Practical Considerations in the Measurement of Outcomes in Healthcare

M.A. Krousel-Wood, MD, MSPH, FACPM
Alton Ochsner Medical Foundation, Outcomes Assessment Department, Research Division

In the current era of accountability in medicine, information regarding outcomes of care can play a pivotal role in medical decision making for physicians, other healthcare providers, patients, and administrators. Although the field of outcomes assessment has not fully matured, a number of tools and methods can be reliably used to produce valid information. Systematic collection and analysis of outcomes data can add to the complexity of the medical encounter. Yet, if appropriately collected, the information derived can facilitate medical decision making and enhance the quality of medical care. This article defines and discusses healthcare outcomes, reviews the relevance of outcomes measurement, describes practical considerations in, and examples of, outcomes measurement in medical practice, and reviews some resources available for outcomes studies or projects.

One of the current "hot" topics in medicine is the measurement and management of healthcare outcomes. Information regarding outcomes of care can play a pivotal role in medical decision making for healthcare providers and patients searching for a means of delivering or obtaining high quality medical care. Advances in outcomes measurement are providing tools that can quantify components of quality such as clinical outcomes, patient satisfaction, and functional status.

Defining Healthcare Outcomes

Although quality in healthcare is difficult to define and often ambiguous, most would agree that it is ultimately rooted in patient outcomes. Healthcare outcomes have been described as "measures of the end result of what happens to patients as a consequence of their encounter(s) with the healthcare system" (1). When end results are assessed for groups of patients, patterns emerge allowing trends to be identified in clinical outcomes and the effectiveness of medical interventions. The systematic assessment of population healthcare outcomes is an application of clinical epidemiology, which is the study of the distribution of disease in human populations, and of the determinants that influence disease distribution (2). The basis of outcomes assessment in epidemiology provides a framework for analyzing and understanding potential relationships between outcomes and medical interventions.

Whatever the outcome of interest, the goals of medical care are the maximization of good outcomes and the minimization of poor outcomes. This can be best accomplished within a framework that facilitates the assessment of the quality of care. Donabedian, a physician and noted health services researcher, has outlined three constructs from which inferences can be made regarding quality: outcome, process, and structure. He noted that information from these constructs can be used to assess quality if they are causally related: "structure leads to process and process leads to outcomes" (3). Major outcomes, process, and structural variables used in medical outcomes studies have been outlined by others (4).

Outcome Construct

Outcomes can be classified into several categories: morbidity, mortality, pain, functional status, satisfaction, and costs (1) (Table 1). Both generic and disease-specific tools can be used to measure immediate or short-term outcomes (e.g. in-hospital mortality for myocardial infarction patients), intermediate outcomes (e.g. 6-12 month functional ability after total knee replacement), or long-term outcomes (e.g. 10-year survival status for patients treated for breast cancer).

Process Construct

What is done in the care of patients and how it is done are assessed by the process construct. The technical components are often readily quantified (e.g. laboratory tests, medications prescribed, and emergency room, clinic, and hospital visits) (4). How healthcare is delivered (e.g. interpersonal style, communication skills) is more difficult to assess and quantify, but it is also important.

At times, outcomes such as intermediate and long-term events are not feasible to collect. Patient records may be incomplete because of changes in healthcare insurance, changes in providers, incomplete documentation, or other reasons. As a result, process measures that are closely related to the outcome of interest are collected and reported as "surrogate outcomes" measures. For example, the development of retinopathy in diabetic patients is an outcome that may take years to develop. Frequent retinal examinations (*process*) can lead to early diagnosis and treatment of retinopathy, thus minimizing or preventing further deterioration. Because long-term longitudinal data may not be available regarding the development of retinopathy, and because frequent retinal examinations (*process measure*) are related to the outcome of interest, frequency of annual retinal examinations in diabetic patients is often reported as a surrogate outcome measure. Although not ideal, surrogate outcomes measures can provide useful information provided they are closely related to the true outcome of interest. The ultimate goal, however, is to collect and track actual outcomes data.

