The Establishment and Dissemination of Quality Goals

Callum G. Fraser, Ph.D.*
Biochemical Medicine
Ninewells Hospital and Medical School
Dundee, Scotland
and
Per Hyltoft Petersen
Clinical Chemistry
Odense University Hospital
Odense, Denmark

***Presenting Author**

Abstract: Many strategies have been proposed for the setting of quality goals in laboratory medicine. Traditional strategies are based on the use of reference intervals, opinions of clinicians, the state of the art, opinions of experts, assessment of the effect of error on clinical characteristics of tests, and biological variation. All have advantages and disadvantages but the use of data on biological variation appears to have many telling merits. The database is large and within-subject biological variation appears generally constant. Goals based on biology are available for bias as well as imprecision and the same theories can be used to generate goals for drug assays. A variety of proposals has recently been advanced for general models involving analyzing the effects of errors on diagnostic efficiency and sensitivity and a model for the allowable difference between two methods; it is of much interest that all of these generate quality goals based upon fractions of biological variation. Although in the U.S., it appears as if the advent of CLIA'88 has caused many concerns, in Europe there has been much recent rediscussion on goal setting. Approaches based on biology are much favored and, in essence, recommendations are that: desirable imprecision is less than or equal to one-half of the average within-subject biological variation, desirable bias is less then or equal to one-quarter of the group (within- plus between-subject) biological variation, and desirable difference between methods (or drift during monitoring) is less than one-third of the average within-subject biological variation. It is important to recognize that imprecision and bias must be considered together and that, when imprecision approaches the goal, bias must be small, and vice versa. Disseminating goals must be more actively pursued by all involved in laboratory medicine including journal editors and referees, industry, and organizers of PT and EQA schemes.

Introduction

Quality management processes, including the essential components of control, assessment, assurance, planning, improvement, and audit, have pervaded manufacturing and service industries in both private and public sectors, including all aspects of health care. However, to use such quality tools correctly, it is necessary first to define the standards of quality required. Perusal of the literature might suggest that problems are few because there are many

papers, reviews, and book chapters dealing with the generation and application of quality goals in laboratory medicine [1]. It does appear, however, that many still have difficulties in deciding the standards which should be attained by laboratories, ideally for all performance characteristics, but particularly for the important reliability characteristics of precision and bias (accuracy). There are plausible reasons for this including the facts that:

- tests are used in many clinical situations and it might be that there is no single set of goals which would make a method suitable for all purposes,
- there are many recommendations in the literature and it might be difficult to choose the most appropriate,
- new recommendations continue to appear, suggesting that there is no professional consensus on the topic,
- there appears to be no evidence that patients have been harmed by current performance,
- in countries which have legislation involving proficiency testing, the focus might be simply on achieving the standards required to pass, and
- industry does not appear to use professionally set goals as major considerations either in development or marketing.

In view of the apparent lack of ubiquitous use of numerical quality goals, these will be briefly critically reviewed here and then the widely accepted current recommendations documented.

Traditional strategies for setting quality goals

Traditional approaches for setting goals for precision (often used for total error, however) are based on use of (I) reference values, (ii) the opinions of clinicians, (iii) the state of the art, (iv) views of expert individuals and groups, (v) analysis of the effect of error on the clinical utility of tests, and (vi) biological variation. All have advantages and disadvantages which have been discussed in detail previously and are only summarized here .[2] Reference intervals are available for most quantities and the strategy is simple, but the fractions of the reference interval chosen to set goals are empirical, and reference intervals depend on the precision and bias of the analytical procedure, the population studied, and the statistical technique used for data reduction. The opinions of clinicians have been mainly obtained by questionnaire involving clinical vignettes but, *inter alia*, the difference between two results (or a result and a reference limit) are not due only to random analytical error as generally supposed but also to within-subject variation and pre-analytical variation, the use of the median result satisfies only half of the respondents, and the probability with which decisions are made is not always $P < 0.05$. The state of the art, even of a selected group of better laboratories, changes with time, laboratories may adopt special techniques in the analysis of samples circulated in proficiency testing (PT) or external quality assessment (EQA) schemes from which the state of the art is usually derived, and the matrix of the samples may not be the same as samples from patients. The views of expert individuals and groups, although interesting,

are often subjective and contradictory. Analyzing the effects of increasing errors on nosological characteristics such as sensitivity and specificity seems appropriate when a test is used in a single well-defined clinical situation. Problems arise, however, in generating clinical guidelines for the use of such test results in that there are often many guidelines for the use of a single test, these guidelines may become outdated or be corrupted in local practice, and they may create an unresponsive attitude to new developments or inhibit new thinking. Goals based upon biology are favored by many and are based upon the postulate of Cotlove et al [3] who suggested that: analytical error < 0.5 biological variation. This concept was expanded at the 1976 Aspen Conference of the College of American Pathologists [4] and it was suggested that, for diagnosis and monitoring - $CV_{analytical} < 0.5\ CV_{within\text{-}subject}$, and for screening - $CV_{analytical} < 0.5\ (CV_{within\text{-}subject}^2 + CV_{between\text{-}subject}^2)^{½}$; the formula in parentheses will be denoted henceforth simply as CV_{group}. This proposal was then accepted by the Sub-Committee on Analytical Goals in Clinical Chemistry of the World Association of Societies of Pathology in 1978.[5] Thus, more than two decades ago, the consensus was that quality goals were best based on biological variation.

The advantages of the approach based on biology

Goals based upon biological variation seem to have become generally accepted, but only slowly. The reasons for this might include the facts that, at least originally, (I) the database encompassed only a few quantities and the experimental work had been done on young healthy subjects, (ii) some of the calculated goals appeared too strict to be achieved with available technology and some appeared too loose, (iii) goals were not available for bias, (iv) goals were not available for exogenous quantities such as drugs, and (v) these goals were based upon statistical considerations and not on the clinical use of tests. These supposed demerits have been negated with the passing of time. Now, good data are available on the biological variation of many quantities, and the estimates seem generally constant and therefore ubiquitously applicable.[6] When goals appear too strict, these should be viewed as worthy targets, other objective goals used as interim measures, and strategies to provide quality laboratory practice, appropriate internal quality control, quality improvement, or investigation of alternative methodology instituted; when goals seem too loose, quality planning and appropriate quality control will save resources. Goals for bias based on biology have been proposed;[7] they showed that, to allow the use of common reference values, bias (as % deviation) < $0.25\ CV_{group}$, if the precision was negligible. Goals for drugs can be calculated [8] using a similar model based on pharmacokinetic theory as:

$$CV_{analytical} < 0.25\ \{[2^{T/t} - 1]/\ [2^{T/t} + 1]\}_{*}100$$

where T is the dosing interval and t the half-life.

Goals based upon recent models

In spite of the work done to refute the alleged criticisms of setting goals using data on biological variation, further general models have been proposed. Harris [9] expanded his earlier work to include bias and suggested that $CV_{analytical} < R(1/80 - 4/5\ bias^2/R2)]$ where R is the reference interval; he suggested as a rule of thumb, however,

that, in the absence of bias, $CV_{analytical} < 0.25$ $CV_{within\text{-}subject}$ for monitoring and $CV_{analytical} <$ 0.1R for diagnosis. Ross[10] considered the effect of errors on loss of diagnostic efficiency as a means of setting goals; it was suggested that, for individual testing, $CV_{analytical} < 0.64$ $CV_{within\text{-}subject}$ and, in further work,[11] it was proposed that $CV_{analytical} < 0.5$ $CV_{within\text{-}subject}$ and that bias < 0.25 to 0.33 $CV_{within\text{-}subject}$. Klee[12] proposed that an error budget, the squared sums of the imprecision and bias, be set to allow less than a 50% increase in the false-positive rate for classification of healthy subjects; the budget was allocated as allowable precision and bias of < 0.18 CV_{group} and < 0.36 CV_{group} respectively. It has also been suggested[13] that the allowable difference between methods used in the same laboratory for a single quantity can be calculated as < 0.33 $CV_{within\text{-}subject}$.

It is interesting that all these models, although none has yet been widely used, propose that analytical goals be based on fractions of biological variation.

