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

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

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

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

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

Introduction

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

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

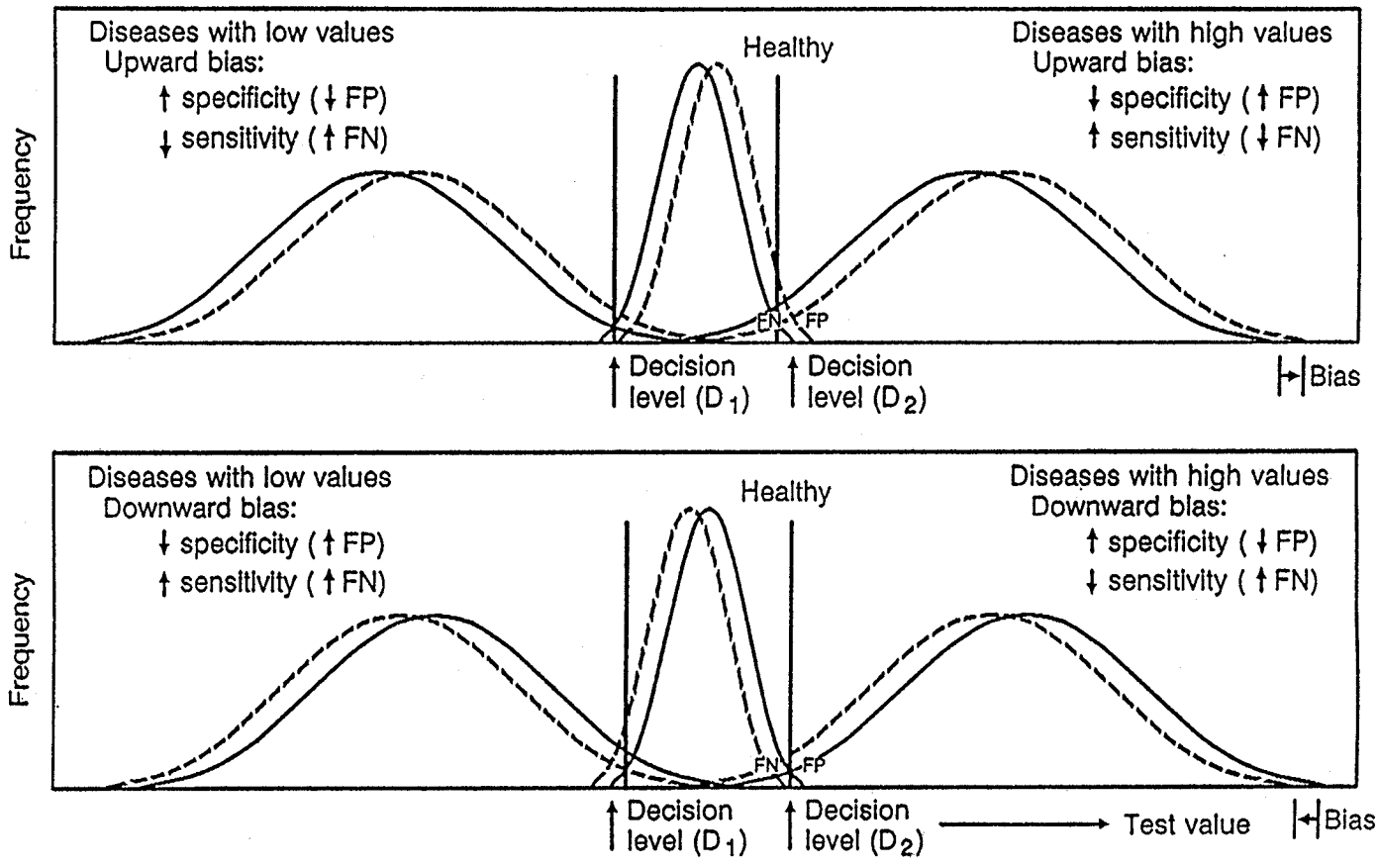

Figure 1. Effect of analytical bias on the classification of patients.

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

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

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

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

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

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

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

Effect of Bias on Sensitivity and Specificity

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

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

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

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

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

Table 1. Effect of bias on sensitivity and specificity.

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

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

Proposed System to Define Medically Important Bias Limits

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

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

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

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

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

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

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

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

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

Discussion

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

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

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

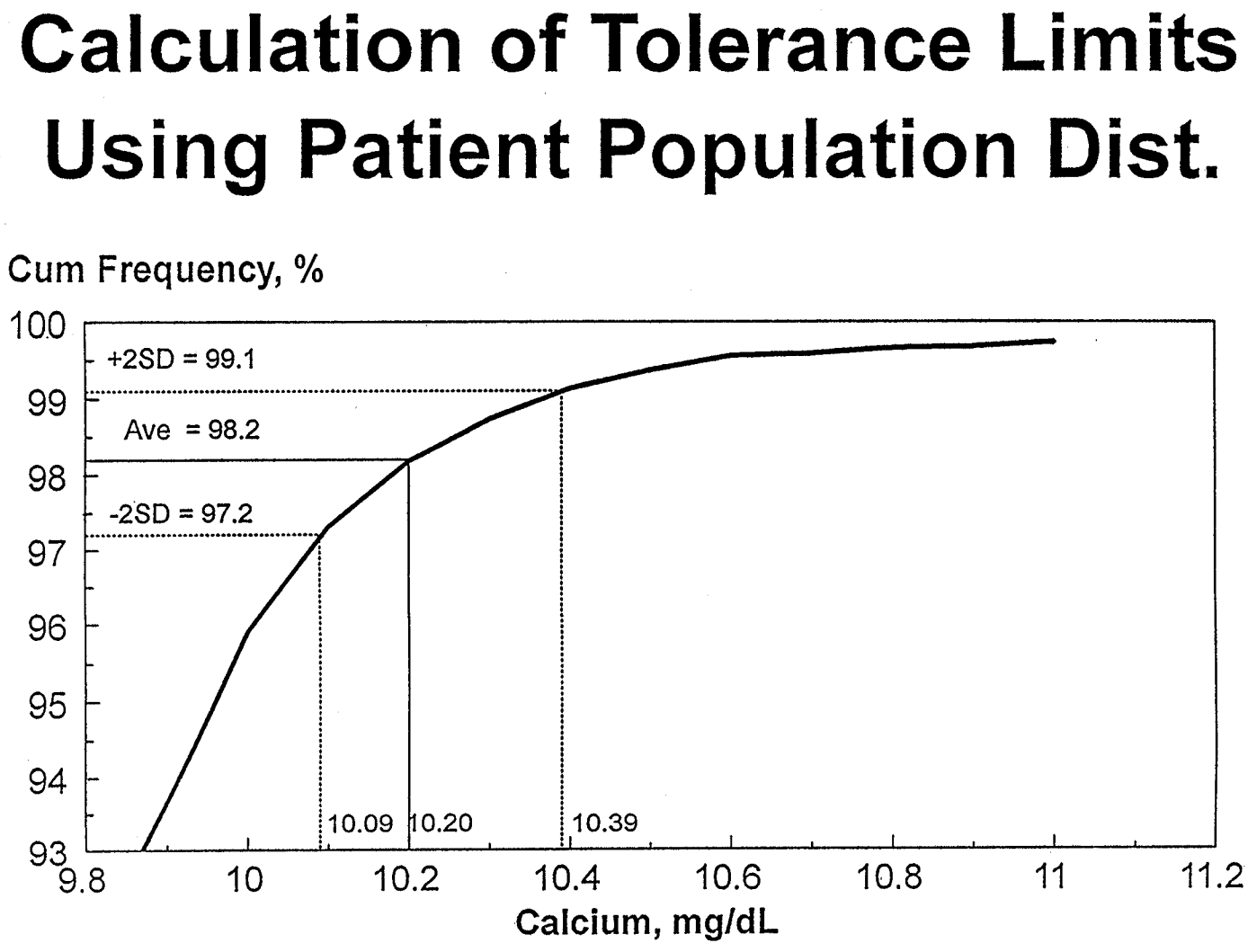

Figure 2. Calculation of tolerance limits using patient population distribution.

met with less control monitoring than currently used.