Table 1. Outcomes Categories and Examples

Category	Examples
Morbidity	Chronic disease Complications
Mortality	Ten-year mortality/survival In-hospital mortality
Pain	Acute Chronic
Functional Status	Physical function Social function
Satisfaction	Overall satisfaction Satisfaction with quality of health care
Cost	Resource utilization Cost-effectiveness, cost-benefits

Structure Construct

The structure of care construct primarily refers to characteristics of the healthcare system, providers, and patients. Although a broader listing has been previously published, system characteristics may include specialty mix and access (4). Provider characteristics may include demographics and specialty training. One key component of the structure construct is the patient characteristics (e.g. demographic, severity of primary disease, and severity of comorbid disease). In comparing outcomes across providers, institutions, or regions, there is often concern regarding the effect of patient "risk" associated with age and severity of primary and/or comorbid disease on the outcome of interest. The underlying theory is that if patients are sicker (*structure*), healthcare providers may alter what is done or not done (*processes*) for the patient (based on the patient's medical status), which in turn might impact the outcome under study. Patient characteristics such as age, severity of primary disease, and/or comorbidity are key components of the structure construct used to quantify "risk." For many diseases, there is an established relationship between patient characteristics and healthcare outcomes. How to appropriately and accurately measure and analyze the risk of patient characteristics remains a challenge. Indicators of risk can be assessed using patient diagnoses, patient medication or laboratory profiles, and/or patient demographics, though patient demographics often do not capture enough of the risk to accurately explore potential differences in outcome. Medication and laboratory data can be used to identify patients with certain chronic and acute diseases; however, access to this data is typically limited and may be incomplete. Patient diagnosis provides useful information when assessing patient risk from severity of disease. A variety of tools have been developed with the goal of quantifying patient risk from different disease conditions. Each has unique scoring and data collection strategies. Careful consideration of patient risk factors in outcomes assessment is key to understanding variation in outcomes. Appropriate adjustment of data for patient risk factors allows medical decision-makers to compare "apples to apples."

A diagram of a framework for the relationships between structure, process, and outcome using Diabetes Mellitus as an example is provided in Table 2. Although the long-term goal is assessment of the outcome variables listed (i.e. glucose control with glycosylated hemoglobin, retinopathy, and/or amputation), oftentimes this information is not readily available. As a result, process measures are frequently reported as surrogates of the outcome of interest.

Table 2. Structure, Process, Outcome Format, Using Diabetes Mellitus as an Example	
Structure ↓	Provider training and certification status Patient risk factors and comorbidities System referral capabilities and diagnostic equipment
Process ↓	Number of patients with HgA_1C tests performed Number of patients with annual retinal exams Number of patients with annual foot exams
Outcome	Mean HgA_1C value (glucose control) Number of patients with diabetic retinopathy Number of patients with extremity amputations
HgA_1C = hemoglobin A_1C (glycosylated hemoglobin)	

Relevance of Outcomes Measurement

Physicians and other healthcare providers have long been interested in what happens to their patients as a result of medical intervention (5,6). Codman presented his "end results idea" in the early 1900s, which encouraged physicians to assess whether outcomes of healthcare were good, and if not, why (5). In recent years, the emphasis on the high cost of medical care and the expansion of managed care has focused more attention on the need for population outcomes information. In addition, variation in medical practice and outcomes that are not readily explained continue to be documented (7). As a result, physicians, other providers, and patients are no longer the only groups interested in outcomes data. Those buying healthcare (including the federal government), accrediting healthcare organizations, or negotiating medical care contracts are requiring that outcomes data be systematically collected, reported, and integrated into medical decision making processes (1).

Although still in its infancy, relatively speaking, outcomes assessment is providing a means for evaluating whether what we do when caring for our patients really makes a difference. It provides a way of tracking and determining whether the right treatments are prescribed to the right patients at the right time and in a cost-effective manner.