Current consensus views

In the U.S., there appears to be much concern with the problems created by introduction of CLIA '88. It is considered that it would be a retrograde move if the standards laid down in this legislation became the analytical goals deemed suitable for use in laboratory medicine. In contrast, in Europe there has been a great interest in the harmonization of all kinds of activities, including the practice of laboratory medicine, and several groups have considered setting quality goals. A working group of the European Group for the Evaluation of Reagents and Analytical Systems in Clinical Chemistry[14] used the concepts based on biology and the need for interim goals for quantities for which these are currently unattainable to present numerical goals for commonly assayed quantities. After due consideration of all available models, it was recommended that: precision should be < 0.5 $CV_{within\text{-}subject}$ or less than the precision attained by the best 0.20 fractile of laboratories, whichever was the less stringent - the latter could be used when data on biological variation were unavailable; bias should be < 0.25 CV_{group} or < R/16 when data on biological variation were unavailable or < $CV_{within\text{-}subject}$ when these goals appeared unattainable with present technology. A further working group organized under the auspices of the European External Quality Assessment Organizers' Group has recently published [1] their views which show the renewed trend towards using biology; it was proposed that, for monitoring patients, $Sd_{analytical} < 0.5$ $SD_{within\text{-}subject}$ when changes in bias were negligible and bias < 0.33 $SD_{within\text{-}subject}$ when precision was negligible, and, for diagnostic testing, $Sd_{analytical} < 0.58$ SD_{group} when bias was negligible, and bias < 0.25 SD_{group} when precision was negligible. It is important to recognize that this working group considered that precision and bias ought to always be considered together and that, when large precision was present, only a small bias was acceptable, and *vice versa*.

Dissemination of quality goals

Although much work has been done on the generation and application of goals, which has been mainly disseminated through the publication of papers, letters, reviews, book chapters, and conference proceedings, there is no doubt that the correct application of goals could be encouraged. The authors of manuscripts concerned with evaluation of new methods, reagent kits, or analytical systems that do not use objective quality

goals as criteria of acceptability ought to be encouraged to apply these by journal editors and referees; moreover, journals could incorporate the use of quality goals as requirements in their instructions to authors. Industry could not only use objective goals to assist in the identifying the quantities for which improvements in methodology are badly needed and to use in method development, but they could assist very much in the disseminating information on quality goals by documenting these in their labeling just as the performance characteristics and reference values are currently included. Organizers of PT and EQA schemes could have a vital role in encouraging the use of objective quality goals through using these as the fixed limits for assessment of laboratory performance and through highlighting acceptable and unacceptable methodology using these criteria.

Concluding remarks

Setting quality goals has been the subject of much discussion for over three decades. There is no doubt that the current consensus is that goals for precision and bias are best based upon biological variation data and that these should be more widely used in many aspects of laboratory medicine. This is not to say that there are no unanswered questions and much work is still required on many topics including goals for tests done close to the patient, for tests done frequently, for other performance characteristics, and for qualitative and semi-quantitative tests. It is to be hoped that these challenges will be actively pursued by professionals in laboratory medicine rather than imposed by legislators.

References

1. Stockl D, Baadenhuisjen H, Fraser CG, Libeer J-C, Hyltoft Petersen P, Ricos C. Desirable routine analytical goals for quantities assayed in serum. *Eur J Clin Chem Clin Biochem.* 1995;33:157-169.

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

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

4. Elevitch FR, ed. Analytical goals in clinical chemistry. Proceedings of the CAP Aspen Conference Aug 25 - 27 1976. Skokie; IL. CAP, 1977.

5. Proceedings of the Sub-Committee on Analytical Goals in Clinical Chemistry, WASP, CIBA Foundation; 1978. Analytical goals in clinical chemistry: their relationship to medical care. *Am J Clin Pathol.* 1979;71:624-30.

6. Fraser CG. The application of theoretical goals based upon biological variation in proficiency testing. *Arch Pathol Lab Med*. 1988;112:404-15.

7. Gowans EMS, Hyltoft Petersen P,

Blaabjerg O, Horder M. Analytical goals for the acceptance of common reference intervals for laboratories throughout a geographical area. *Scand J Clin Lab Invest.* 1988;48:757-64.

8. Fraser CG. Desirable performance standards for therapeutic drug monitoring. *Clin Chem*. 1987;33:387-9.

9. Harris EK. Proposed goals for analytic precision and accuracy in single point testing. *Arch Pathol Lab Med.* 1988;112:416-20.

10. Ross JW. A theoretical basis for clinically relevant proficiency testing evaluation limits: sensitivity analysis of the effect of inherent test result variability on acceptance method error. *Arch Pathol Lab Med*. 1988;112:421-34.

11. Ross JW, Fraser MD. Analytical goals developed from inherent error of medical tests. *Clin Chem*. 1993;39:1481-94.

12. Klee GG. Tolerance limits for short-term analytical bias and analytical imprecision derived from clinical assay sensitivity. *Clin Chem.* 1993;39:1514-8.

13. Hyltoft Petersen P, Fraser CG, Westgard JO, Lytken Larsen M. Analytical goal-setting for monitoring patients when two analytical methods are used. *Clin Chem.* 1992;38:2256-60.

14. Fraser CG, Hyltoft Petersen P, Ricos C, Haeckel R. Proposed quality specifications for the imprecision and inaccuracy of analytical systems in clinical chemistry. *Eur J Clin Chem Clin Biochem.* 1992;2-30:311-7.

Analytical Goals in New and Non-Traditional Clinical Laboratory Techniques

Charles R. Handorf, M.D., Ph.D.
Department of Pathology
Methodist Hospitals of Memphis
Memphis, Tennessee

Abstract: There can be absolutely no debate that, in the current economic and social climate, the demands on the medical care delivery system are changing. Yet, in the rush to develop faster and cheaper systems for care delivery, every effort must be made to maintain and--where possible--to improve the quality of care given to patients. The clinical laboratory serves as a very effective model for evaluating the changes and challenges associated with keeping the patient first among priorities which may sometimes be confusing and even conflicting.

Definitions of analytical quality vary somewhat depending upon the observer's point of view; it is important to affirm that none of these views is incorrect. Quality as defined by the traditional clinical laboratorian, some other analyst, the clinical care giver, or the patient will impact the appropriateness of analytical goals. Each of these definitions of quality may be correct, but each is also likely to be incomplete.

Alternative site/point of care laboratory testing strategies have been introduced to the testing armamentarium amid great confusion about how to define and assure quality. Little thought was given to analytical goals and how to achieve them before most alternative site testing strategies were introduced. A close analysis of the history of analytical goal setting in blood glucose monitoring gives insight into the ways in which analytical goal setting, monitoring, and assurance should be approached as new and diverse approaches such as nanotechnology and molecular pathology are introduced into common use.

Introduction

The delivery of medical care in the United States is changing; whether we like it or not, a variety of economic and sociopolitical forces are forcing us to reevaluate and, in some cases, to radically reengineer the manner in which we deliver care. Technological advancements, improved computer-based information management, and a consolidated systems approach should allow the clinical laboratory to survive and even to thrive under whatever delivery system evolves. The major challenge facing all players in the laboratory industry, however, will continue to be this: Putting all personal, professional, and parochial motives aside, how can we develop approaches which deliver the best possible care to our patients? Major concerns still exist regarding the part to be played by alternative site/point of care testing strategies such as bedside glucose monitoring, coagulation testing, and blood gas and electrolyte evaluation. Rapidly evolving techniques such as molecular pathology and in vivo monitoring represent special challenges in goal setting.

Developing appropriate analytical goals is heavily dependent upon the prior development, knowledge, and understanding of applicable clinical goals. Analytical goals must not be disjoined from the real world

Central Laboratory Testing	Point-of-Care Laboratory Testing
Many tests; few sites; few instruments	Few tests; many sites; many instruments
Large runs; “factory” environment	Small runs; “boutique” environment
Few highly trained analysts	Many inexperienced analysts
Analysts with restricted tasks in the testing cycle	Analysts with more general tasks in the testing cycle
Longer turnaround time	Shorter turnaround time
Controlled physical environment for reagents and instruments	Less controlled physical environment for reagents and instruments
Error types: “shifts and trends”	Error types: “sporadic”

Table 1. Comparison of analytical systems in central laboratory testing vs point-of-care laboratory testing.

application of the clinical test, lest we fall into the trap of spending unacceptable amounts of our resources searching for an analytical equivalent of the holy grail--an analytically pure and absolutely "correct" answer. The 1976 College of American Pathologists Aspen Conference on Analytical Goals in Clinical Chemistry developed a primary recommendation that analytical goals can only be defined in terms of needs for patient care,[1] a goal that is all too easy to lose sight of. Medical care is probably most effective when data are derived in approximately the following proportions: 70% from the clinical history, 20% from the physical examination, and 10% from laboratory tests.[2] The part played by the laboratory in medical decision making is important, but not preeminent. On the other hand, clinicians must also understand and be willing to adhere to appropriate clinical goals; uncritical reliance upon new technology and unnecessary focus upon speed of the results as a primary virtue must be vigorously challenged. It has been suggested that under any circumstances, faster results are preferable to slower results. This defies logic, especially if the faster results are unacceptably expensive or are incorrect.