Conclusion

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

Prevention Effectiveness Case Study: Institutionalizing Prevention of Group B Streptococcal Infections

Anne Schuchat, M.D.
National Center for Infectious Diseases
Centers for Disease Control and Prevention
Atlanta, Georgia

Abstract: The value of cooperation between clinicians and laboratorians is scarcely controversial, yet failure to assure adequate laboratory participation in implementing new programs can have serious clinical consequences. The recent example of preventing group B streptococcal (GBS) disease in newborns is illustrative. Neonatal GBS infection can lead to death, long term disability, and substantial direct and indirect health care costs, while GBS prevention programs, when implemented appropriately, are cost-saving. Much perinatal GBS disease is potentially preventable through using prenatal screening cultures and administering antimicrobial prophylaxis intrapartum to mothers at increased risk of transmitting the infection to their newborns. Efforts by clinician groups to promote prevention strategies during 1992 appear to have had no measurable impact on disease incidence. GBS-related sepsis and meningitis continue to strike nearly 8,000 newborns each year in the United States. A series of investigations of the apparent failure of recent prevention efforts has identified potential pitfalls of implementing clinical programs without laboratory involvement and/or acceptance. For example, most clinicians reported collecting screening cultures, but few used appropriate culture sites. Most microbiology laboratories surveyed were using insensitive culture methods to process the specimens, limiting the validity of reported results. Results of late gestation screening cultures are needed by clinicians at the time and place of delivery. Thus, even when cultures are collected and processed appropriately, lapses in reporting can limit the ability of clinicians to optimally manage the mother-infant pair. Evaluating program effectiveness requires monitoring clinical outcomes of interest, many of which are best identified through laboratory-based surveillance systems (e.g., surveillance for invasive GBS disease and for infections due to antimicrobial resistant pathogens). As in this example, successful clinical management as well as disease prevention programs require a commitment from clinicians and laboratory personnel to coordinate practices and program evaluation.

Introduction

The value of cooperation between clinicians and laboratorians is scarcely controversial, yet failure to assure adequate laboratory participation in program implementation can have serious clinical consequences. In this session, we are concerned with patient outcomes. I will explore the evaluation of a prenatal screening program aimed at preventing perinatal infections as one example of how laboratory practice research can contribute to improving clinical outcomes.

The clinical outcome I will discuss is every family's nightmare: following a normal pregnancy, a few hours after delivery, a parent is faced with a critically ill newborn. The baby is infected with group B

Streptococcus, a leading cause of sepsis and meningitis in newborns in the United States. The parents have never heard of group B streptococcal (GBS) disease and are surprised to learn that many of these infections are preventable. To understand what might have gone wrong for these parents, I will focus on laboratory and clinical practices that may contribute to failed preventive measures. By reviewing the process by which the Centers for Disease Control and Prevention (CDC) has been evaluating the effectiveness of GBS disease prevention in the United States, I hope to highlight the interconnected nature of laboratory testing, clinical management, and patient outcomes.

The Disease

Group B *Streptococcus* first emerged as an important pathogen in the 1970s.[1] About 7500 cases of GBS sepsis and meningitis in newborns are reported each year. [2] The burden of perinatal group B streptococcal disease extends beyond neonatal illness and death, and includes long term disabilities such as hearing loss, impaired vision, and developmental problems.[3,4] Maternal morbidity from GBS includes sepsis, amnionitis, postpartum wound infections, and stillbirths.[5,6] The direct costs of neonatal disease alone in the U.S. have been estimated as $300 million annually.[7]

Most GBS disease among newborns results from maternal to infant transmission during labor and delivery. Many women are asymptomatically colonized by GBS in the genital and gastrointestinal tracts. About half the infants born to colonized mothers are themselves colonized on the skin and mucosal surfaces. Most of these infants, 98%, are asymptomatic. About 2%, however, will develop early onset disease, presenting with sepsis, pneumonia or meningitis in the first few days of life.

The Screening Test

To identify women with an increased risk of transmitting GBS to their newborns, clinicians would like to know which women are colonized with GBS in the genital or gastrointestinal tract. GBS colonization is not static, though, and women can acquire or lose carriage during the course of pregnancy.[8] Although clinicians would like to know maternal colonization status at the time that labor begins or membranes rupture, results of cultures collected at labor onset will not be ready before most women deliver, while the intervention--antibiotic prophylaxis--is ineffective in preventing transmission unless initiated before a woman delivers. Clinicians currently rely on prenatal cultures to predict intrapartum GBS colonization. The laboratory test, therefore, is at best an indirect measure of the intrapartum colonization status, but risk analysis in a large cohort study suggested that women with GBS identified by prenatal cultures had 29 times higher risk of delivering an infant with early onset GBS disease, compared with women whose prenatal cultures were negative.[9]

Prevention Strategies

Efforts to prevent perinatal GBS disease have focused on antimicrobial chemoprophylaxis. During the 1980s, investigators demonstrated that giving antibiotics to women early in pregnancy, or to infants after birth, was not effective in preventing GBS disease. Giving antibiotics during labor, however, proved to be extremely effective in reducing maternal infections and preventing early onset disease in newborns. During the 1990s, debate grew

over which women should receive antibiotics. Efforts to focus antibiotics on women who could benefit most but avoid exposing millions of low risk women to antibiotics led to several potential prevention strategies. Authorities have considered giving antibiotics to all maternal carriers,[10] to women with obstetric risk factors for GBS disease regardless of colonization status,[11] or to GBS carriers who also have obstetric risk factors.[12]

In 1992, clinical organizations began promoting prevention strategies. The American College of Obstetricians and Gynecologists (ACOG) published a technical bulletin on prevention in July 1992,[13] followed by two clarifications of their position in 1993.[11,14] ACOG recommended against prenatal screening cultures, and stressed that intrapartum prophylaxis should be given to all women with specific obstetric risk factors. The American Academy of Pediatrics (AAP) published guidelines for prevention in November of 1992,[12] recommending an approach that combined prenatal screening cultures with intrapartum treatment of GBS carriers who developed obstetric risk factors. Although the strategies differ, it was expected that consistent application of either approach would prevent from 60 to 75% of early onset cases. CDC undertook a series of investigations to assess whether effective prevention was occurring, and if not, why not.

Disease Detection

To collect population-based information on GBS disease, CDC has been collaborating with investigators in academic institutions and state health departments on active laboratory-based surveillance for invasive bacterial disease. Through regular contact between regional surveillance officers and microbiology and infection control personnel in all acute care hospitals in each surveillance area, simple clinical and demographic information is collected on all cases of illness where GBS is isolated from a usually sterile site, such as blood or cerebrospinal fluid. Surveillance in the multi-state population identified no reduction in incidence of early onset GBS disease between 1991 and 1993.

Prevention Effectiveness

To determine why the 1992 prevention statements had no discernible impact on disease occurrence, we conducted a series of surveys. The objective of these surveys was to identify potential barriers to effective GBS prevention. CDC first collaborated with the Georgia Department of Human Resources on a survey of obstetric caregivers in this state.[15] The survey of clinicians revealed that most respondents screened at least some of their prenatal patients for GBS. However, only 9% reported that they cultured the optimal sites -- vagina and rectum. Many clinicians were collecting cervical cultures for GBS, but cervical cultures are often negative when vaginal or rectal cultures are positive. Few clinicians knew what methods their laboratories used to process the clinical specimens, so in 1994 we queried the microbiology laboratories serving hospitals in our multi-state active surveillance system.[16] Results from a survey of over 200 clinical labs in five states suggested that very few labs used selective broth media, although use of this method can increase recovery of GBS by about 50%.

These surveys identified several problems with prevention practices. Results suggested that clinician and laboratory practices during 1993 and 1994 would not have been likely to reduce early onset cases, and thus explained

the trends identified in the CDC surveillance. These surveys also illustrate how tremendous resources may be invested in health services--diagnostic, preventive, or therapeutic--to influence clinical outcomes, but without careful coordination between laboratory practitioners and clinical providers, resources may be wasted and morbidity left unabated.

