Furthermore, this method would not be expected to pass proficiency testing unless the systematic error were eliminated or a peer group established where the target mean compensated for the systematic error of the method. If systematic error were eliminated, then method performance would be considered to be good. ED performance would be classified as marginal,

owing to the smaller systematic error; reducing imprecision would still be desirable to provide a more cost-effective process for daily operation.

SODIUM

The CLIA PT criterion is 4 mmol/L, which is 2.857% at a target value of 140 mmol/L. For simplicity in preparing the MEDx chart, let's round TEa to 3.0%. At a concentration of 140 mmol/L, the replication experiment showed CVs of 0.49% for CP staff and 0.76% for ED staff. The CP comparison study for 103 patient samples gave a regression line of y=30.7 mmol/L + 0.772X (r^2=0.805), whereas the ED comparison study for 574 patient samples gave a regression line of y = 27.25 mmol/L + 0.805x (r^2=0.751). Because of the narrow range of data, as revealed by the low correlation coefficient, let's evaluate performance near the mean of the data (approximately 140 mmol/L) to minimize any problems with the reliability of the regression coefficients. At 140 mmol/L, the systematic error in the CP study is 1.22 mmol/L or 0.87%, whereas the ED study shows a systematic error of only 0.05 mmol/L, which is essentially 0%.

Figure 6 shows a MEDx chart with operating points for the CP (y=0.87%, x=0.49%) and ED studies (y=0.0%, x=0.76%). Performance is classified as excellent for CP and good for ED based on the precision and accuracy observed in the middle of the analytical range.

To gain practice in applying the MEDx chart, you may find it useful to work through the additional applications for potassium, chloride, and urea nitrogen that are included in the study.[11] Potassium appears to provide excellent performance. Chloride precision seems to be excellent, but the systematic error may be a limitation, depending on the decision level you choose. Urea nitrogen shows some limitations for both imprecision and inaccuracy.

The authors of the study concluded that the portable clinical analyzer showed good accuracy when compared with a standard clinical chemistry analyzer. This conclusion represents a judgment of the performance for the system as a whole since no analytical quality requirements were defined for the individual tests. If the CLIA PT criteria were used as analytical quality requirements, some tests would be assessed to have good to excellent performance and some as poor to marginal performance. This example shows why each laboratory needs to be able to make its own decisions on method acceptability, even when evaluation studies have been published in the scientific literature. When considering your own situation, you might decide that this analyzer provides acceptable performance for sodium and potassium, but not for some of the other tests.

DISCUSSION

The method evaluation studies for this portable clinical analyzer illustrate the importance of defining quality requirements for individual tests and making decisions about the performance observed relative to the quality required for those tests. Collecting the right data and calculating the right statistics aren't enough. The clinical laboratory scientist must also make correct decisions based on the calculated statistics. These decisions should consider whether the performance observed is satisfactory for the quality requirements of the individual tests.

The MEDx chart provides a simple graphical tool for comparing the precision and accuracy observed for a method with an analytical quality requirement that is stated in the form of an allowable total error. Analytical errors estimated from method evaluation experiments can be displayed as individual components by indicating their magnitude on the axes or as combined imprecision and inaccuracy by plotting an operating point (inaccuracy as the y-coordinate, imprecision as the x-coordinate). Estimates of total imprecision from the day-to-day replication experiment and bias from the comparison of methods experiment are recommended as the most realistic values for the operating point. Method performance can then be classified as poor, marginal, good, or excellent, based on the location of the operating point relative to the different total error criteria.

Final judgments on acceptability depend on each laboratory's situation and the particular application of the method. A method with poor performance should be considered unacceptable for routine operation in almost all cases; the exceptions may be if there is no alternative method available or if this test is the worst case in a group of good to excellent performing tests on a multi-test analyzer. A method having marginal performance may be acceptable if performed by a highly controlled instrument system that is extremely stable and seldom has problems, or if performed in a laboratory by well-trained and experienced technologists who diligently perform preventive maintenance, carefully monitor instrument function checks, perform periodic validation tests, and skillfully interpret statistical control charts. Methods with good performance should be generally acceptable, but may still require a high level of preventive maintenance and daily quality control to assure routine performance is acceptable. Methods with excellent performance should be more readily controllable by common statistical QC procedures with 2 to 4 control measurements per run and should provide the most cost-effective operation for routine service.

Concerning the issue of which total error criterion is most appropriate, the relationship between the different total error criteria and the operating specifications needed to adequately control a testing process demonstrates the validity of the most demanding recommendation (bias+4s). In fact, an even more demanding criterion, such as bias +5s, would be advantageous to assure that a method could be adequately controlled with the CLIA minimum requirement of 2 control

measurements per run. Earlier work[9] shows that the minimum systematic error that would be medically important can be related to the TE criterion in the following way: a bias+2s TE criterion corresponds to a critical systematic error (ΔSE_{crit}) of $0.35s_{meas}$, bias+3s to $1.35s_{meas}$, bias+4s to $2.35s_{meas}$, bias+5s to $3.35\ s_{meas}$, and bias+6s to $4.45\ s_{meas}$. Critical systematic errors of 3 and greater can generally be detected by using 2 control measurements per run, as documented by power function graphs for commonly used control rules.[9] Thus, a criterion of bias+5s would lead to even more cost-effective methods.