Many of us trained in laboratories where a sign on the wall stated: "Speed, quality, and low cost--you can have any two." A major challenge in the coming environment is to change this approach, so that instead of sacrificing one or two of these desirable attributes of laboratory tests, we find systems which optimize all three. If our clinical goal may be summarized as "take good care of patients," then our analytical goal may be synthesized as follows: To get the best quality answer possible in a clinically appropriate time frame at the lowest cost attainable. Focusing on analytical quality alone as the only domain of analytical goals will no longer be acceptable.

Blood Glucose Testing

Possibly because blood glucose is a common analyte, much has been written about medical relevance and analytical goals for glucose. Additionally, glucose has presented a special challenge since it is the most frequent analyte evaluated at the bedside, through the use of glucose reflectance meters. Thus, glucose may serve as a test case for analyzing the effectiveness of implementing of alternative site testing strategies. From this case we may learn much.

Table 1 offers a contrast between the analytical systems inherent in central laboratory testing versus testing at the bedside. Although this table was developed with blood glucose testing in mind, it generalizes rather well to the range of laboratory testing done at the point of care. Each type of testing has strengths and weaknesses. Although analytical goals for in-hospital glucose testing should be the same regardless of where or how the test is done, this table helps to emphasize that the quality assurance systems appropriate to assure analytical success may well differ depending on the type of testing system. Application of the concepts of process control which are quite appropriate in the central laboratory may not be quite so appropriate for testing at the bedside. Another point to be made from evaluating these contrasting elements is the important, though different, role of the analyst in each setting. There is a need in the central laboratory to broaden the technologist's perspective so that there is more focus on the patient and less focus on the test as an end in itself; on the other hand, there is a great need to integrate the bedside analyst more into the quality evaluation of the testing process. Neither testing at the point of care nor testing in central laboratories is implicitly good or bad. Each type of testing has inherent strengths and weaknesses; each may hold specifically different requirements to assure that analytical goals are achieved.

If analytical goals are dependent upon clinical goals for relevance and if clinical goals should be more specific than just "to take good care of patients," it is useful to dwell for a moment upon the question of how clinical goals are set. In many institutions, blood glucose monitoring in the central laboratory has not significantly decreased when bedside testing has been introduced. Rather, the entrance of bedside testing has been more or less additive to the total amount of testing done. One suspects that, although clinical goals including such approaches as critical pathways and protocols of care should incorporate issues such as frequency and type of glucose testing necessary for adequate patient monitoring, they do not. This suspicion is verified by anecdotal evidence indicating that in some hospitals, blood glucose monitoring is routinely performed twice daily, in others it is performed four times daily, and in others it may be performed as often as once an hour. Critical pathways should never be so rigid as to disallow clinical judgment needed to care for patients, but one suspects that in this and many other instances we have not even begun to establish guidelines for what is necessary for good care. It is little wonder, then, that our efforts to establish appropriate analytical goals are splintered and somewhat feeble. The first step in establishing analytical goals should be articulating clinical goals; as we integrate new testing strategies into the care of patients, it is vitally important that clinical goals be established and understood. Until now we have not done very well in this regard.

Fraser has done much of the work to help us understand how analytical goals should be set.[3] Basically, goals are set biologically, experimentally, or experientially and generally have relied upon statistical evaluation of the coefficient of variation of multiple observations as a reflection of imprecision or random error. Biological goals have generally grown from the strategy of Cotlove, which holds that the allowable coefficient of variation for an analyte should be less than one half of the observed biological variation.[4] Based upon this approach, the analytical goal for blood or serum glucose may be demonstrated to be in the range of 2 to 3 percent. Experimental approaches to analytical goal setting include evaluating reference values, the state of the art, and analyzing the effect of errors on clinical decision making; experiential goal setting may reflect the opinions of clinicians or of expert committees or, in the case of glucose evaluation, may even reflect analytical goals derived from the opinions of patients themselves.[5] Each of these approaches, when applied to the glucose issue, gives a somewhat different view. Cumulatively, desirable coefficient of variation may range from 2 to 15 percent depending upon which approach is taken. It is clear that we really do not know what our analytical goal should be. This no doubt reflects our confusion about clinical goals, and is a state of affairs which is not likely to improve until more cogent clinical goals are established.

We have a number of tools at our disposal which allow us to monitor our achievement of goals; as new technologies develop and are implemented into the care of patients, it will be critically important that we use all of the quality systems which are at our disposal, since each of them will tell us something different about the sort of job we are doing. There has been a tendency to rely heavily upon data from external proficiency testing programs in discussing analytical goals; this is not inappropriate but must be supplemented with evaluation of other data such as integrated hospital quality assurance data, quality control data, review of external and internal inspections, and review of data obtained from operator training, evaluation, and certification. Many data and resources are available to help us understand analytical goals of new testing approaches, but we must begin to understand how better to gather and apply these data. Resources include but are not limited to College of American Pathologists Surveys, special studies such as CAP Q-Probes and data deriving from the Laboratory Management Improvement Program (LMIP), and information from the CAP Laboratory Accreditation Program. The use of all of these information sources will help us to avoid taking too narrow a view of analytical goals and how they should be evaluated; new technologies will require new approaches to goal setting and evaluation.

An example of the usefulness of these information sources is the report of Jones et. al. upon the report of the 1991 Q-Probe on bedside glucose monitoring; precision measurements based on 15,950 observations in 569 institutions were evaluated.[6] The authors concluded that programs demonstrated better performance if laboratorians were involved, appropriate operator training was instituted, if an internal quality comparison program was in place, and if an external proficiency testing program was used. On a global basis, this sort of observation is invaluable in helping us to assure quality performance.

Another tool is the CAP Laboratory

Deficiency	% of Surveyed Laboratories Deficient
In the absence of on-site supervisors, are the results of tests performed by personnel reviewed by the laboratory director, POCT section director, general supervisor, or the person in charge of the POCT section on the next routine working shift?	16.50
When applicable, are all patient results reported with accompanying reference (normal) ranges?	14.04
Are all reagents properly labeled with the following elements, as applicable and appropriate? 1. Content and strength, concentration or titer, 2. Storage requirements, 3. Date prepared or received, 4. Date placed in service, 5. Expiration date	13.75
Is linearity of the instruments/reagent system verified initially and at least semi-annually, or when calibration fails to meet the laboratory's acceptable limits?	13.05
Is there a documented system in operation to detect clerical errors, significant analytical errors, and unusual laboratory results?	9.36
Is quality control evaluated daily?	8.37
Is there documented evidence that quality control checks are performed on all tests each day of use with suitable positive and where appropriate, negative reference samples?	7.65
Is the laboratory enrolled in available CAP Surveys (Interlaboratory Comparison) program for the patient testing performed?	6.96
Is there evidence of corrective action when control results exceed defined tolerance limits?	6.90

Table 2. Inspection data from CAP Laboratory Accreditation Program, January through August, 1995. N = 589.

Accreditation Program; participation in laboratory accreditation activities as an inspector and as an inspected laboratory can be a very valuable experience, but what may we learn from a global evaluation of the data derived from the program, especially as regards point of care testing? Reviewing data from 589 laboratories inspected from January through August of 1995 shows that the most commonly cited deficiencies involve inadequate result review, failure to post normal ranges, failure to label reagents, and failure to check linearity (Table 2). Review of such data should help direct our efforts

for quality enhancement programs in testing at the bedside.

Hospitals may carry out their own programs relative to the monitoring and attainment of analytical goals, and indeed should do so as new and emerging technologies are encountered. Laposata conducted a series of elegant studies to show the usefulness of regular inspections of bedside glucose testing sites to maintaining quality results.[7] He has demonstrated that fast- paced clinical settings do more poorly overall with quality maintenance and that the most common quality assurance violation is failing to properly perform proficiency testing.

At Methodist Hospitals of Memphis, review of frequency of bedside blood glucose testing by analysts has shown wide disparity--some analysts perform only one or two tests per month, whereas others perform many hundreds. Such data are useful to help focus upon specific analysts who, through infrequency of testing, may require more intensive retraining or proficiency testing monitoring. In fact, as we attempt to "get the right answer" in an environment of improving technology and increasing numbers of less sophisticated analysts, reaching analytical goals may require that more emphasis be placed on monitoring the analyst than on monitoring the instruments and reagents. We may, for instance, develop proficiency testing paradigms in which relatively infrequent analysts are required to perform proficiency testing on a more frequent basis than are analysts who perform quality control and actual patient assays on a more frequent basis.