Economic Considerations

Today, cost considerations exert a strong influence on clinical and laboratory practices. In the Georgia survey, one of the most common reasons cited by clinicians who did not screen their patients was the belief that prenatal screening was not cost-effective. Laboratories considering the use of selective broth media over nonselective methods for GBS isolation may similarly question whether the more elaborate media are economically justifiable. Participants in specific components of the health care system are less likely to consider the societal perspective but focus instead on their own bottom line. To evaluate the economic impact of routine prenatal screening for GBS and antibiotic treatment of high risk mothers, CDC[7] and others have conducted economic analyses which assess the costs of a screening and prophylaxis program vs. the costs of treating cases that could otherwise be prevented. Consistently, these studies indicate that prevention programs--including those incorporating prenatal screening cultures--save money compared with treating GBS cases that would occur without these interventions. Because laboratory costs are often borne by the hospital, while costs of caring for the acute illness and chronic sequelae of neonatal infection are borne by numerous parties, assessing the economic impact of prevention programs ideally should incorporate the societal perspective.

Public Health Response

To address the specific problems identified in clinical practice and promote effective prevention programs, CDC prepared guidelines for prevention of GBS disease. To develop the guidelines, CDC solicited substantial input from outside experts and published a draft version of the guidelines for public comment in the Federal Register. The CDC guidelines addressed specific concerns identified by the practice surveys, including cost-effectiveness.[17] The guidelines also stressed the importance of appropriate culture methods. We received thousands of letters in response to the draft guidelines from families, clinicians, professional organizations, and others. The issue also has entered the political arena, since several state legislatures have considered bills that would mandate prenatal education or prenatal screening for GBS. In 1994, the California legislature, after reviewing one such proposal, passed a bill requiring the California health department to hold a consensus conference on the topic. CDC convened the meeting with the California Department of Health Services, and numerous organizations participated, including a community-based parent advocacy group, the Group B Strep Association. Revised recommendations are the product of this process, and the laboratory practice research on this issue played a major role in developing the new recommendations.

Conclusions

Specific recommendations for GBS prevention seek to enhance the effectiveness of prevention activities, a critical component of which is ensuring participation by supporting laboratories in designing and implementing prevention programs. As I

have tried to indicate, clinical outcomes depend on a mutual understanding by clinicians, laboratory personnel, and patient groups, of the goals of a testing program, interpretation of test results, the action plan necessitated by test results, and limitations of the program when optimally implemented. Communication between clinicians and laboratory personnel is critical even if clinicians collect the appropriate specimens and the laboratory processes specimens in the optimal fashion, clinical outcomes depend on information being available to providers at the time and place of delivery. As clinicians become more dependent on offsite laboratory services, perhaps in the context of expanding managed care, assuring communication systems for the prompt reporting of laboratory results to multiple facilities will be critical.

To monitor program performance, an ongoing commitment is needed to conduct surveillance for clinical outcomes of substantial importance (in this example, for cases of perinatal GBS disease). Because most hospitals are too small for meaningful trends in invasive GBS disease to be detected, surveillance will ideally involve larger populations. Managed care organizations and other groups of affiliated hospitals should provide ideal circumstances to monitor program effectiveness through laboratory-based surveillance systems.

References

1. McCracken GH. Group B streptococci: the new challenge in neonatal infections. *J Pediatr*. 1973;82:703-6.

2. Zangwill KM, Schuchat A, Wenger JD. Group B streptococcal disease in the United States, 1990: report from a multi-state active surveillance system. In: CDC Surveillance Summaries. v. 41: MMWR, 1992; No. SS-6:25-32 (November 20, 1992).

3. Baker CJ, Barrett FF, Gordon RC, Yow MD. Suppurative meningitis due to streptococci of Lancefield group B: a study of 33 infants. *J Pediatr.* 1973;82:724-9.

4. Baker CJ, Edwards MS. Group B streptococcal infections. In: Remington J, Klein JO, eds. Infectious diseases of the fetus and newborn infant. 4th ed. Philadelphia: W.B. Saunders, 1995:980-1054.

5. Pass MA, Gray BM, Dillon HC. Puerperal and perinatal infections with group B streptococci. *Am J Obstet Gynecol.* 1982;143:147-52.

6. Yancey MK, Duff P, Clark P, Kurtzer T, Frentzen BH, Kubilis P. Peripartum infection associated with vaginal group B streptococcal colonization. *Obstet Gynecol.* 1994;84:816-9.

7. Mohle-Boetani J, Schuchat A, Plikaytis BD, Smith D, Broome CV. Comparison of prevention strategies for neonatal group B streptococcal infection: a population-based economic analysis. *JAMA*. 1993;270:1442-8.

8. Boyer KM, Gadzala CA, Kelly PD, Burd LI, Gotoff SP. Selective intrapartum chemoprophylaxis of neonatal group B streptococcal

early-onset disease. II. Predictive value of prenatal cultures. *J Infect Dis.* 1983;148:802-9.

9. Boyer KM, Gotoff SP. Strategies for chemoprophylaxis of GBS early-onset infections. *Antibiot Chemother.* 1985;35:267-80.

10. Garland SM, Fliegner JR. Group B streptococcus and neonatal infections: the case for intrapartum chemoprophylaxis. *Aust N Z J Obstet Gynaecol.* 1991;31:119-22.

11. Hankins GV, Chalas E. Group B streptococcal infections in pregnancy: ACOG's recommendations. *ACOG Newsletter.* 1993;37:2.

12. American Academy of Pediatrics. Guidelines for prevention of group B streptococcal infection by chemoprophylaxis. *Pediatr.* 1992;90:775-8.

13. American College of Obstetricians and Gynecologists. Group B streptococcal infections in pregnancy. Washington, D.C.: ACOG, 1992 (ACOG Technical Bulletin; No. 170, July 1992).

14. Larsen JW, Dooley SL. Group B streptococcal infections: an obstetrical viewpoint. *Pediatr.* 1993;91:148-9.

15. Gaffer HS, Schuchat A, Hailstone R, Whitney C, Toomey K, Wenger J. Barriers to prevention of perinatal group B streptococcal disease. *Pediatr Infect Dis J.* 1995;14:662-7.

16. Whitney CG, Schwartz B, Deaver K, et al. Institutionalizing prevention of perinatal group B streptococcal infections. In: Abstracts of the 35th Interscience Conference on Antimicrobial Agents and Chemotherapy. San Francisco: ASM, 1995:322.

17. Centers for Disease Control and Prevention. Prevention of perinatal group B streptococcal disease: a public health perspective. MMWR 1996;45(No. RR-7):1-24.

The Development of Measures of Outcomes from a Clinical Laboratory Database

R. Peter Mallon, Ph.D.
Director, Medical Applications
Corning Medical Informatics
Rutherford, New Jersey

Abstract: Dramatic changes are rapidly occurring in the health care industry. One force driving change involves the fact that the entity paying for health care services increasingly is not the direct user of those services. However, this new payer has needs beyond the requirements associated with routine delivery of services. The need to be able to measure provider performance and assess the value of those services is among the new payer needs. Measuring the outcomes of clinical practice is increasingly becoming a measure of the value and quality of the health care service.

Clinical outcomes involve linkages. Quality and value involve giving the right care to the right patient at the right time. Outcomes research involves identification of what works best for whom.

This paper will present definitions of outcomes research and analysis. It will also discuss the change in paradigm as laboratory medicine moves from systems of transactional delivery of service to managing the wellness of a population. The paper will conclude with two specific examples of outcomes measures derived from a laboratory data repository. These examples will involve identifying the best practice benchmarks and establishing measures of preventive health.

This paper presents two examples of outcomes measures that have been derived from clinical laboratory databases. One example is cited from the literature, while the other has been developed from a large relational database of clinical laboratory test results and patient demographic data. A brief discussion of outcomes is also presented.