Practical Considerations in Outcomes Measurement in Medical Practice

Over the last several decades, much has been accomplished in developing and validating tools and analytic methods that facilitate outcomes measurement. Although much work is still to be done regarding refinement and generalizability of the methods, there is a solid foundation for selected outcomes measurement.

Topic Selection

The first step in outcomes studies and projects is the selection of a topic. When selecting disease conditions or procedures for study, several criteria may be applied. Conditions or procedures that are high volume (or occur frequently), high cost (not only in dollars expended but also in lost patient productivity), or have a functional impact on patients are typically good candidates for further study. In addition, conditions or procedures that are well defined, have quantifiable measures, have potential for improvement, and for which there is an evidence-base for medical intervention are well suited for outcomes studies and related projects (8).

Study Design

Once a topic has been selected, it is important to consider the following:

1. What is the study question? (e.g. what do you hope to find out?)
2. What are the goals or purposes of the study? (e.g. address a research question, meet accreditation requirements, submit results for publication?)
3. What data are available? (e.g. medical records, patient/physician surveys, administrative data?)
4. What are the appropriate data collection tools and have they been determined to be valid and reliable tools for the intended use?
5. What is the appropriate and feasible study design to answer the question? (e.g. randomized controlled study, cohort study, or case-control study?)
6. What are the appropriate statistical tests to utilize in the outcomes data analysis (e.g. chi square, regression analysis?)
7. What are the limitations of the study? (e.g. generalizability of the results, access to all relevant records and information?)

Consideration of these questions up front provides a guide for the study and increases the likelihood of obtaining useful information from the project. Consultation from a statistician or epidemiologist or both is often helpful, particularly in the early phases of the project when planning the study design and statistical analysis.

Validity, Reliability, and Data Collection Tools

A frequent error in outcomes studies is the use of data collection tools that have not been demonstrated to be valid and reliable methods for the specific project's purpose. Validity and reliability are critical in data collection.

Simply put, validity refers to the ability of a tool to measure what it is supposed to measure. Reliability refers to the accuracy and reproducibility of the data collected with any given tool. A non-medical example of these concepts is presented in Figure 1. A person's goal in a dart game is to hit the bull's-eye on Target B. After six tries, he consistently hits the bull's-eye on Target A. The results of the dart game are reliable (i.e. he hit the bull's-eye consistently), but not valid (he hit the wrong bull's-eye) (1). The relevance of validity and reliability extends beyond data collection tools; outcomes and process measures used to assess quality of medical care should be valid and reliable indicators of what actually happens in medical practice. Care must be taken in the selection and use of tools and measures if the information collected is to be used appropriately in decision-making.

In evaluating various tools for outcomes measurement it is also helpful to consider a number of factors (9). Availability is important. If a study or project is to be practice-oriented, it is also relevant to consider the ease of interpretation of the data collected with the tool, along with ease of the tool's administration; data that are cumbersome to collect and interpret are more likely to be set aside. Clinical usefulness of the data should also be addressed up front. Feedback from collected data that can be directly related to clinical care will increase the likelihood of continued measurement and monitoring of outcomes information, which can lead to improved quality of care.

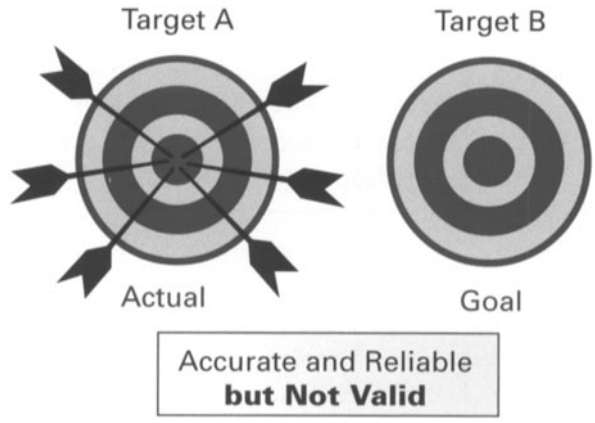

Figure 1. Nonmedical example regarding reliability and validity.