Summary and Conclusions

In summary, for adequate analytical goals to be set and monitored for new and emerging technologies, it will be critically necessary that we establish on the front end the need for such testing and that we are certain that new technology is really required to perform the task at hand. A current example is the observation that well-managed pneumatic tube systems to the central laboratory can replace point of care technology at a fraction of the cost and can maintain superior quality and turnaround times.[8] Another emerging theme in the diffusion of laboratory testing away from the central laboratory is the need to maintain excellent training and monitoring of analysts and, wherever possible, to keep the number of analysts to a minimum. Furthermore, new approaches to management of quality will be essential. We must devise practical quality control, proficiency testing, and training programs that are demonstrated to serve established analytical goals and, at the same time, to be cost-efficient.

Two emerging areas which are impacted by these issues are molecular pathology and in vivo patient testing. Each of these technologies represents high-cost testing areas for which we still must establish clinical and analytical goals. Through well conducted trials, we must decide how much quality management is sufficient to serve these goals and not attempt to apply old quality management paradigms to these technologies. Molecular pathology basically represents the challenge of how to benchmark non-numeric data, a challenge which has not yet been answered. In many ways, the failure of analytical goal setting in molecular pathology parallels a much older problem---that of analytical goal setting in microbiology. The challenge remains to broaden the concepts of goal setting beyond the establishment of numerical biological norms; we may need to strive much harder in

areas of epidemiological study to help understand what our goals of overall laboratory performance must be so that we serve the needs of whole populations of patients without spending resources needlessly. More and more, the analytical question which we answer may become "do I need to do this test at all?" rather than "how good an answer do I need to produce?"

An even more difficult emerging problem may be in the area of in vivo, continuous sensing technology. So-called nanotechnology is rapidly evolving, and very soon we will progress into an arena in which a considerable number of analytes may be evaluated on a patient in a real time, continuous monitoring mode. At this time, there still is some advantage since most of these technologies are still in a developmental or testing stage; there still may be time for thoughtful dialogue and a scientific approach to developing broad based analytical goals for these technologies. Laboratorians will need to resist the temptation to bind themselves to old approaches to goal setting, since in vivo technologies do not, at this time, conform very well to old laboratory paradigms of internal or external quality control by introducing a pseudo-sample. Indeed, for most in vivo technologies, the concept of an analyst is largely irrelevant. For this and other reasons, developers of in vivo technologies may be tempted to say that such technologies are not laboratory testing at all, but are merely "monitoring"; it therefore may be implied that such activities should not come under the scientific or regulatory purview of laboratory specialists. Such semantic dodging must be vigorously opposed, so that the patients of the future may have access in a systematic way to the most appropriate level of laboratory care. If we have learned anything from the experience of the implementing point of care laboratory testing in hospitals, it should be that when a disorganized, non-systematic approach is taken to the introduction of a new technology, it can take many years of effort to untangle the mess.

References

1. Elevitch FR, ed. Proceedings of the 1976 Aspen Conference on Analytical Goals in Clinical Chemistry. Skokie, IL, College of American Pathologists, 1977.

2. Cutler P. Problem Solving in Clinical Medicine from Data to Diagnosis. Baltimore, Williams and Wilkins, 1985, p 13.

3. Fraser CG, Hyltoft Petersen P. Desirable performance standards for imprecision and bias in alternate sites: the views of laboratory professionals. *Arch Pathol Lab Med*. 1995;119:909-913.

4. Cotlove E, Harris EK, Williams, GZ. Biological and analytic components of variation in long-term studies of serum constituents in normal subjects, III: physiological and medical implications. *Clin Chem*. 1970;16:1028-1032.

5. Weiss SL, Cembrowski GS, Mazze RS. Patient and physician analytical goals for self-monitoring blood glucose instruments. *Am J Clin Path*. 1994;102:611-615.

6. Jones B, Bachner P. Q probe 91-09A: bedside glucose

monitoring--data analysis and critique. Northfield, IL, College of American Pathologists, 1992.

7. Laposata M, Lewandrowski KB. Near patient blood glucose monitoring. *Arch Pathol Lab Med*. 1995;119:926-928.

8. Green M. Successful alternatives to alternate site testing: use of a pneumatic tube system with a central laboratory. *Arch Pathol Lab Med*. 1995;119:943-947.

Use of Analytical Goals by Health Care Manufacturers

Jan S. Krouwer, Ph.D.
Evaluations & Reliability
Ciba Corning
Medfield, Massachusetts

Abstract: Analytical goals are the *quantitative* requirements that the product must meet. Goals must be specific and have a clear success/failure criterion. Ideally, goals should include a protocol and data analysis method. There are three categories of analytical goals. First, clinical acceptability is the total analytical error and total analytical error sources. Second, there are a variety of regulatory goals depending on the approval required (510k, PMA, PLA). Finally, there is a list of marketing goals spawned by the competitive nature of business.

Setting goals involves: defining a metric, setting its target, and specifying a protocol and data analysis method. Whereas many of the metrics are defined, the target setting process is still difficult for manufacturers. Laboratorians may know what they need, but effective communication of these needs to manufacturers is lacking.

To set targets, manufacturers perform surveys (open ended questions, multiple choice questions, focus groups, and conjoint analysis). They also use performance data for similar, released assays (CAP data, internal data, published evaluation data).

Reviewing existing goals reveals inadequacies such as non existent goals, non quantitative goals, goals without a meaningful success/failure criterion, and unsupported goals. The goal setting process can be improved by deciding on and gaining experience with a metric, preparing cause and effect diagrams, and challenging existing goals.

Goals and claims are different. Manufacturers have internal goals. Upon product release, these goals are transformed into "claims", which may be different from the internal goals. Different manufacturers state claims differently, leading to confusion. Claims are: 1) The "typical data" claim - Half of the customers are expected to observe better performance, and the other half, worse performance. 2) The "warranty" claim - Here, all customers are guaranteed performance better than the limit. What is needed is a common vocabulary for the laboratorian and manufacturer.

Introduction

Manufacturers of diagnostic assays have a key milestone in the product development process: product release. The analytical performance goals they set are often quite different before and after product release. Before release, goals are often called *(internal) specifications* and, after release, *product claims*. Laboratorians never see the internal specifications. Sometimes there is confusion as to which goals are under consideration. This paper focuses on goals before product release. They can be divided into three conceptual categories: clinical acceptability, regulatory, and competitive goals.

Lab results have error. *Clinical acceptability* goals define how bad the error can be before it causes diagnostic problems. Laboratory assays are regulated. For

manufacturers, this means that assays must be FDA approved. *Regulatory* goals depend on the approval required (510k, PMA, PLA). For a laboratory, regulatory agencies require acceptable performance on proficiency surveys. Hence, assays must achieve a certain performance level with proficiency survey controls. Finally, companies must be *competitive* to remain in business. This spawns a list of marketing analytical performance goals.

Definition of an Analytical Goal

Analytical goals are a subset of the quantitative requirements that an assay must meet. The terms specifications, target values, and requirements are synonyms of analytical goals. Goals must be specific and have a clear success/failure criterion (e.g., there must be a metric). Ideally, goals should specify a protocol and data analysis method. This assures that not only the right type and amount of information will be collected but also describes how the data will be analyzed and reported. An analytical goal example is: the total precision CV should be less than 10% throughout the 50-500 mg/dL range as determined by the NCCLS protocol EP5.

A protocol and analysis method is recommended as part of a goal because the analytical performance of an assay differs from directly measurable assay properties such as the size and weight of an instrument. Analytical performance cannot be exactly determined - the true performance values (the "true state of nature") can only be estimated by experiments. Variation in the experimental results prevents their direct determination. The resulting data from these experiments has information in it that a properly designed analysis procedure will extract. Without a correct analysis and reporting procedure, interpretation of the data will be difficult if not impossible.

Without protocols and analysis methods it is unclear how to determine if a goal is being met. For example, for a glucose assay with a range of 5-1000 mg/dL: do we need to evaluate precision at 5 concentration levels, every 20 mg/dL, every mg/dL? Can we spike and dilute samples? If we dilute, what should be the diluent? To test interferences, what glucose level(s) should be used? If bias is evaluated, what is the criterion for meeting the goal, the point estimate? Its 95% confidence interval, its 99% confidence interval, every data point within the goal? NCCLS evaluation protocols help address some but not all of these issues. Experience shows that agreement for these issues helps to prevent questions after the data have been collected.