As a background, clinical outcomes involve linkages. Specific conditions are linked to specific types of care for identifying what works best for whom. Outcomes have become important because delivery of health care will increasingly be measured upon value. Outcomes are the measure of value.

The health care delivery system is experiencing new challenges. We are challenged not only to provide the best care and treatment in response to illness; we are now challenged to more systematically prevent illness. Thus, the infrastructure of the health care delivery system is being transformed. Current health care systems have evolved so as to manage and respond to the consequences of disease prevalence within a population. As the dynamics of the health care reorganization progress, the system is being challenged to not only to respond to, but also reduce, the rate of illness within a population. Health care is being challenged to better manage wellness. Outcomes will be the engine as health care moves from transactional medicine to managing wellness.

The goals of outcomes research are to identify variations in patterns of care; evaluate costs of care; make decisions about resource allocation; assess quality of care. The tools of outcomes research are the tools

of total quality management:

- Observation
- Measurement
- Diagnostic Journey
- Remedial Journey
- Goals

Outcomes are measured as rates. An outcome always has a numerator and a denominator. Outcome rates must be constructed from explicit criteria that apply to specific conditions. Apples must be compared with apples to make meaningful decisions about benchmarks.

Two examples of laboratory based outcomes as contributions to outcomes initiatives are presented below:

- Analytical process
- Clinical Process

Outcomes Measure of an Analytical Process

This outcomes measure of an analytical process is taken form the Durand-Zaleski et al. (1993) article titled "Outcomes study of ordering patterns for tumor marker tests." The focus of their study was to establish measures in order to better manage the unnecessary use of laboratory tests, specifically, tumor marker tests. These investigators established measures that indicated:

- Requests for tumor markers accounted for 50% of all immunoassays
- Twenty-nine percent (50 / 170) physicians accounted for 80% of requests
- Laboratory test requisition was typical of most order forms in that test names were merely listed
- A separate booklet provided semi-quantitative information concerning the clinical relevance of the assay
- Survey indicated that 50% of request lacked clinical justification

Researchers set out upon a diagnostic journey to find the root cause of the various measures of variation. They made some critical assumptions in that they recognized that physicians do not deliberately order the wrong test. Physicians, moreover, like most workers, do want to do the correct thing. Physicians, also as do most individuals, require real time process control feedback mechanisms in order to assuredly accomplish what is expected.

Analysis indicated that critical point to manage better feedback mechanism was at the time of the test request. Providing feedback at the time of the laboratory report was ineffective in changing ordering behavior. To accomplish this, the lab test requisition was modified. The requisition was changed from a standard form listing tests available to a matrix. In this matrix the target organ of concern was listed as column headings and the specific tumor marker of choice was listed as the row labels. This format formed a grid of squares. The boxes that matched the appropriate assay for the specific organ of monitoring remained clear. Boxes that associated inappropriate testing request were "blacked out" thus not permitting the inappropriate test for a particular organ of interest (Table 1).

What tools were used?

- Observation
- Measurement Pareto Analysis
- Assignable Cause Identification
- Fool-proof the system
- Measure the outcome by Cost of Poor Quality (COPQ)

OUTCOME		
	BEFORE	**AFTER**
Test Requests/Requisition	**2.5**	**1.9**
Max	**11**	**7**
Min	**2**	**1**
SD	**2**	**2**
COPQ		***$50,000/YR***

Table 1. Effect of changing requisition form on test ordering.

Specimen	Date	Ratio Result	Glycohgb Result	Sex	Age, Years
123456	1/1/96	7.3	10.9	M	45
234567	½/96	6.3	8.4	M	53
345678	1/3/96	6.4	4.9	F	71
456789	1/4/96	7.1	7.6	F	43
567890	1/5/96	8.4	10.9	M	
678901	1/6/96	5.9	8.9	F	22
789012	1/7/96	6.0	6.9	F	29
890123	1/8/96	4.4	7.1	F	65
901234	1/9/96	5.9	9.1	M	22

Table 2. Laboratory results aggregated within a population.

Outcomes Measurement of a Clinical Process

As discussed previously, transactional medicine is evolving to management of wellness. Wellness management involves the delivery of health care management not only to individuals in response to progressed illness, management of wellness very simply involves scheduling. Successful management of wellness of a population involves an integrated health care delivery system that appropriately screens individuals for the early detection of disease, effectively diagnoses illness and then effectively monitors therapy. The following demonstrates the evolution from transactional care to wellness management:

A lab request frequently initiates the process:

From: Dr Jones
To: The lab
Re: Do These Tests
Date: N
HDL, Glycohgb, Lyme, ANA

A lab report is the response:

From:	The lab
To:	Dr Jones
Re:	Specimen 123456
Date:	N + 1
HDL	7.3
Glycohgb	10.9
Lyme	Neg
ANA	Neg
Other	

But within the context of a population individual lab results may be aggregated in a table (Table 2).

The data in a table (Tables 3 and 4) can provide more meaningful information as a distribution of results. An individual result is assessed in context to its relative position within a common population's distribution of results. What is the frequency of results within the risk levels of this measure of disease?

Then the data can then be viewed in the context of a relationship of related results (Table 5).

A set of patterns can then become apparent for the population (Table 6).

The pattern evolves to identify the proportion of individuals within the population that have the poorest measure of glycemic control and the highest risk for heart disease (Table 7 and 8).

A trend is suggested. How do individuals progress into the upper right quadrant of highest risk. Individuals for the most part are not born with these measures suggesting high risk. Individuals progress towards these high risk conditions over a lifetime. What is the direction of this progression? Is it clockwise, or in a counter clockwise direction, or is it a straight-line path? Is there more than one mechanism (Table 9)?

In summary, outcomes measures are derived from data, and these data are available from many lab sources. The discipline of outcomes will become an important management tool of the future. The tools are changing. Increasingly an important tool of health care will be the tool used for the analysis of populations of data, i.e., the personal computer.

Selected References

1. United Health Care Corporation. The Language of Managed Health Care and Organized Health Care Systems. United Health Care Corporation, Opus Center, Minnetonka, MN.

2. Guadagnoli E, McNeil BJ. Outcomes research: Hope for the future or the latest rage? *Inquiry*. 1994;31:14-24.

3. Epstein AM. The outcomes movement - Will it get us where we want to go? *N Engl J of Med.* 1990; 323:266-270.

4. James, BC. Quality improvement in the hospital: Managing clinical processes. *The Internist*. 1993; 34:11-17.

5. Laffel G, Blumenthal D. The case for using industrial quality management science in health care organizations. *JAMA*. 1989;262:2869-2873.

6. Juran J, Gryna FM. Juran's quality control handbook. 4th edition. McGraw-Hill Book Company. New York, NY. 1988.

7. National Committee for Quality

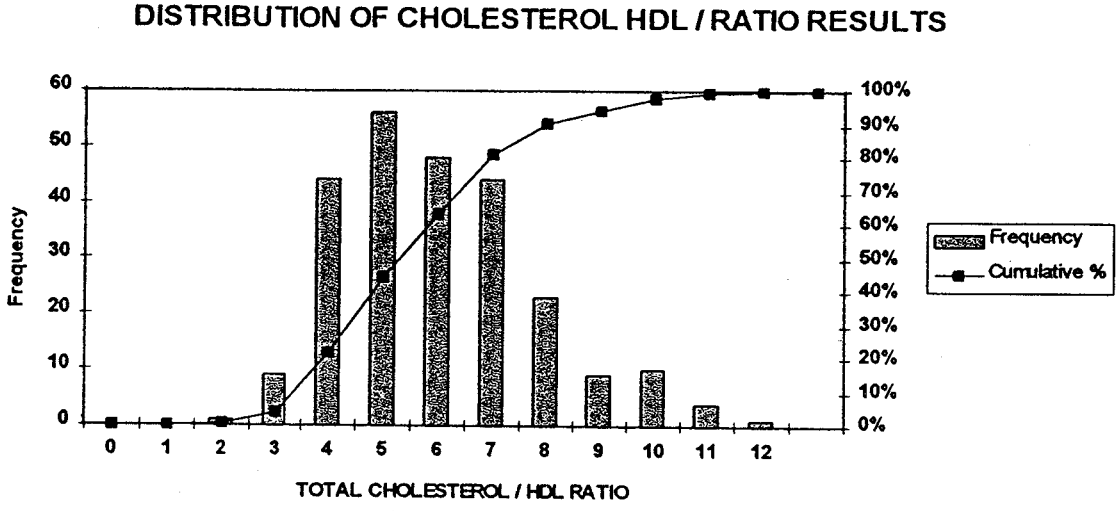

Table 3. Distribution of Cholesterol/HDL Ratio results.