Examples of Outcomes Studies at Ochsner

The Outcomes Assessment Department of the Research Division of the Alton Ochsner Medical Foundation was established in 1990. Since that time, a number of projects have been undertaken and completed in the areas of functional status, predictive modeling, and case-mix or severity adjustment.

Functional Status

One of the areas of increasing popularity in outcomes assessment is functional status measurement. A number of generic tools have been developed and used over the years, such as the Sickness Impact Profile (SIP) (10,11) and the COOP Charts (12,13). One of the current generic tools that has been widely studied in different disease conditions and different languages, and whose validity and reliability have been demonstrated is the Short Form 36 (SF 36) (9,14,15). The 36-question functional status assessment tool was developed for self-administration in patients 14 years of age and older and takes approximately 15 to 30 minutes to complete. It provides a quantitative estimate of nine areas: [1] physical function, [2] mental function, [3] role function-physical, [4] role function-mental, [5] social function, [6] energy/vitality, [7] pain, [8] health perception, and [9] change in health. We have utilized the SF 36 in a number of outcomes studies at Ochsner (16-19) and found that patients with different disease conditions appear to have unique patterns of functional status. Independent cross-sectional studies of patients with systemic lupus erythematosus (18,19) and patients with complaints of low back pain (16) revealed differences in functional status between these two patient populations and the general population of the United States (14) (Figure 2). These differences may potentially serve as a basis for medical intervention or for prompting further tracking and study. Variant patterns by other chronic disease types (e.g. hypertension, diabetes) have also been reported (20).

Functional status, as assessed with the SF 36 tool in patients with certain diseases (such as hypertension), however, has not been shown to differ substantially from the functional status of patients with no chronic conditions (20). Generic functional status information may be more useful in tracking outcomes in some diseases than in others. Nevertheless, collection of functional status information is increasing. Recent developments have led to the Short Form 12 (SF 12), a subset of the SF 36 (21). In addition, physical and mental summary scales have evolved that may be useful in evaluating patient functioning (21,22). Incorporation of functional status assessment into routine practice may provide useful information in the care and management of patients with various conditions.

Predictive Modeling

In a predictive modeling study of patients with low back pain, functional status (as measured with the SF 36) was used as a predictor and not as an outcome variable (16). The outcome variable was physician-determined fit-for-work status. A number of variables were collected from patient questionnaires, self-administered upon check-in for his or her scheduled clinic appointment. Logistic regression analysis was

