Users' Guides to the Medical Literature
JAMA; Feb 21, 1996; 275, 7; AMA Titles
pg. 554

Users' Guides to the Medical Literature

X. How to Use an Article Reporting Variations in the Outcomes of Health Services

C. David Naylor, MD, DPhil; Gordon H. Guyatt, MD, MSc; for the Evidence-Based Medicine Working Group

CASE SCENARIO

Your patient, a 78-year-old retired internist, has been complaining of increasing symptoms of benign prostatic hyperplasia. He has long-standing hypertension and coronary artery disease, with remote anterolateral myocardial infarction and bypass surgery 10 years ago. His left ventricular ejection fraction was recently documented at 20%, and he has been started on an angiotensin-converting enzyme inhibitor. Rectal examination confirms a moderately enlarged prostate, without irregularities, nodularity, or tenderness. As you discuss management options, your patient insists that transurethral prostate surgery is dangerous and that international studies of thousands of patients have proved that, as he puts it, "old-fashioned open prostatectomy is safer than that keyhole surgery." You prescribe a trial of an α-blocker, terazosin, and arrange to see him again. However, the retired internist sounds so convinced that you also resolve to look into the evidence about the two forms of prostatectomy.

THE SEARCH

Later, you sit down in the hospital library, using a program that contains the MEDLINE database from January 1990 to October 1994. You start from "Explode Prostatic Hyperplasia," limit the search to English-language articles on human subjects, and then combine the resulting set with "transurethral" and "mortality" as text words. This yields 27 citations. Browsing through the resulting abstracts, two appear to address your patient's concern. One, by a Danish group,[1] addresses the long-term outcomes of transurethral vs "open" (suprapubic or transvesical) prostatectomy using hospitalization data linked to vital status data for the entire Danish male population from 1977 to 1985. The study relies on administrative data and massive population-based numbers (38 067 men) and shows excessive mortality among patients undergoing transurethral resection of the prostate (TURP). The other report, by Concato et al,[2] offers long-term outcomes data on only 252 patients who underwent either procedure at a Yale teaching hospital in New Haven, Conn, between 1979 and 1981. However, a detailed chart audit was undertaken, and the results suggested that patients undergoing the more extensive open procedure had lower long-term mortality because they were healthier at the outset.

INTRODUCTION

Over the last decade, changes in health care delivery have broadened the range of groups interested in the outcomes of medical care. Concern with costs and with dramatic interregional or international differences in practice among clinicians and institutions have focused the attention of administrators and politicians on the interplay between the processes and outcomes of health services. The evolution of managed care has sharpened interest in measuring and managing the quality of care delivered by individual practitioners, hospitals, and other institutions.

Implicitly, the questions about quality of care and the best way of delivering health services are issues of optimal treatment. For example, once a patient's problem is identified, the primary care physician first determines what intervention, if any, should be undertaken, and may then face the quality-related issue of choosing a specialist or institu-

From the Institute for Clinical Evaluative Sciences, Ontario, North York, the Clinical Epidemiology and Health Care Research Program, Sunnybrook Unit, and the Departments of Medicine and Surgery, University of Toronto (Ontario) (Dr Naylor); and the Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario (Dr Guyatt).

A complete list of members (with affiliations) of the Evidence-Based Medicine Working Group appears in the first article of this series (*JAMA*. 1993;270:2093-2095). The following members contributed to this article: Eric Bass, MD, MPH; Hertzel Gerstein, MD, MSc; Daren Heyland, MD, MSc; Ann Holbrook, MD, PharmD, MSc; Virginia Moyer, MD, MPH; Tom Newman, MD; Andrew Oxman, MD, MSc; W. Scott Richardson, MD; Peter Tugwell, MD, MSc; and John Williams, Jr, MD, MHS.

Corresponding author: C. David Naylor, MD, DPhil, Institute for Clinical Evaluative Sciences in Ontario, G-106, Sunnybrook Health Science Centre, 2075 Bayview Ave, North York, Ontario, Canada M4N 3M5.

Reprint requests to Room 2C12, McMaster University Health Sciences Centre, 1200 Main St W, Hamilton, Ontario, Canada L8N 3Z5 (Dr Guyatt).

Users' Guides to the Medical Literature section editor: Drummond Rennie, MD, Deputy Editor (West), *JAMA*.

tion to offer that service. From a prior Users' Guide[3] you've learned that decisions about what treatment to provide are best made in light of evidence from randomized studies with complete follow-up. However, investigators are generally not going to be able to randomize patients to different practitioners or hospitals, and focusing on the outcomes associated with these differences in care will require strategies other than randomized trials. Increasingly, investigators have looked to large administrative or other observational databases to examine the outcomes of care associated with different procedures, practitioners, or institutions. Under what circumstances should you believe the inferences made on the basis of such studies?