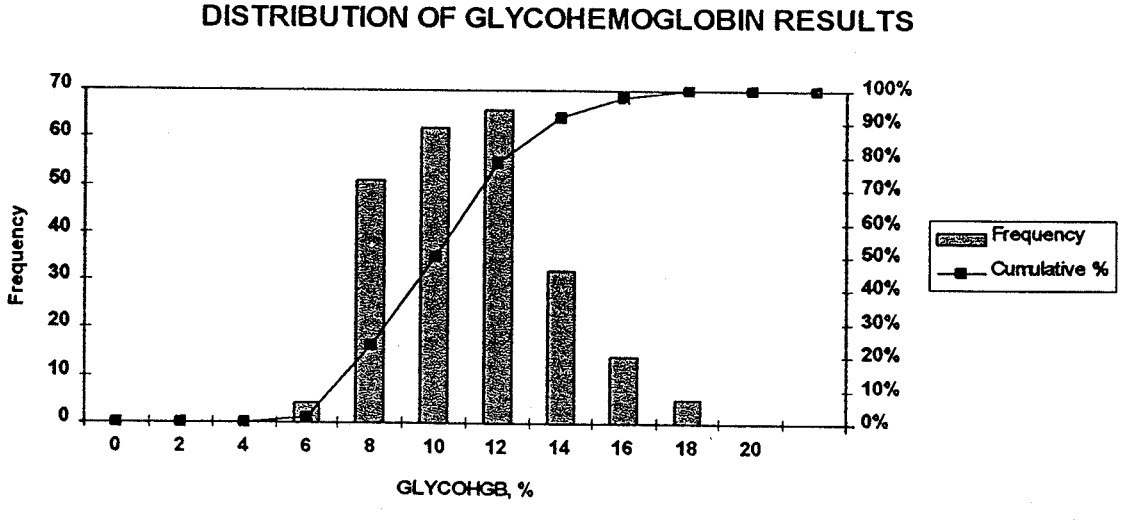

Table 4. Distribution of glycohemoglobin results.

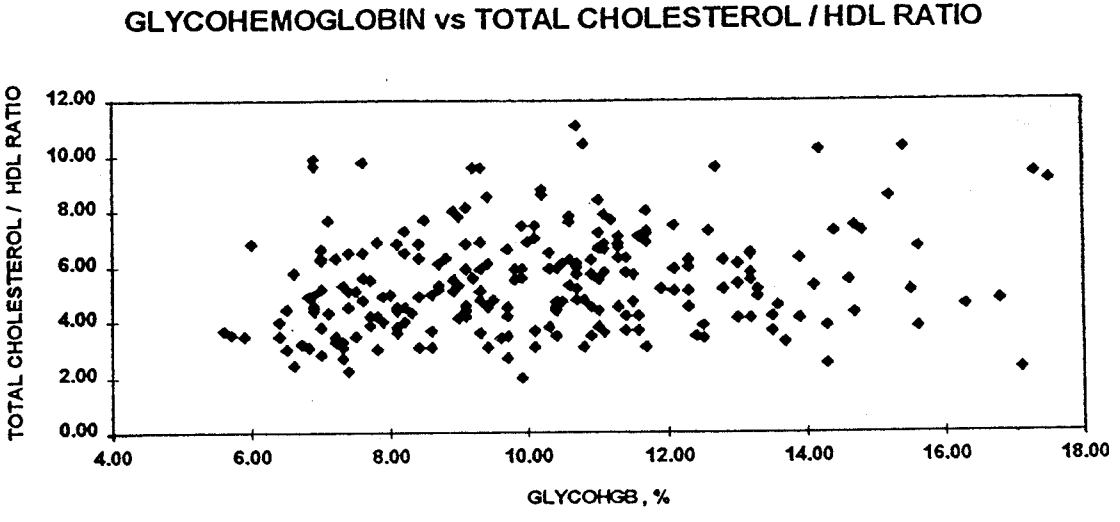

Table 5. Glycohemoglobin vs Total Cholesterol/HDL Ratio

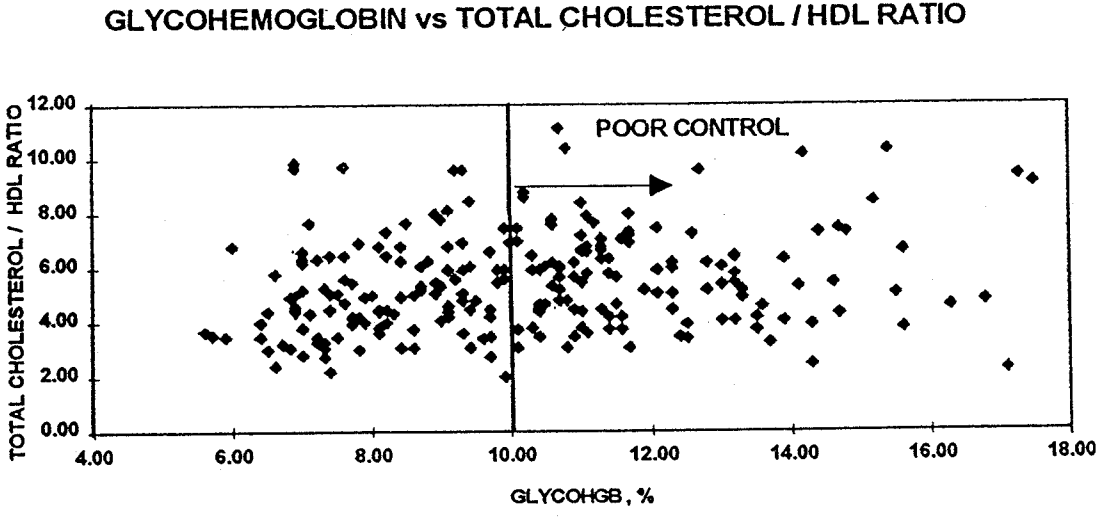

Table 6. Population in poor control.

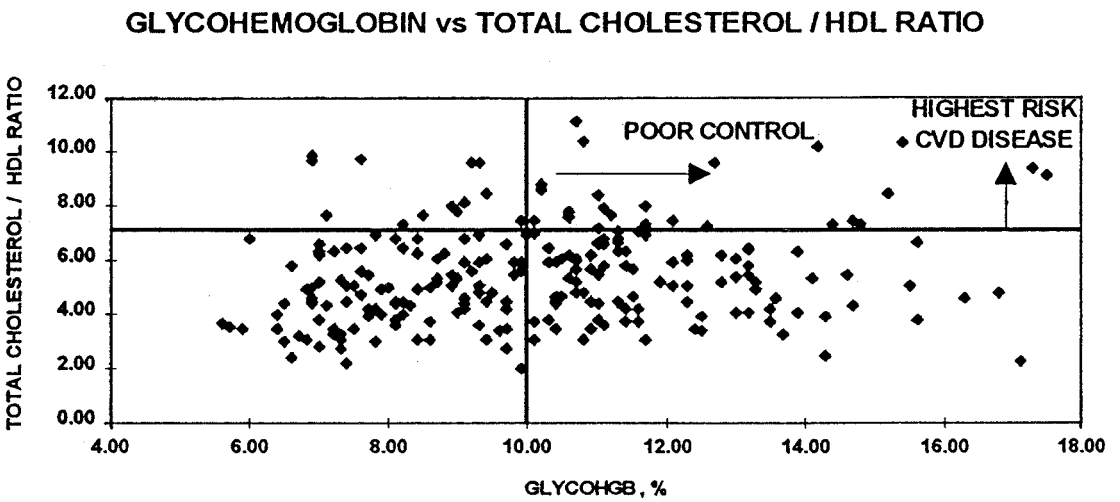

Table 7. Population at highest risk for heart disease.