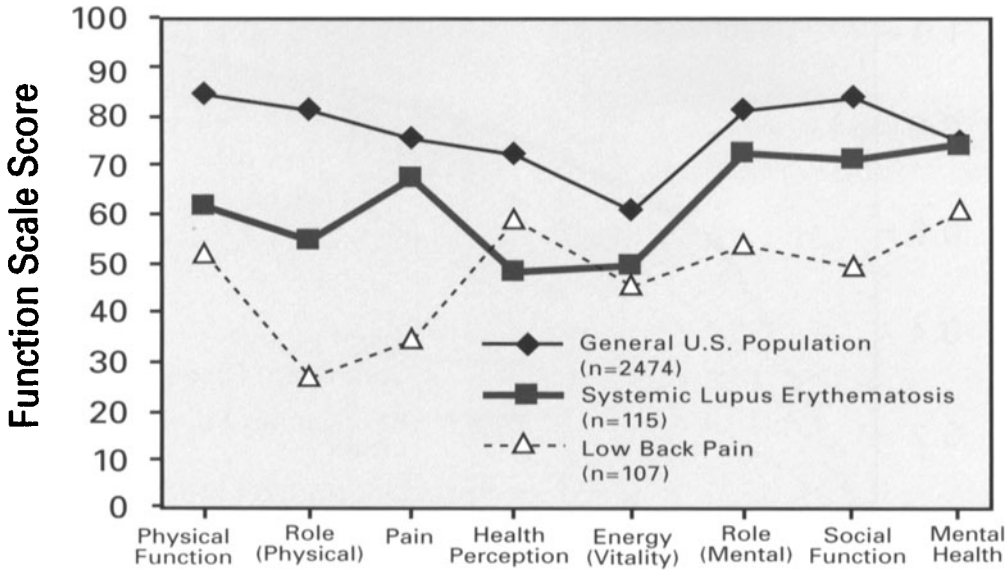

Figure 2. Comparison of function scale scores: low back pain patients versus systemic lupus erythematosus versus general U.S. population (17).

performed on data from 107 patients with complaints of lower back pain. The final model contained the following variables: physical function scale score (SF 36), employment status, smoking status, and physical function (SF 36) by gender interaction. With the model, 90% of the NOT fit-for-work cases and 77% of the fit-for-work cases were correctly classified according to physicians' fit-for-work status classification; the overall correct classification with the model was 85%. The authors concluded that this model, utilizing four patient-reported variables, was a powerful predictor for the outcome of interest. In addition, the model may be useful as a screening mechanism to be used by other healthcare professionals not formally trained in occupational medicine but required to perform fit-for-work evaluations.

Risk Adjustment

As described earlier, adjustment for differences in patient risk is critical in the assessment and comparison of outcomes data. In medical management, questions often arise with respect to variation in outcomes for patients undergoing different procedures for the same condition. In some cases, the medical status or "risk" of the patient drives the selection of the procedure. In order to have a valid comparison of the outcomes for variant procedures for the same condition, one must attempt to control for the potential differences in patient risk.

One study used three unique indices to assess comorbid illness in patients undergoing prostatectomy procedures for benign prostatic hyperplasia (BPH) (23): Charlson Index (24), the Kaplan Feinstein Index (25), and Index of Coexistent Disease (26,27). Although many physicians, including urologists, noted that patients undergoing transurethral prostatectomy (TURP) procedures were likely to be older and sicker than those patients undergoing open procedures, Roos and co-workers had reported an increased 5-year mortality rate for patients undergoing TURP (vs an open procedure), even in a subset of risk-adjusted data (28). This result was unexpected, and uncertainty surrounded the effect of the measure of comorbidity used on the outcome in the Roos study (28).

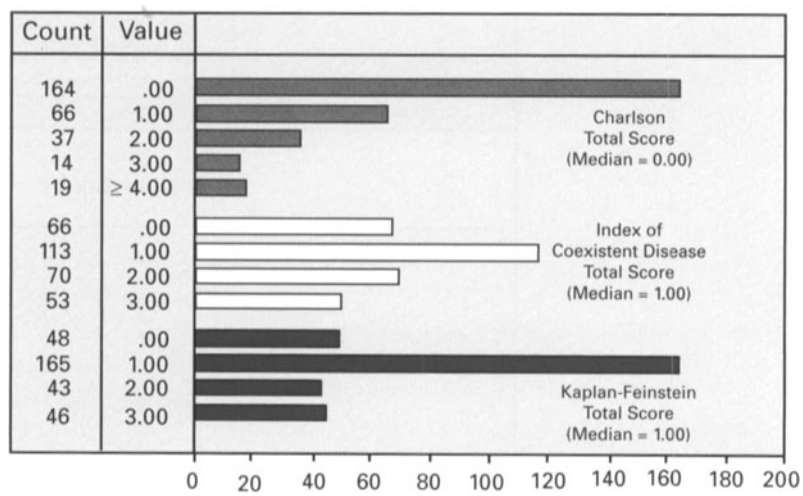

Figure 3. Frequency distribution of three comorbid index scores [Reprinted by permission of Blackwell Sciences, Inc.] (23).