Constructing goals

Constructing goals involves:

- defining a metric (e.g., % CV precision)
- setting the metric's target (e.g., < 5% CV precision)
- defining a protocol to evaluate the metric (e.g., 2 observations per day for 20 days)
- defining an analysis and reporting procedure for the metric (e.g., ANOVA)

Consider that most assays are developed by manufacturers for sale to clinical laboratories. Laboratories run the assays and provide clinicians with results. Clinicians use the results to help answer the question, "Should I treat or not treat the patient?"

To a clinician, **total analytical error** is

the only parameter of importance. Total analytical error is defined as a percentage, (often 95% or 99%) of the distribution of differences in concentration between the test and reference method. Wrong results that cause misdiagnosis are just as harmful whether they are caused by random or systematic error.

Besides total analytical error, laboratories need to know **total analytical error sources** because these sources contribute to total analytical error and some error sources are specified by regulatory agencies. The manufacturer needs to know both total analytical error and total analytical error sources because he must satisfy all clinical and laboratory needs. Knowledge of error sources leads to improved assay performance which helps meet competitive goals.

Hence, there are two analytical goals for manufacturers:

1. total analytical error, used to validate the clinical use of an assay
2. total analytical error sources, used to improve assay quality

Setting Analytical Goal Targets

The reason that it is hard for manufacturers to set targets for goals is:

- manufacturers don't know how to ask for targets
- laboratorians don't know how to talk about targets

Surveys

Surveys would seem to be an easy way to set clinical acceptability targets. One simply asks clinicians. However, there are pitfalls. If one asks clinicians *open-ended* questions, such as "what are the clinically acceptable limits for a cholesterol assay?", one could get responses such as "no error" or "error that doesn't cause diagnostic problems." One remedy to this is to offer *multiple choice* questions. However, this is not foolproof either. Responses can be checked off without a guarantee that the question was correctly understood by the respondent. Moreover, respondents tend to want "the best." Thus, given a choice for total error for cholesterol to be 1%, 3%, or 5%, many respondents will simply check off 1%. Skenzel overcame many of these difficulties in a cleverly constructed survey.[1]

The problem is that in real life, one must make tradeoffs. One wants a car that is both luxurious and low priced. For laboratory assays, one wants low total error, low cost, high ease of use, high reliability, etc. For any of these situations, there will be acceptable compromises among the desired values of the goals. Conjoint analysis is a form of marketing research that provides a protocol and analysis method for estimating these tradeoffs.[2] Its idea is to present a clinician with a series of assays, each with different values for various attributes. The clinician ranks his preference for each of the assays. With several clinicians performing this ranking, the value of each attribute can be found by statistical analysis.

Theory

Studies have been made to set analytical performance goals by relating biological variation and analytical error to diagnostic decision making. Manufacturers keep track of these studies and try to ascertain to what extent the results are used by laboratorians.

Use of Current Performance Data

Given an assay that is in service, one can ask if the complaint rate is sufficiently low. A yes answer may signify that the assay's analytical performance is adequate. One can then measure this performance and use it as a

goal for a new version of the same or similar assay. Assay performance data sources can include: CAP or other proficiency survey data, published evaluations, or in-house studies. Of course, this method will not work for new analytes, for which there are no data. Moreover, the problem has not really been solved, it has been transferred. One must now decide what is a sufficiently low complaint rate.

Allocating Total Error into Goals for Total Error Sources

Setting goals for error sources that contribute to total analytical error creates a rather complicated problem. Goals for total analytical error sources have the constraint that the sum of combined values of the individual sources cannot exceed the total analytical error goal. Error modeling (propagation) can be used to achieve this.

Inadequate Goals

A **non existent goal**, while a rather obvious category, crops up surprisingly often. An example is lack of an outlier goal. Outliers are values that are so far away from the true values that they almost always cause problems. Yet, there is seldom a goal describing how far off a value must be to called an outlier and how many outliers are acceptable. Ideally, there should never be outliers, but an implied goal of zero outliers has its own problems. Realistically, there will always be a small but measurable frequency of outliers, even for the best assays. A goal of zero will be an **unrealistic goal** and not useful.

Aids to Improve Goal Setting

A goal requires a clear pass/fail criterion. This implies that a metric is in place which will be used to base the decision. **Deciding on a metric** is a step in agreeing on goals. A metric should be objective, easy to understand, and relevant to the goal. Not all metrics are appropriate. For example, one might use intuition[1] as a metric: "I've seen the data and feel the assay is OK." Metrics can be evaluated to assess their validity. This could be done with simulations (e.g., without actually running assays). For example, the correlation coefficient is sometimes used as a measure of agreement between two methods, in spite of reports that caution this metric's limitations. One can simulate what values of correlation coefficients would be observed with various expected datasets to see how well the correlation coefficient predicts agreement between methods. Alternatively, one could retrospectively analyze real data with the proposed metric, again testing its predictive value.

Preparing **cause and effect diagrams** (also called fishbone or Ishakawa diagrams) helps to highlight potential analytical problems and focus goal setting. The universe of potential goals must be narrowed down to those that have a reasonable likelihood of causing problems. Otherwise, the list of goals to be tested could be endless.

One can go further with a cause and effect diagram by **mathematically modeling** error sources such that sources combine to yield an estimate of total error. This method is also called propagation of errors. It allows goal limits to be assigned to total analytical error sources. One study showed that for a cholesterol assay, traditional analysis underestimated total analytical error

[1]This does not mean that one should not use intuition. This is an old statistical adage: "Beware of the following: *Statistics on - Brain off*"

compared with a method that estimates total analytical error and its sources.[3,4]

How goals are and should be used

Manufacturers, like mostly everyone, are faced with yes/no decisions. Should we release or not release the product (meaning has it met or not met its goals)? Consider two assays, however, where assay **A** is just inside and assay **B** is just outside of spec. From a manufacturer's standpoint, assay A has *full value* (i.e., identical to an assay that is perfect), whereas assay B has *zero value*. From a customer standpoint, the two assays have *similar* performance (and value). Yet, manufacturers must still treat these two assays that are similar to a customer, as either having full or zero value.

There is no easy way to deal with this problem. What happens in practice is that the yes/no region, while conceptually clear (i.e., accept if precision is less than or equal to 4.0% CV, reject if greater than 4.0% CV) becomes fuzzy: accept if precision less or equal to 4.0% CV; conduct further discussions if precision is between 4.0 and 5.0% CV; and reject if greater than 5.0% CV. Thus, targets set at the beginning of a project are revisited throughout the project development cycle and especially near product release. Since many specs are set close to the technical capability of a system due to competitive pressures, the situation occurs frequently.

Goals during the commercialization of the assay (the claims that are made to customers)

When the product is released, internal goals are transformed by the manufacturer into customer claims, which may or may not be the same as the internal goals. The claims represent a data source for customers, who can compare different manufacturers claims. Of course, as most consumers are aware, not all claims are always met! There is an additional problem. A claim can be stated in a way that is not clear, leading to confusion. Basically there are two types of claims:

1. The **"typical" data claim** - Half of the customers are expected to observe better performance, and the other half, worse performance. With statistical tests, one can determine whether performance is unreasonably far from the claim.

2. The **"warranty**" claim - Here, all customers are guaranteed performance better than the limit.

An NCCLS subcommittee is trying to address these problems by providing guidelines for standard language to be used for claims.

References

1. Skendzel LP Barnett RN and Plat R. Medically useful criteria for analytic performance of laboratory tests. *Amer J Clin Pathol.* 1985;83:200-5.

2. Green PE and Wind Y. New way to measure consumers' judgments. Harvard Business Review 1975;107-17.

3. Krouwer JS. Estimating total analytical error and its sources. Techniques to improve method evaluation. *Arch Pathol Lab Med.* 1992;116:726-31.

4. Krouwer JS and Monti KL. A simple, graphical method to evaluate laboratory assays. *Eur J Clin Chem Biochem.* 1995;33:525-7.

Analytical Goals, the Total Testing Process, and Patient Outcomes

Derek P. Lehane, Ph.D., DABCC
Technical Manager, Market Development
DuPont Diagnostics
Wilmington, Delaware

Abstract: Specifications are routinely required to address the total testing process that includes all pre-analytical, analytical, and post-analytical variables. Rigorous specifications must be developed to address the issues of collection of appropriate specimens as well as the increasing use of bar-coding, optical scanning and information technology links in sample identification. Efforts to fully automate sample accessioning and processing are creating new needs for goals to drive emerging engineering concepts. As electronic reporting of data becomes commonplace, there is a growing need for clinically relevant specifications to connect expert systems and object-oriented database management systems to improved patient outcomes.