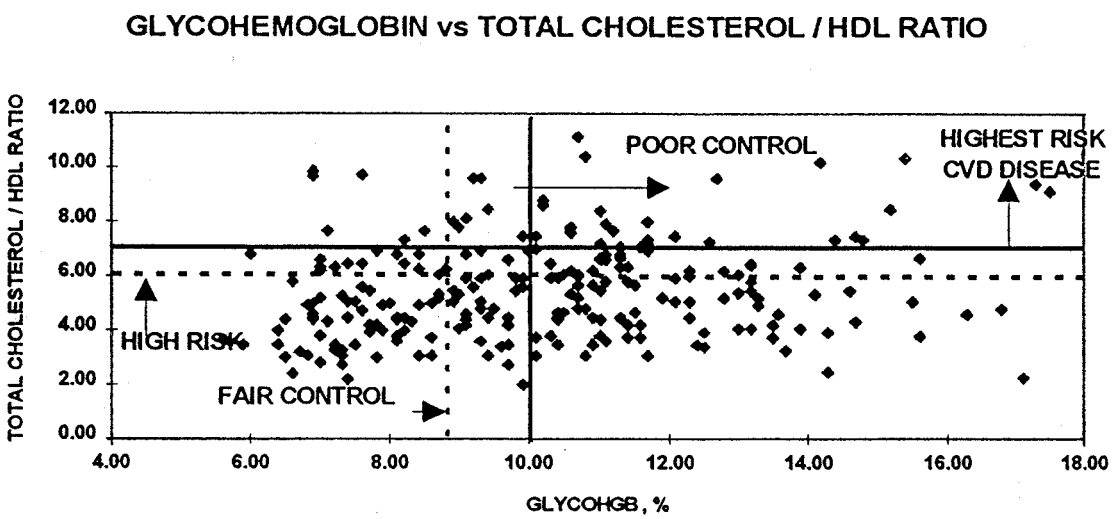

Table 8. Population in fair control.

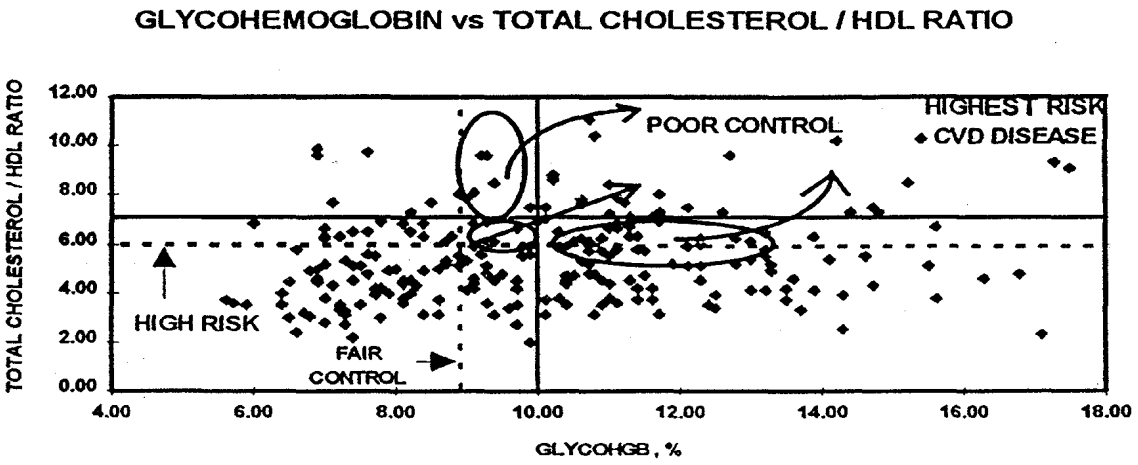

Table 9. Possible directions toward high risk of CVD over life time.

Assurance. HEDIS: Health plan employer data and information set and users manual. Version 2.5: Updated specifications for HEDIS 2.0. National Committee for Quality Assurance, Washington DC.1993.

8. Lawthers AG, Palmer RH, Edwards JE, Fowles J, Garnick DW, Weiner JP. Developing and evaluating performance measures for ambulatory care quality: A preliminary report of the DEMPAQ project. *Journal on Quality Improvement.* 1993;19:552-565.

9. Berwick D. Continuous improvement as an ideal in health care. *New Eng J of Med* . 1989;320:53-56.

10. Durand/Zaleski I, Rymer JC, Roudot-Thoraval F, Revuz J, Rosa J. Reducing unnecessary laboratory use with new test request form: Example of tumour markers. *The Lancet.* 1993;342:150-152.

11. Garnick DW, Fowles W, Lawthers AG, Weiner JP, Parente ST, Palmer RH. Focus on quality: Profiling physicians' practice patterns. *Journal of Ambulatory Care Management.* 1994;17:44-75.

12. Sloane MD. How to lower health care costs by improving health care quality. Results-based continuous quality improvment. ACQC Press. 1994.

Suggested Texts and Journals

1. "The 1993-1994 Medical Outcomes & Guidelines Source book. A Progress Report and Resource Guide on Medical Outcomes Research and Practice Guidelines: Developments, Data and Documentation." Published by Faulkner and Gray (1993).

2. "Report on Medical Guidelines & Outcomes Research" published by Capitol Publications Inc., 1101 King St., Suite 444, Alexandria, Virginia, 22314-2968,(800)655-5597.

3. "Outcomes Measurement and Management" published by The Zitter Group, 90 New Montgomery Street, Suite 820, San Francisco, California, 94105, (415)495-245.

4. "Medical Outcomes & Guidelines Alert" published by Faulkner and Gray Health care Information Center, 1133 Fifteenth Street, NW, Suite 450, Washington, DC, 20005 (215)967-7060.

Laboratory Tests for Case-Finding in Ambulatory Care: Application of a Conceptual Framework

Marc D. Silverstein, M.D.
Division of Area General Internal Medicine and Section of Clinical Epidemiology
Mayo Clinic
Rochester, Minnesota

Abstract: Recent efforts to reduce health care costs have focused attention on laboratory tests, and methods for evaluating laboratory tests for case-finding are needed. Diagnostic yield (proportion of patients with laboratory tests that resulted in a new diagnosis), therapeutic yield (proportion of patients with laboratory tests that resulted in a new therapy), cost per new diagnosis, and cost per new therapy were adapted from a published conceptual framework for evaluating laboratory tests to evaluate frequently used laboratory tests for case-finding in the ambulatory setting. The therapeutic yield of routine tests in ambulatory patients is approximately 15% for lipid tests, 3% for the chemistry profile, and less than 1% for the Complete Blood Count (CBCs), sensitive Thyroid Stimulating Hormone (TSH), and urinalysis. The cost per new diagnosis associated with treatment is approximately $100 for lipids, $300-$1100 for CBC, chemistry profile and urinalysis, and $2000 for the sensitive TSH. The application of this framework and information on the diagnostic yield, therapeutic yield, cost per new diagnosis, and cost per new therapy can be used to make informed decisions about the use of laboratory tests for case-finding in ambulatory adults in a managed care setting.

Introduction

Health care reform discussions center around the increasingly high costs of medical care. The importance of evaluating the health services provided to patients as well as the outcomes they experience is widely recognized, as payers try to reduce costs and providers strive to make informed choices about health care resources used to maintain quality of care for their patients.

Efforts to reduce costs will not spare any cost center in the new health care environment. Laboratory tests are a natural focus of efforts to reduce costs. Although the cost of individual tests is low, in aggregate substantial health care resources are devoted to laboratory tests. A likely early focus of efforts to reduce the use and therefore costs of laboratory tests is the ambulatory setting, because many of those tests are for "screening" or "case-finding" and result in negative or normal values in most patients.