For practical purposes, the 2s, 3s, and 4s criteria are sufficient to classify methods today, but a 5s criterion might be added in the future to further distinguish the performance of different methods. The MEDx chart could be constructed to include 4 criteria, perhaps classifying a method that satisfies the bias+5s criterion as superior. Or, the current classifications could be applied to 3s, 4s, and 5s TE criteria. The advantage of the MEDx chart is that it can incorporate three or four criteria and provide suitable classifications for each. Multiple criteria allow the laboratorian to make the most appropropriate judgment for whatever situation arises.

Given the importance of QC in establishing appropriate method acceptance criteria, a logical extension and improvement of the MEDx chart is the OPSpecs chart[10], which can be used both to assess the adequacy of method performance and to select (or validate) control rules and numbers of control measurements. Use of an OPSpecs chart integrates the practices of method evaluation and QC validation with the same graphical tool and the same quality requirement. This approach should assure that the same level of quality will continue to be achieved in the routine operation of the testing process.

OPSpecs charts are more difficult to prepare because they require additional information about the performance of QC procedures, but they are readily available in a workbook compilation[12] and can easily be generated by electronic spreadsheets or by specialized software, such as the QC Validator program[13] [Westgard QC, Ogunquit ME]. An advantage of computer preparation is that clinical quality requirements in the form of medically significant changes in test results can also be used to derive operating specifications.[14,15] Thus, decisions on method performance could be directly linked to the intended clinical use of a test result to further improve the decision making process.

A practical strategy for improving decisions on method performance is to begin now to use the MEDx chart; you only need a sheet of graph paper to prepare the chart and apply the approach in your laboratory. The next step is to learn about OPSpecs charts and their additional capabilities for selecting and validating QC procedures. In the future, you should develop the capability to use both analytical and clinical OPSpecs charts for validating both method and QC performance.

REFERENCES

1. Barnett RN, Youden WJ. A revised scheme for the comparison of quantitative methods. Am J Clin Pathol 1970;54:454-62.

2. Westgard JO, Hunt MR. Use and interpretation of common statistical tests in method comparison studies. Clin Chem 1973;19:49-57.

3. Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method deelopment and evaluation. Clin Chem 1974;20:825-33.

4. Westgard JO, deVos D, Hunt MR, Quam EF, Carey RN, Garber CC. Concepts and practices in the evaluation of clinical chemistry methods. Part I. Background and approach. Am J Med Technol 1978;44:290-300. Part II. Experimental procedures. Am J Med Technol 1978;44:420-430. Part III. Statistics. Am J Med Technol 1978;44:552-70. Part IV. Decisions on acceptability. Am J Med Technol 1978;44:727-42. Part V. Applications. Am J Med Technol 1978;44:803-13.

5. Westgard JO. Precision and accuracy: Concepts and assessment by method evaluation testing. CRC Critical Reviews in Clinical Laboratory Sciences 1981;13:283-330.

6. Fraser CG, Hyltoft Petersen P, Ricos C, Haeckel R. Proposed quality specifications for the imprecision and inaccuracy of analytical systems for clinical chemistry. Eur J Clin Chem Clin Biochem 1992;30:311-7.

7. Stöckl D. European specifications for imprecision and inaccuracy compared with US CLIA proficiency-testing criteria. Clin Chem 1995;41:120-1.

8. Ehrmeyer SS, Laessig RH, Leinweber JJ, Oryall JE. 1990 Medicare/CLIA final rules for proficiency testing: Minimum interlaboratory performance characteristics (CV and Bias) needed to pass. Clin Chem 1990;36:1736-40.

9. Westgard JO, Burnett RW. Precision requirements for cost-effective operation of analytical processes. Clin Chem 1990;36:1629-30.

10. Westgard JO. Charts of operational process specifications ("OPSpecs charts") for assessing the precision, accuracy, and quality control needed to satisfy proficiency testing criteria. Clin Chem 1992;38:1226-33.

11. Woo J, McCabe JB, Chauncey D, Schug T, Henry JB. The evaluation of a portable clinical analyzer in the emergency department. Am J Clin Pathol 1993;100:599-605.

12. Westgard JO. OPSpecs Manual: Operating specifications for imprecision, inaccuracy, and quality control. Ogunquit ME, Westgard QC, 1995.

13. Westgard JO. A program for assessing statistical control procedures. Med Lab Observ 1994;26(2): 55-60.

14. Westgard JO, Hytoft Petersen P, Wiebe DA. Laboratory process specifications for assuring quality in the U.S. National Cholesterol Education Program (NCEP). Clin Chem 1991:37:656-61.

15. Westgard JO, Seehafer JJ, Barry PL. Allowable imprecision for laboratory tests based on clinical and analytical test outcome criteria. Clin Chem 1994;40:1909-14.