Analytical goals also need to reflect diagnostic impact and patient outcomes. In developing cardiac protein assays (CKMB, cardiac troponin) or thyroid hormone assays (TSH, free thyroxine), the specification of specificity and functional sensitivity are a necessary first step. However, only rigorous clinical trials and use of receiver operator characteristic plots can answer key diagnostic questions. Can mass CKMB as a single biochemical marker (together with clinical findings and EKG) appropriately diagnose myocardial infarction, or is concurrent testing with LD1 or cardiac troponin required for optimal diagnostic efficacy? What level of functional sensitivity is required to diagnose hypothyroidism in hospitalized patients with nonthyroidal illness? Is use of a test with rigorous analytical goals sufficient to contribute to improved health of the patient?

Introduction

This presentation has one overarching theme: seizing opportunities in and developing standards and specifications for the management of information in improving patient care. Development of standards in bar coding and data streams has immeasurably improved the laboratory's ability to identify specimens and transfer data with virtually no errors and high levels of productivity. Interchange of clinical information among independent health care-oriented computer systems is now a reality through standards developed for peer-to-peer data transfer or for the use of smart interfaces. Automation of sample preparation and distribution is the focus of work in developing standards and specifications so that robotic systems and clinical laboratory analyzers can effectively interface. Expert systems - under used in laboratory medicine - can be extremely helpful in managing analytical processes and in analyzing complex laboratory data.

Managing information to better meet clinical needs is discussed with reference to cardiac markers and thyroid-stimulating hormone (TSH). Statistically measuring diagnostic performance of, for example, creatinine kinase MB isoenzyme (CKMB) effectively requires the use of well designed clinical trials and of receiver operator characteristic (ROC) curves. Only such approaches can determine whether use of

CKMB alone, or in combination with other cardiac proteins, can effectively aid in diagnosing myocardial infarction. Performance guidelines developed by the American Thyroid Association[1] for thyroid-stimulating hormone are effectively challenging our views on the value of the clinical information provided by first-, second-, and third-generation immunoassays.

Information Management

Altschuler[2] tells us that the practice of medicine consists largely of information management, that health professions do not always use objective data appropriately and, when data are not used appropriately, care is often poorer and costs higher than they would otherwise be.

In analyzing mistakes occurring in laboratory testing, Boone and Ross[3] showed that only 7% were due to analytical problems, while 93% were due to pre- and post-analytical errors. However, recent advances in information technology have significantly improved the areas of sample identification and data streams, while robotics are advancing rapidly in improving control of pre-analytical variables and expert systems are emerging to deal with post-analytical variables.

The American Society for Testing and Materials (ASTM) has developed a standard (E 1466-92) for the form, placement and content of bar code labels on specimen tubes that are used on clinical analyzers. By specifying the use of Code 39 with standard check digit or code 128 in place of older, error-prone symbologies, unparalleled levels of reliability in sample identification are now achievable in clinical laboratories. ASTM standard E 1394-91 covers the transfer of information between clinical instruments and computer systems. If widely adopted by manufacturers of laboratory information systems (LIS) and clinical analyzers, the standard would obviate the need for developing LIS-specific interfaces and would provide a true '"plug and play" environment. An "interpretation box," such as provided by Dawning Technologies, can yield the same reliability even if a manufacturer's data stream is not compatible with E 1394-91.

While information technology is available to dramatically improve the reliability of sample identification and data streams, it is found in less than half of U.S. hospitals. Far fewer hospitals have moved to the next stage, the electronic interchange of patient demographics, orders and results for laboratory tests, imaging studies, etc., among multiple sites. This can be achieved by using a global peer to-peer data transfer standard such as Health Level Seven or by using smart interfaces that adapt to various protocols and legacy (i.e., existing) systems. Medical imaging has taken the lead in this area and, through the use of customized digital imaging systems, captures images from multiple types of diagnostic equipment and display and print them on file anywhere in the network, allows the radiologist to view images and provide consultation from home or office, and provides greater coverage and greater utilization of human resources. The implications for the clinical laboratory are obvious.

The significant unmet need in clinical laboratories is automating sample preparation and distribution. As very large and very expensive robotic systems arrive to meet this need, a parallel need arises for worldwide standards to facilitate optimizing, interfacing and integrating clinical analyzers. The Clinical Testing Automation Standards Steering Committee, formed in 1994 by representatives from clinical laboratories and

manufacturers, has embarked on a multi-year process to develop standards for laboratory automation.

Expert systems, also known as decision-support systems, are now seen with increasing frequency in industrial settings[4] but, disappointingly, are rarely encountered in clinical laboratories. Focused on an appropriate problem, the rule-based, decision tree-based or case-based reasoning expert systems can be powerful tools in managing laboratory processes, in troubleshooting clinical analyzers in analyzing laboratory data and in diagnosing illness. The reasons why expert systems seem to fail is either that technology integration into an already turbulent environment is difficult or that the system is being grafted onto an existing, outmoded workflow, or that the system is not continually "refreshed" with new knowledge.

Analytical Goals and Patient Outcomes

Typical analytical goals (reproducibility, assay linearity, sensitivity, accuracy, correlation, etc.) are essential in developing an assay but alone may not be sufficient to deliver a product with the requisitive diagnostic sensitivity, specificity and efficiency. This will be illustrated using cardiac proteins and thyroid function tests as examples of how to improve the diagnostic information provided by these tests and, in turn, to positively impact patient outcomes.

Well designed and executed clinical trials are an essential first step; in assessing an assay for free thyroxine (FT4), for example, diagnostic performance must be assessed in all patient groups usually encountered, including patients with non-thyroidal illness and extreme binding-protein anomalies.[5] The second essential step is to analyze the clinical data so collected using ROC plots and predictive value theory.

ROC plots[6] provide a graphical description of test performance representing the relationship between the true-positive fraction (sensitivity) and the false-positive fraction (specificity). Clinical accuracy, in terms of sensitivity and specificity, is displayed for the entire spectrum of decision levels. In understanding the value of diagnostic information provided by various cardiac protein assays, ROC plots have been used to optimize clinical performance.

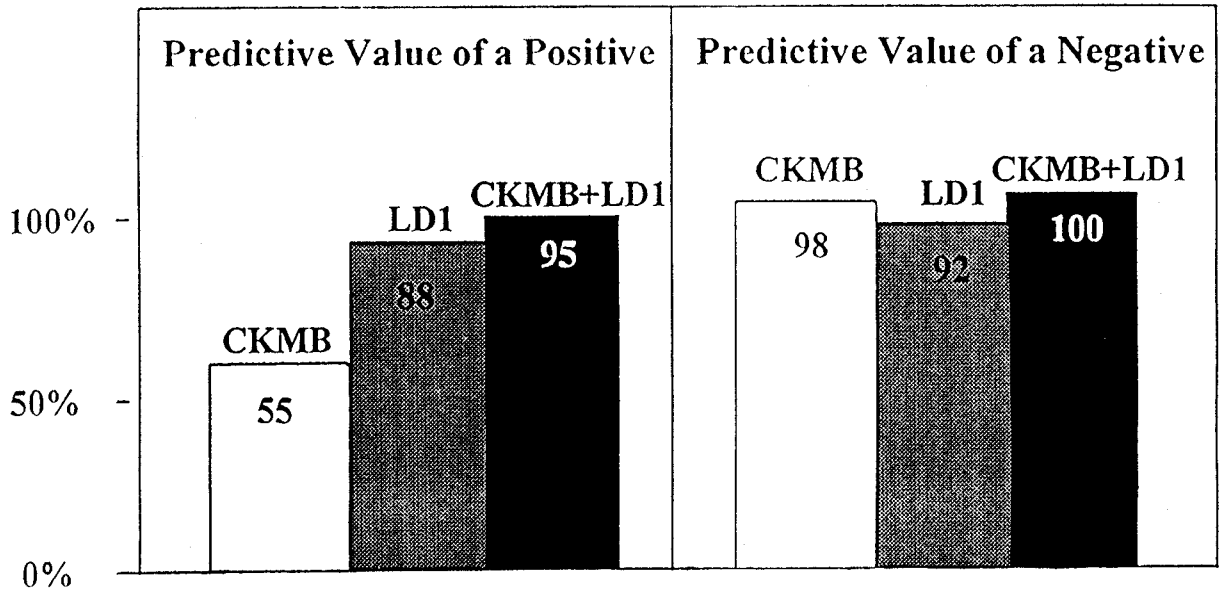
The Power of Receiver Operator Characteristics (ROC) Plots

- ROC plots allow direct comparison of assays using equivalent or different units
- ROC plots provide a means of comparing various reporting units for a given assay
- ROC plots allow direct comparison of clinical performance between assay having significant biases
- ROC plots allow optimization of clinical performance

Ideally, while we want to use a single, powerful test to help rule in or rule out a disease state, applying predictive value theory to the clinical trial data demonstrates the virtual necessity of some form of combination testing. The commonly used cardiac protein tests (myoglobin, CKMB, troponin I (cTnI), troponin T (cTnT) and lactate dehydrogenase isoenzyme 1 (LD1)) all demonstrate distinct temporal rise and fall curves. For example, applying predictive value theory demonstrates the diagnostic performance of CKMB testing alone or in combination with LD1 testing (Table 1).