Elsewhere the author has presented a conceptual framework for evaluating laboratory tests for case-finding in ambulatory patients [1]. The focus of this paper is on how to use the diagnostic yield, therapeutic yield, and costs of laboratory tests to make informed decisions about resource use in the ambulatory setting. The "diagnostic yield" is the proportion of patients tested who have a new diagnosis, and the "therapeutic yield" is the proportion of patients who have a new diagnosis associated with a change in their therapy, where change may mean initiating a new therapy, modifying an existing therapy, or stopping a therapy.

Ambulatory Testing

Tests are obtained in the ambulatory setting to search for early disease in asymptomatic patients, to evaluate presenting problems in symptomatic patients, to monitor patient responses to therapy, to watch for adverse effects of medications and other treatments, and to obtain information about the severity of disease and prognosis. The evaluation of laboratory tests in the ambulatory setting must take into account the purpose for which a test is ordered, since different criteria may be used in evaluating a test that was obtained as part of a diagnostic evaluation or for monitoring therapy. This paper focuses on tests obtained in ambulatory patients to search for a condition or disease that is not related to the reason for the patient's visit, that is, "case-finding."

"Case-finding" and "screening" often are used interchangeably, but the two terms have different meanings and implications. Ideally, "case-finding" should be reserved to describe the test ordering by the physician in the office setting to identify conditions in asymptomatic patients and "screening" should be reserved to describe programs to test the general population in other settings. Clinicians should be aware of criteria that are used to evaluate screening procedures and tests in populations, because these same criteria may be applied to evaluate the use of laboratory tests obtained for case-finding in the ambulatory setting.

Several years ago, the World Health Organization (WHO) formalized six basic criteria for screening [2], which have been reformulated and restated in many different ways but are still widely accepted. Screening is indicated when the condition, test, and treatment fulfill the following criteria: (1) the condition for which the screening test is performed should have a significant effect on the quality or quantity of life; (2) the condition should have an asymptomatic period during which it can be detected; (3) the incidence of the condition should be sufficiently high to justify the costs of screening; (4) an acceptable test should be available; (5) acceptable methods of treatment should be available; and (6) treatment when the patient is asymptomatic should yield superior results to treatment when the condition would become symptomatic.

Randomized controlled trials of multiphasic screening have produced information to test these criteria. Perhaps the best known study of routine tests for case-finding is the Kaiser Permanente study of the multiphasic health checkup, which used a randomized design to compare a comprehensive, multiphasic checkup to usual care. The study compared the outcomes of 5156 plan participants who were urged to have yearly multiphasic health checkups with the outcomes of a control group of 5557 who had access to the multiphasic checkup but were not actively encouraged to complete a yearly checkup. The cost of this multiphasic health checkup was approximately 10% of the annual per capita health care costs in 1967 and 1968. Follow-up reports of the outcomes of this trial were published in 1978 [3] and 1986 [4]. There were no differences in all-cause mortality or disability in the group that received yearly multiphasic examinations compared to the usual care group. There was 30% reduction in "postponable" causes of death, predominantly those due to colorectal cancer, hypertension, hypertensive heart disease, and stroke. No difference in any outcomes was noted that could be attributed to a routine laboratory test.

Another randomized controlled trial of

multiphasic screening was performed in 574 families from three socioeconomic groups in the Salt Lake City area in the 1970's. The use of health services, morbidity, health status, and attitudes were measured over one year. There were no significant differences in morbidity or attitudes; an increase in hospitalization was observed in the multiphasic health checkup group compared with the usual care group[5].

Although routine laboratory tests can identify early manifestations of several conditions, the WHO criteria for screening are not met for most conditions that could be detected by routine laboratory tests in the ambulatory setting. The Canadian Task Force on the Periodic Health Examination, the American College of Physicians, and the U.S. Preventive Services Task Force have used similar methods to review available evidence and make recommendations about the use of the CBC, chemistry profile, lipid tests, sensitive TSH, and urinalysis as case-finding tests in ambulatory adults. The table summarizes the recommendations from these expert panels (Table 1).

Perhaps the most striking fact is the general agreement that routine CBC, chemistry panel, and thyroid disease screening tests are *not* routinely indicated in the otherwise generally health adult. Also of note is the recommendation that the interval for obtaining a serum cholesterol level should be every 5 years. The expert panels differ in their recommendations for routine urinalysis for case finding. The urinalysis has been recommended for pregnant women and in patients with diabetes by the U.S. Preventive Services Task Force, which also has stated that it may be "prudent" to obtain a urinalysis in preschool children and in older adults.

These recommendations are important because they have implications for laboratories. The data from a Mayo Clinic prospective study of 531 ambulatory adults suggests that if these or similar recommendations were followed, the number of laboratory tests ordered would be substantially reduced. The reduction would be 65% for CBC, 57% for chemistry profile, 50% for lipids, 65% for sensitive TSH, and 38% for urinalysis. Clearly, if these guidelines were used in a managed care environment for routine case-finding in ambulatory adults, there would be significant impact on laboratory test volumes[6].

Although these expert panels have recommended against using routine laboratory tests for case-finding in otherwise healthy adults, patient expectations for the tests remain high, and physicians still frequently obtain these tests in the ambulatory setting. As insurers raise barriers to use of tests and treatments to try to reduce health care costs, it is imperative that clinicians and laboratories make informed decisions about test use. Obstacles exist that make this process difficult. The condition, test, or treatment may not meet WHO criteria for screening, and the test may be judged as unnecessary or inappropriate. The needed information may not be available: Patient outcomes may occur days, weeks, or even years after the laboratory test has been obtained and treatment initiated. Laboratory tests alone may not be sufficient to produce good patient outcomes. Laboratories have no control over the treatment decisions, quality of treatment provided, or patient compliance with provider orders. Therefore, the relationship between the diagnostic test and the therapeutic outcome must be evaluated.

Test	American College of Physicians	Canadian Task Force on the Periodic Health Examination	US Preventive Services Task Force
CBC	Not Recommended	Not Recommended	Not Recommended
Chemistry	Not Recommended	Not Recommended	Not Recommended
Lipids	Males & females, age 18 or older, every 5 years	Males ages 30-59, every 5 years	Males & females, age 18 or older, at least every 5 years
Thyroid	Not Recommended	Not Recommended	Not Recommended
Urinalysis	Pregnancy	Not Recommended	Pregnancy, Diabetes Mellitus & possibly children < age 5 & adults >age 60

Table 1. Recommendations of Expert Panels for Laboratory Tests for Case-Finding in Ambulatory Patients

Diagnostic Yield and Therapeutic Yield

Recent studies from Basel, Switzerland, evaluated the use of routine CBC and chemistry for case-finding. The CBC was studied in a university clinic in an unselected cohort of 595 adults, predominantly young men (mean age 40; 62% male). The CBC was analyzed as consisting of four components (hemoglobin, mean cell volume, leukocyte count, and platelets). 65% of tests were obtained for screening or case-finding. Abnormal values were noted in 6%. Five patients had new diagnoses as a result of the laboratory test, and three patients were treated. Thus, the diagnostic and therapeutic yield of the CBC in this population was less than 1% [7].

The routine chemistry profile was evaluated in an unselected cohort of 493 adults (likely a subset of the same patients in the previous study). The chemistry profile consisted of 23 individual tests. 89% of the chemistry profiles were obtained for screening or case-finding. Abnormal results were noted in 11% of patients tested. 35 new diagnoses were made, 25 of which were associated with therapy. The diagnostic yield was approximately 7 % and the therapeutic yield was approximately 5%, with most of the yield due to lipid tests [8].