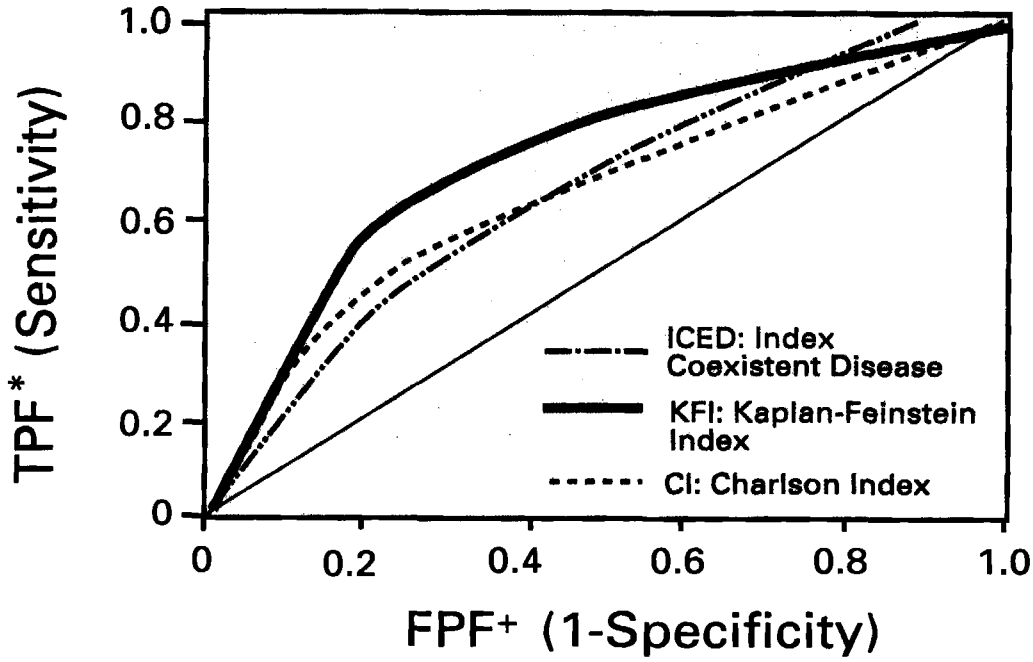

Figure 4. ROC curves for three comorbidity indices (TURP patients only: n=253) [Reprinted by permission of Blackwell Sciences, Inc.] (23). * TPF = True Positive Fraction; + FPF = False Positive Fraction

Table 3. Internet References and Resources for Outcomes Studies and Projects

Web sites	
www.ahcpr.gov	Agency for Health Care Policy and Research
www.ama-assn.org	American Medical Association- clinical quality improvement
facct.org	Foundation for Accountability in Healthcare
nih.gov	National Institutes of Health
nlm.nih.gov/nichsr/nichsr.html	National Library of Medicine (NLM) web-based outreach and training material and link to NLM databases (HSTAT Healthstar, HSR Proj, and DIRLINE)
www.outcomes-trust.org	Medical Outcomes Trust
www.sf-36.com	SF-36 (Short Form 36) - Functional Status Survey

Reference Books

Measuring Health: A Guide to Rating Scales and Questionnaires. 2nd ed. McDowell I, Newell C, editors. New York: Oxford University Press, 1996.

Risk Adjustment for Measuring Healthcare Outcomes. 2nd ed. Iezzoni LI, editor. Chicago: Health Administration Press, 1997.

Health Measurement Scales: A Practical Guide to Their Development and Use. 2nd ed. Streiner DL, Norman GR. New York: Oxford University Press, 1995.