Similarly, Wu and colleagues[7] have used

Effect of Combination Testing on Predictive Values



The value of using LD1 beomes apparent when the predictive value of the combination, CKMB+LD1, is compared to each test individually. The chart above shows that the predictive value of a positive combination test (CKMB and LD1 both positive) is greater than either test alone. This means that when both CKMB and LD1 are positive at any time during the sampling period, the probability that the patient has had an MI is 95%. The predictive value of a negative for the combination is also improved to 100% versus either test alone.

Table 1. Effect of combination testing on predictive values.

ROC plots to show that the diagnostic information provided by CKMB testing is superior to that provided by CTnT from 6 to 24 hours after acute myocardial infarction (AMI), that the information provided by both tests is equivalent from 24 to 48 hours after AMI and that cTnT provides more information from 48 to 96 hours after AMI. They also conclude that CKMB is more specific for diagnosing AMI and propose that cTnT is more sensitive to myocardial injury. These findings appear to indicate that combination testing of CKMB and cTnT yields optimal diagnostic information.

In thyroid function testing, the quest for improved diagnostic information has been aided by the development of the American Thyroid Association's performance guidelines for TSH.[1,8,9] These guidelines are perhaps the most definitive consensus specifications available for an analyte, and particularly challenge the assay developer with regard to assay reproducibility at subnormal TSH levels. Spencer[8,9] has popularized the concept of first-, second- and third-generation TSH assays, based on functional sensitivity performance.

Functional Sensitivity of TSH Assays

The lowest TSH concentration that achieves an interassay CV of 20%:

Generation	Functional Sensitivity Limit
First (RIA)	1.0 - 2.0 mIU/L
Second (Immunometric)	0.1 - 0.2 mIU/L
Third (Immunometric)	0.01 - 0.02 mIU/L

Functional sensitivity in clinical practice is usually suboptimal in comparison with that reported by the manufacturer.

By combining data from ROC plots and by assessing functional sensitivity, various authors who have studied newer TSH assays raise two intriguing points about the quality (and cost) of diagnostic information provided by these assays.[10,11]

- Second generation TSH assays with appropriate functional sensitivity can match the diagnostic information provided by third-generation assays, but at lower cost.

- TSH values alone, even when obtained from third-generation assays, may not always indicate whether a patient is hyperthyroid; thus, combination testing (TSH and FT4) may be an appropriate strategy to maximize diagnostic information.

Postscript

The current focus on managing diagnostic information, whether through using information technology to minimize pre- and post- analytical variables or developing more powerful assays that maximize diagnostic information, is providing major benefits in laboratory medicine. Five areas of information management, however are suggested that are expected to provide ample benefits for laboratorians:

- driving the benefits of bar-coded sample identification and laboratory information systems into most clinical laboratories

- developing systems for peer-to-peer information transfer that combine and display laboratory data, imaging studies, etc., across multiple geographic sites

- encouraging the use of decision-

support tools to improve laboratory productivity and to maximize the informational content of laboratory data

- seeking a consensus on the design of clinical trials and a uniform manner of expressing the resulting performance data

- achieving a consensus on the coherent use of cardiac proteins in diagnosing AMI.

References

1. Hay ID, KIee GG. Linking medical needs and performance goals: clinical and laboratory perspectives on thyroid disease. *Clin Chem*. 1993;39:1519-24.

2. Altschuler CH. Data utilization, not data acquisition, is the main problem. *Clin Chem*. 1994;40:1616-20.

3. Boone DJ. Governmental perspectives on evaluating laboratory performance. *Clin Chem*. 1993;39:1461-7.

4. Port 0. Computers that think are almost here. Business Week, July 17, 1995;68-72.

5. Christofides ND, Sheehan CP. Multicenter evaluation of enhanced chemiluminescence labeled - antibody immunoassay (Amerlite - MAB™) for free thyroxine. *Clin Chem*. 1995;41:24-31.

6. Assessment of the clinical accuracy of laboratory tests using receiver operator characteristic (ROC) plots. NCCLS Document GP10-A. Villanova, PA 1995.

7. Wu HB, Valdes Jr. R, Apple FS et al. Cardiac troponin - T immunoassay for diagnosis of acute myocardial infarction. *Clin Chem*. 1994;40:900-907.

8. Nicoloff JT, Spencer CA. The use and the misuse of the sensitive thyrotropin assays. *J Clin Endocrinol Metab*. 1990; 71:553-8.

9. Spencer CA. Interlaboratory/intermethod differences in functional sensitivity of immunometric assays of thyrotropin (TSH) and impact on reliability of measurement of subnormal concentrations of TSH. *Clin Chem*. 1995;41:367-374.

10. Roden M, Nowotny P, Hollenstein U et al. Equivalent discrimination among states of thyroid function by immunochemiluminimetric and immunoradiometric determination of thyrotropin *Clin Chem*. 1993;39:544-54.

11. Taimela E, Tahtela R, Koskinen P. et. al. Ability of two new thyrotropin (TSH) assays to separate hyperthyroid patients from euthyroid patients with low TSH. *Clin Chem*. 1994;40:101-105.

Summary of Workshop 5: Establishing Analytical Performance Goals

Facilitator: Peter Howanitz, M.D.
Director, Clinical Laboratories
UCLA Medical Center
Los Angeles, California

CDC Liaison: Thomas L. Hearn, Ph.D.

Key Questions:

1) How are analytical performance goals established and evaluated for new technologies?
2) How should such goals be established and evaluated?

Presentations and discussions in the workshop on establishing analytic performance goals focused on answering the above two key questions.

To set the stage for discussion of the issues among the 23 workshop participants, Presentations were made by Drs. Callum Fraser, Jan Krouwer, Charles Handorf, and Derek Lehane. Dr. Fraser provided historical and international perspectives on analytical goals as well as a challenge to journal editors, industry, and external quality assurance organizers to be more active in disseminating analytical goals. The use of analytical goals by health industry manufacturers and the problems of developing and communicating goals between the industry and laboratorians was described by Dr. Krouwer. Dr. Krouwer emphasized the importance of metrics, measurement-based protocols, and reiterative processes for establishing, monitoring, and achieving analytical goals. Dr. Handorf discussed analytical goals within the context of the total health and medical management system, reviewed what our track record has been in the use of analytical goals, and provided a look forward at how analytical goals could be applied to newer technologies such as point-of-care and molecular pathology. Finally, Dr. Derek Lehane put analytical goals into a larger perspective--the perspective of the total testing process and the perspective of patient care. Opportunities to incorporate engineering and electronic information systems were highlighted and the need to merge assessment of analytical goals with clinical outcomes measures were highlighted by Dr. Lehane. Manuscripts from the four speakers are included within this chapter.

In response to *how are analytical performance goals established and evaluated for new technologies and how should analytical goals be established and evaluated,* the answer appears to depend on who is using the analytical goal and who is developing the analytical goal. Two major groups are certainly involved, laboratorians and instrument manufacturers, but clearly, there is at least one more.

Laboratorians, as reviewed by Dr. Fraser, have developed a wide variety of different strategies. At least 17 different strategies have been used in the last approximately 30 years. These strategies continue to change, continue to be improved, and continue to be honed. Laboratorians usually use the

coefficient of variation (CV) as the statistic of choice for a measure of the analytical goal with inaccuracy and imprecision as the parameters used to describe analytical goals. In some cases, for example, qualitative analyses, no goals exist whatsoever.