A Mayo Clinic retrospective study of 100 adult patients from Olmsted County or the surrounding area who had a comprehensive general medical evaluation reported the diagnostic yield and therapeutic yield of the CBC, chemistry profile, a thyroid test (usually sensitive TSH, but less often thyroxine), and urinalysis. The patients' mean age was 59 years, and approximately 60% were females. The patients had an average of 2.3 serious medical conditions, and the mean interval between the current

examination and their last comprehensive medical evaluation was just under 3 years. Approximately 70-90% of patients had routine tests for case-finding, and the majority of test results, as expected, were normal. The diagnostic yield of the lipid tests was greatest (12.3%), followed by thyroid (2.8%), CBC and chemistry (2.2% each), and urinalysis (1.1%). The highest therapeutic yield also came from the serum lipid tests, 9.2%. The therapeutic yield of the chemistry profile was 2.2%, and the therapeutic yield of the CBC, thyroid disease screening test, and the urinalysis were all less than 2% [9]. Similar results were found in a Mayo Clinic prospective study of 531 patients, although the contrast was greater: Therapeutic yield of lipid tests was 16.5%, chemistry profile 2.8%, and the others less than 1% [6].

Cost Per New Diagnosis and Cost Per New Therapy

In a managed care environment, profit centers such as laboratories become cost centers. Providers of care are responsible for the costs of providing services for the patients in their care. Costs of routine laboratory tests for case-finding were estimated using the methodology recommended by the College of American Pathologists and standard accounting practices and estimates of diagnostic yield and therapeutic yield in ambulatory adults. The cost per new diagnosis associated with treatment is approximately $100 for lipid tests, $300-1100 for CBC, chemistry profile, and urinalysis, and $2000 for sensitive TSH.

Charges to payers are, of course, higher than laboratories' costs of performing the tests. When estimates of the actual medical charges are used, charges (in a fee for service system) per new diagnosis and per new therapy from the use of these laboratory tests for case-finding in ambulatory adults can be calculated. The charges per new diagnosis associated with initiating therapy are lowest for serum lipid tests at approximately $325; between approximately $2000 and $3000 for the CBC, chemistry profile, and urinalysis; and almost $11,000 for the sensitive TSH. Within managed care systems or other integrated health systems, a corresponding relationship of costs per new diagnosis associated with initiating therapy can be assumed.

Unfortunately, data are not directly available on patients' outcomes resulting from treatment initiated as a result of using the laboratory test for case-finding. Until such data are available, decisions about the use of laboratory tests for case-finding in ambulatory patients must be based on information on diagnostic yield and therapeutic yield, as defined in this paper.

Conclusions

More than half of "routine" laboratory tests in the ambulatory setting are now obtained for case-finding. Decisions about the use of laboratory tests for case-finding in ambulatory patients reasonably could be based on costs per new diagnosis or cost per new therapy. The therapeutic yield of routine tests in ambulatory patients is approximately 15% for lipid tests, 3% for chemistry profile, and less than 1% for CBCs, sTSH, and urinalysis. Actual costs per new diagnosis or new diagnosis associated with treatment are modest in light of other health care costs ($100-$2000 per new diagnosis treated). More information on patient outcomes resulting from laboratory tests for case-finding is needed to improve the decision-making process. Especially needed is information on

outcomes of case-finding for hypercholesterolemia, diabetes mellitus, and other conditions for which case-finding is potentially indicated. Laboratorians and clinicians must work together to control costs and to measure outcomes to ensure that we are actually achieving the benefits of effective treatments that are initiated for conditions detected by the use laboratory tests for case-finding.

References

1. Silverstein MD, Boland BJ. Conceptual framework for evaluating laboratory tests: case-finding in ambulatory patients. [Review]. *Clinical Chemistry* 1994;40:1621-1627.

2. Wilson JMG, Jungner G. *Principles and practice of screening for disease*, Geneva, World Health Organization; 1968:1-163.

3. Dales LG, Friedman GD, Collen MF. Evaluating periodic multiphasic health checkups: a controlled trial. *Journal of Chronic Diseases* 1979;32:385-404.

4. Friedman GD, Collen MF, Fireman BH. Multiphasic Health Checkup Evaluation: a 16-year follow-up. *J Chron Dis.* 1986;39:453-463.

5. Olsen DM, Kane RL, Proctor PH. A controlled trial of multiphasic screening. *N Engl J Med.* 1976;294:925-930.

6. Boland BJ, Wollan PC, Silverstein MD. Value of Case-Finding Laboratory Tests in the Comprehensive Medical Examination. *J Gen Intern Med.* 1995;10:S39.(Abstract)

7. Ruttimann S, Clemencon D, Dubach UC. Usefulness of complete blood counts as a case-finding tool in medical outpatients [see comments]. *Ann Intern Med.* 1992;116:44-50.

8. Ruttimann S, Dreifuss M, Clemencon D, di Gallo A, Dubach UC: Multiple biochemical blood testing as a case-finding tool in ambulatory medical patients. *Amer J of Med.* 1993;94:141-148.

9. Boland BJ, Wollan PC, Silverstein MD. Review of systems, physical examination, and routine tests for case-finding in ambulatory patients. *Amer J Med Sci.* 1995;309:194-200.

Summary of Workshop 4: Detection of Problems Affecting Patient Outcome

Facilitator: Robert Kisabeth, M.D.
Medical Director
Mayo Medical Laboratories
Rochester, Minnesota

CDC Liaison: Richard A. Keenlyside, M.D.

Key Questions:

1) Are methods available to detect problems, to measure their impact on patient care, and to provide feedback for preventing future occurrences?
2) What new methods are needed and how could they be implemented?

The participants in this workshop discussed approaches to identifying and investigating problems in laboratory practice as they affect patient outcomes. The questions posed to participants to frame the discussion were:

"*How can outcomes be measured*?
What methods are useful in detecting laboratory problems affecting patient outcome ?
How can we better communicate our findings to our colleagues? "

Priority areas for study

It was clear that there are many possible approaches for measuring the relationship between laboratory testing and patient outcomes. The group suggested that, to begin with, research should focus on medical conditions that require high volume testing, expensive tests, highly prevalent conditions with significant morbidity, and those where intervention may have some impact. Studies should also take advantage of existing information systems more than in the past. Soundly based multi-center epidemiological studies were encouraged. In addition, some established provider research networks are ideally suited to studies of laboratory tests and testing and collaboration with these was encouraged.

Outcomes measurement

The challenge of laboratory medicine research is to develop robust and meaningful measures that monitor the consequences of testing. These may be health outcomes (patient benefit) or system-based process measures (organizational, cost/utility etc.). Health outcomes are determined through clinical review, especially chart review, but process measures in laboratory medicine can be used as surrogate measures in lieu of these.

Critical pathway analysis is now increasingly used to evaluate the effectiveness of clinical care and the performance of health systems. It provides a context and a controlled environment to study the impact of various interventions. This approach has great potential for studying the role of laboratory practice in clinical care.

Data sources Reliable and useful data for

studying laboratory practice are commonly highly fragmented, inaccessible and expensive to obtain. Most available data are not accumulated over time and not linked to other outcomes of interest such as health care utilization. Uniform, standardized datasets that combine laboratory testing data, clinical information and health care utilization measures (collected prospectively) are needed. To achieve this, close collaboration must exist between the research and health care communities to identify a core data set to be accumulated prospectively from health provider networks. New databases such as The Health Care Employer Data and Information Set (HEDIS) that are now used by many managed care organizations are useful models for this.

Communication with colleagues

In the future, laboratorians will be asked to participate in collaborative research with others to demonstrate the diagnostic and therapeutic effectiveness of our clinical testing. This will involve collaboration between laboratory professionals and clinical colleagues as well as payers and management professionals. To be effective in this role, laboratory professionals will therefore need better training in research design, epidemiology and biostatistics.

Timely communication with our clinical colleagues about all aspects of laboratory testing is an essential part of quality care. This extends from the rationale for ordering tests to the interpretation of results. In support of this, improved laboratory decision support systems are needed to identify problems that may affect clinical decisions such as systematic analytic bias. Modeling and bench-marking would also be helpful in providing feedback to colleagues.

Finally, the group believed that research questions and study results should be more widely disseminated beyond the conventional pathology journals in publications read by clinicians of all specialties and health care managers.