Table 4. Examples of Funding Sources for Outcomes Projects and Studies	
Government Agencies	• Agency of Health Care Policy and Research • National Institutes of Health • Centers for Disease Control and Prevention
Foundations	• Robert Wood Johnson Foundation • Pew Charitable Trust
Pharmaceutical Companies	
Managed Care Organizations	

The goal of the Ochsner study was to determine if any or all of the aforementioned indices could be used in the assessment of an observed excess mortality for patients undergoing a TURP compared with an open prostatectomy. The data collected for the generation of each index was obtained from the medical records of the patients meeting the study inclusion criteria. Several findings were reported:

- Each index resulted in a unique classification of the study population with regards to severity of comorbid illness (Figure 3). [It is important to note that there are index-dependent differences in coding and scoring, and the values are not equivalent across indices. Risk adjustment methodologies are not all alike, which could result in inconsistencies in the prediction of outcomes for different disorders.]
- A Receiver Operating Characteristic (ROC) Curve analysis which was performed to assess index sensitivity and specificity with respect to survival status for patients undergoing TURP revealed that there were no significant differences in the index curves (i.e. the areas under the curves were not significantly different) (Figure 4).
- The logistic regression analysis predicting survival differences between TURP versus open procedure patients revealed that each index had a similar effect in rendering the increased mortality associated with TURP insignificant. The reduction in risk differed depending on the index used.

The authors concluded that their analysis supported earlier work, which indicated that severity of illness accounts for higher 5-year mortality associated with patients undergoing TURP procedures (23). Further work is required to understand the relationship between index variation and risk reduction. Nevertheless, assessment of the risk associated with patient severity of illness is an important factor in outcomes assessment and analysis.

Resources Available for Outcomes Measurement

Tables 3 and 4 provide brief listings of some of the resources available for outcomes studies and projects. The Internet is a key source of current information regarding various outcomes tools, projects, and funding organizations (Table 3). In addition, a number of different groups and organizations provide funding for various types of projects (Table 4). Accessing and utilizing appropriate resources can facilitate the implementation of appropriate outcomes studies and projects.

Conclusion

The measurement of healthcare outcomes for groups of patients is becoming a key factor at all levels of medical care management. Although the field of outcomes assessment has not fully matured, a number of tools and methods can be reliably used to produce valid information. Systematic collection and analysis of outcomes data can add to the complexity of the medical encounter. Yet, if appropriately collected, the information derived can facilitate medical decision-making and the enhancement of quality medical care.