Analytical goals are developed entirely differently by manufacturers. There is some interest paid in using the clinical goals established by laboratorians, but regulatory and competitive needs are also very important to manufacturers in establishing goals. Manufacturers have internal specifications which include data and warranty claims and rely on a system of metrics, targets, protocols, and analyses to develop, implement, and monitor progress on the goals. Thus, the workshop concluded that laboratorians and manufacturers approach goals in entirely different ways.

A third group with a special role in analytical goal setting is the clinician group. Some in the workshop commented that clinicians don't actually set analytical goals and have not been as involved with others in setting them as is desirable.

How should analytical goals be established and evaluated? Again, laboratorians and instrument manufacturers do it two different ways. Laboratorians believe that analytical goals based on biology is best. For example, a CV, which is expressed in some fraction such as .5 of the biological CV, was described as an appropriate analytical goal for some scientists. Current consensus was that biology-based goals should be used by all. Some participants also pointed out that this consensus might, in fact, be geographic; that is, clearly in Europe a consensus exists, but in the United States we continue to refine goals that we've had.

Some laboratorians also mentioned that clinical goals should precede analytical goals--and that analytical goals should take into consideration other sources of data such as **quality** assurance, proficiency, and training. Pursuant to this point was some discussion about the merits of performing daily quality control when testing is conducted in non-traditional sites and when single use devices are employed. One person noted that in some of these situations, individuals performing quality control always obtained the same results. In those circumstances, continuing education about the importance of achieving the analytical goals may be helpful. One idea that emerged is that editors of Clinical Chemistry and similar journals, which provide reviews of emerging technology, should begin to require use of the analytical goal as part of their evaluation of new technology and either include this as part of the acceptance process.

Having manufacturers jointly establish their goals with laboratorians was an important theme emphasized not only by the manufacturers but also by the laboratorians. Manufacturers' goals should also be based on consideration of the total testing process and the implications of the errors throughout the total testing process--for instance, the use of bar coding may, in fact, address one of the major problems in defining a result from a laboratory test. An error in the bar coding step, the specimen identification step, certainly may be far greater than the error in the measurement step itself Additionally, clinical goals should be considered when manufacturers establish analytical goals.

Participants also stressed that the analytical goal should be included as part of the information management system and that (a) laboratorians evaluate the information and (b) present the information for clinicians incorporating the analytical goal within the

presentation.

The workshop participants suggested strategies and methods for developing and establishing analytical goals. One part of the discussion was the sense of urgency for having some carefully constructed research protocols. The remainder of the discussion centered more on a strategy for choosing analytical goals. For example, if goals based on biological variability are selected then there are some potential problems--in reviewing the literature some goals are too strict, others are insufficiently so. The need for a fall-back position for each and every procedure, that is, an alternative goal which could be used if a primary goal does not meet the required specifications, was suggested.

The analytical goals that are in place today are nearly entirely for quantitative tests. The workshop was essentially unanimous in believing that developing methods for determining analytical goals for non-numeric (qualitative) tests should be a very high priority. Examples of these situations are molecular biology testing where test results are reported as a "plus or a minus"-positive or negative--and for microbiology. One suggestion was that the measurable goal in these cases could be based on efficacy, but others expressed that clearly much more thinking on this needs to be done.

Two other issues that were raised within the workshop dealt with the roles of information systems and of quality management schemes in analytical goal development. A need for some demonstration projects indicating how systems might be developed to transfer information from peer to peer was suggested. With regard to quality management systems there was some discussion of quality management and how analytical goals fit into the quality management scheme, but no clear agreement about how to establish and measure quality management goals. Again, much more needs to be done in this arena.

There was a discussion that perhaps we needed some other alternative metrics on how to describe analytical goals. As an example, one manufacturer indicated using a metric called “capability indexes” in developing his goal within his company.

The workshop participants sought a definition of quality management for non-traditional testing, for example, point-of-care testing. What should the goals be" Should they be different from what they are in the clinical laboratory? Should they be tied to patient care? And how does one decide at those locations when enough is enough? Again, participants believed that practical clinical goals should be established for new technologies. There was a discussion about the interpretation of tests in relationship to clinical goals and a suggestion that diagnostic algorithms might be useful. A model suggested was measuring cardiac enzymes for acute myocardial infarction and evaluating the usefulness of analytical goals on a diagnostic algorithm.

While there is often concern about inability to meet analytical goals, there are tests and testing laboratories which find that they can exceed the goals, far surpassing the medical need. These are important opportunities for reducing costs in terms of reducing the frequency of quality control. Workshop participants recognized that decreasing the regulatory personnel standards introduced the potential for reducing testing quality. In some cases, laboratories which once met analytical goals now may no longer be able to achieve the

analytical goals if poorly trained and educated laboratorians are employed. Within the framework of the discussion about personnel concerns, one instrument manufacturer revealed that in the last few years they have found laboratory supervisors and directors unwilling or unable to send laboratory technologists for training on new equipment. In fact, there was great concern raised by instrument manufacturers and laboratorians that the changing quality of personnel might inadvertently influence the analytical goals that were established in the past. These goals may be remarkably different in the future.

The last segment of the workshop was spent addressing funding mechanisms, untapped data and information resources, and future collaborations. In seeking to identify who could be responsible for funding analytical goal research initiatives there were no new ideas. There was agreement, however, that new non-traditional sources of research funding are essential because of decreased resources among traditional providers and because of the need to include segments of the health care system (for example, managed care payers) who are buyers of laboratory services.

Data already collected on analytical performance by instrument manufacturers and by laboratorians appear to be rich ground for assessing the state of analytical goals and where the gaps exist. There were suggestions that scientific community should begin using the information, that manufacturers should start sharing this information with each other, and that large data bases could be assembled and shared not only among manufacturers but also among laboratorians and others involved in the collection of this data. Because much of the manufacturers' data are from clinical trials, it was felt that consensus should be sought on how to design and conduct clinical trials so that information was obtainable and analyzable in some fairly if uniform ways. Marked differences among how manufacturers currently approach clinical trials makes the linking of analytical goals to such things as receiver/operator curves almost impossible today.

In a discussion about what laboratorians should do in developing, implementing, and assessing analytical goals, some manufacturers felt that laboratorians haven't done their fair share. Moreover, instead of directing all of their attention to what else needs to be done, laboratory scientists need to begin to examine what is of little merit or value, that is, provides no benefit. In particular the laboratory community has been barraged with a large variety of different regulatory processes--some of questionable value. These must be re-evaluated and, if are found to have no value, we should to abandon them. Laboratory scientists, manufacturers, and users and purchasers of laboratory services should make their voice known so those responsible for developing regulations can hear the collective voice and can take appropriate action.

One important difference that was raised in the workshop was the markedly different approach that academicians and laboratorians used for establishing goals versus instrument manufacturers' approach of relying on market driven goals. There was some discussion that medically driven goals should replace market drive goals. Instrument manufacturers should no longer be looking at improving their penetration into a market by offering technologies that far exceed analytical goals as a way to give them a “leg up” on the competition, and laboratorians, as part of this process, should

say "enough is enough. We have enough precision, we have enough accuracy, and we are wil1ing to pay for the additional accuracy and precision."

A fair question to ask is "who cares about analytical goals?" After long discussion, workshop participants concluded that we all must care ... sometimes. Limiting focus to analytical goals may be directing attention to the wrong component of quality. Analytical goals should be based on other performance characteristics. So, the reason we must all care "sometimes" is that there are situations in which the analytical goals far exceed what is needed clinically. If a major error occurs in another process in the pre-analytical or post-analytical phase, then we should not be concerned about the analytical goal, but about the errors. Also, if tests are being performed that have little or no value in patient care, then analytical goals for those procedures are of limited value.

In closing, despite the lack of a consensus about analytical goals for laboratory procedures there was a sense that useful information was being provided. A slide which a manufacturer presented illustrating one way the industry established goals for its customers--laboratorians--stimulated thoughts about a paradigm for the future. Manufacturers give laboratorians choices. Among the choices are cost, sample volume, and precision (that is, the analytical goal). In a focus group format, laboratorians are then asked to choose which of the two they want. In the future, as the way laboratory medicine is practiced, as the way health care is delivered, and as health care reimbursement schemes evolve the laboratory community will be called upon to make more choices. In making good choices laboratory scientists should incorporate their unique expertise and knowledge about the components of quality in the testing system to derive reasonable, health- effective analytical goals.

In conclusion, another important discussion point raised was that despite having long analytical goals, we still can get the right answer. Again, I thank the participants and panelists who devoted exceedingly large amounts of their time to this topic.