References

1. Krousel-Wood MA. Outcomes assessment and performance improvement: Measurements and methodologies that matter in mental healthcare. In: Rodenhauser P, editor. Mental Health Care Administration: A Guide For Practitioners. Ann Arbor: University of Michigan Press, (in press) 1999.
2. Gordis L. Epidemiology. Philadelphia: W.B. Saunders Company, 1996.
3. Donabedian A. The role of outcomes in quality assessment and assurance. QRB Qual Rev Bull 1992; 18: 356-360.
4. Tarlov AR, Ware JE Jr, Greenfield S, et al. The Medical Outcomes Study. An application of methods for monitoring the results of medical care. JAMA 1989; 262:925-930.
5. Codman EA. The product of a hospital. Surg Gynecol Obstet 1914; 18:491-496.
6. Iezzoni LI, editor. Risk Adjustment for Measuring Health Care Outcomes. Ann Arbor, Michigan: Health Administration Press, 1994.
7. Wennberg JE, Cooper MM, editors. The Dartmouth Atlas of Health Care in the United States. Chicago: American Hospital Publishing, Inc., 1998.
8. McGlynn EA, Asch SM. Developing a clinical performance measure. Am J Prev Med 1998; 14:14-21.
9. McHorney CA, Ware JE Jr, Rogers W, et al: The validity and relative precision of MOS short- and long- form health status scales and Dartmouth COOP charts. Results from the Medical Outcomes Study. Med Care 1992; 30 (5 Suppl): MS253-MS265.
10. Follick MJ, Smith TW, Ahern DK. The sickness impact profile: a global measure of disability in chronic low back pain. Pain 1985; 21:67-76.
11. Bergner M, Bobbitt RA, Carter WB, et al. The sickness impact profile: development and final revision of a health status measure. Med Care 1981; 19:787-805.
12. Landgraf JM, Nelson EC, Hays RD, et al. Assessing function: does it really make a difference? In: Lipkin M, editor. Functional Status Measurement in Primary Care. New York, NY: Springer-Verlag NY Inc, 1990:159-161.
13. Nelson EC, Landgraf JM, Hays RD, et al. The COOP function charts: a system to measure patient function in physicians' offices. In: Lipkin M, editor. Functional Status Measurement in Primary Care. New York, NY: Springer-Verlag NY Inc, 1990:97-121.
14. Ware JE. SF-36 Health Survey Manual and Interpretation Guide. Boston MA: The Health Institute, New England Medical Center, 1993.
15. Shiely JC, Bayliss MS, Keeler SD, et al. SF-36 Health Survey Annotated Bibliography. 1st ed. (1988-1995). Boston MA: The Health Institute, New England Medical Center, 1996.
16. Krousel-Wood MA, McCune TW, Abdoh A, et al: Predicting work status for patients in an occupational medicine setting who report back pain. Arch Fam Med 1994; 3:349-355.
17. Krousel-Wood MA, Re RN: Health status assessment in a hypertension section of an internal medicine clinic. Am J Med Sci 1994; 308:211-217.
18. Quinet R, Davis W, Krousel-Wood MA, et al. Striking inverse correlation of each functional status scale (SF-36) with degree of disease activity in SLE (SLAM-R) (abstract). Arthritis Rheum, 1996; 39(6), R20.
19. Krousel-Wood MA, Quinet R, Davis W, et al. Predictors of SF-36 physical role functioning in patients with systemic lupus erythematosus (SLE). Health Services Research: Implications for Policy. Health Care Delivery and Clinical Practice (Abstract Book). Atlanta: Association for Health Services Research, 1996:171.
20. Stewart AL, Greenfield S, Hays RD, et al. Functional status and well-being of patients with chronic conditions. Results from the Medical Outcomes Study. JAMA 1989; 262: 907-913.
21. Ware JE, Kosinski M, Keller SD. SF-12: How To Score the SF12 Physical and Mental Health Summary Scales. 2nd edition. Boston, MA: The Health Institute, New England Medical Center, 1995.
22. Ware JE, Kosinski M, Keller SD. SF-36 Physical and Mental Health Summary Scales: A User's Manual. Boston MA: The Health Institute, 1994.
23. Krousel-Wood MA, Abdoh A, Re R: Comparing comorbid-illness indices assessing outcome variation: the case of prostatectomy. J Gen Intern Med 1996; 11:32-38.
24. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987; 40:373-383.
25. Kaplan MH, Feinstein AR. The importance of classifying initial co-morbidity in evaluating the outcome of diabetes mellitus. J Chronic Dis 1974; 27:387-404.
26. Greenfield S, Apolone G, McNeil BJ, et al. The importance of co-existent disease in the occurrence of postoperative complications and one-year recovery in patients undergoing total hip replacement. Comorbidity and outcomes after hip replacement. Med Care 1993; 31:141-154.
27. Nicolucci A, Cubasso D, Labbrozzi D, et al. Effect of coexistent diseases on survival of patients undergoing dialysis. ASAIO J 1992; 38:M291-M295.
28. Roos NP, Wennberg JE, Malenka DJ et al. Mortality and reoperation after open and transurethral resection of the prostate for benign prostatic hyperplasia. N Engl J Med 1989; 320: 1120-1124.

M.A. Krousel-Wood, MD, MSPH, FACPM, is a Reseacher in Ochsner's Outcomes Assesment Dept and a member of the Ochsner Journal Editorial Board.