There is a parallel here with studies assessing potential harm to patients: it is impossible to randomize people to smoke or not, or to various levels of air pollution, and so observational studies or "natural experiments" are used as sources of insight. In a previous Users' Guide[4] we provided criteria for validity for the observational studies that investigators must use when exploring issues of harm. The challenges are fundamentally the same for comparing outcomes of two or more sets of health care practitioners or delivery systems. However, observational studies using administrative databases are growing in scope and importance and have their own particular challenges. Therefore, we devote this Users' Guide to these issues. Table 1 revisits our criteria for assessing an article about harm, modified here for examining associations between variations in processes and outcomes of health care in the real-world setting.

Table 1.—Three Core Questions to Ask About a Study Using an Observational Design to Examine Sources of Difference in Patients' Outcomes

Are the outcome measures accurate and comprehensive?
Were there clearly identified, sensible comparison groups?
Were the comparison groups similar with respect to important determinants of outcome, other than the one of interest?

ARE THE OUTCOME MEASURES ACCURATE AND COMPREHENSIVE?

A randomized therapeutic trial must have valid and reliable outcome measures; so must any observational study assessing patients' outcomes. The easiest outcomes for health researchers to measure are those that are defined objectively and usually captured in large insurance databases or computerized hospital administrative data, eg, death, in-hospital complications of surgery that are routinely coded, or readmissions to the hospital. Linkage to vital status registries is also performed to track out-of-hospital deaths. However, other outcomes, eg, disability, discomfort, distress, and dissatisfaction[5] are important to patients. Functional status and quality-of-life measures are needed to capture these burdens, but these measures are not applied in routine clinical care, and if applied, their results are not incorporated into administrative databases. Incorporating these measures into routine care and administrative databases, moreover, may generate more questions than answers. Researchers have begun to understand some of the factors that predict, for example, increased risk of mortality after various types of elective surgery. However, there is no similar understanding of the factors that predict functional status and quality of life.

In sum, many large databases are not designed for clinical research and may either mismeasure patients' outcomes or fail to capture outcomes that are important to patients and their physicians. Researchers should therefore report on the quality and comprehensiveness of the data source. Ideally there should be independent cross-checks to ensure that the same outcomes are measured consistently and completely for whatever unit of comparison is used, eg, verifying that data on ascertainment or cause of death are accurate or confirming hospital readmission rates after a specific surgical procedure in a quality-of-care study.

How did our two studies of prostate surgery perform in these respects? Andersen et al[1] used vital status data for the entire population of Denmark, and therefore mortality was measured in a reliable and unbiased fashion across all groups for comparison. Concato et al[2] reported on all-cause mortality data within 5 years of the procedure obtained by hospital chart review and, where those data were inconclusive, from the national vital status registry.

The complete resection attained by open prostatectomy obviously eliminates the need for repeat procedures as occasionally occurs with TURPs. However, neither study compared the two procedures with respect to various outcomes of interest to patients and physicians, eg, effectiveness in relieving obstructive or irritative symptoms of benign prostatic hyperplasia, overall recovery time, rates of complications such as impotence or incontinence, and so forth. Careful prospective data collection is necessary to capture these outcomes and provide a more complete tally of the burdens and benefits of the two treatments being compared. Even with those data, moreover, there would be uncertainty about the weights that patients would give to diverse benefits and harms, and a major challenge in determining how different outcomes related to each other and to patients' pretreatment characteristics.

Table 2.—Factors That May Systematically Affect Outcomes

What service was provided* For example, variations among two or more management strategies with respect to use of drugs, doses, devices, type of procedure, and the like
Who provided the service For example, variations among procedural specialists; nurse practitioners vs family physicians; by level of experience (house staff vs qualified specialists); by volume of service delivered (high-caseload vs low-caseload practitioners)
Where the service was provided For example, variations among hospitals or clinics; between wards in a hospital; between a step-down unit and a conventional intensive care unit; home vs hospital care; by city; by county; by region or nation
When the service was provided For example, variations in timing of service (eg, day or evening, weekend vs weeks, the July phenomenon for house staff effects); according to length of stay in hospital; across months (seasonal effects) or years (broad temporal trends)

*These questions are best addressed using randomized trial methods; see Guyatt et al.[3]

WERE THE COMPARISON GROUPS SIMILAR WITH RESPECT TO IMPORTANT DETERMINANTS OF OUTCOME OTHER THAN THE ONE OF INTEREST, AND WERE RESIDUAL DIFFERENCES ADJUSTED FOR IN THE ANALYSIS?

Clinicians and health care managers are interested in a variety of determinants of outcome, the major categories of which are shown in Table 2. One type of comparison examines differences that may be due to variations in quality of care across individual practitioners or institutions providing care in a specific city or region. State agencies now publish some provider- or institution-specific outcomes, and researchers sometimes relate these outcomes to the provider- or institution-specific volume of the services under scrutiny. This reflects a belief that "practice makes perfect"—all things being equal, centers (and by, inference, physicians or surgeons) with a higher caseload will generally achieve better outcomes than lower-volume centers. For example, various studies suggest that in-hospital postoperative mortality after aortic aneurysm surgery,[6] percutaneous transluminal coronary angioplasty,[7] and coronary artery bypass graft surgery[8,9] is lower for centers or surgeons managing more patients. On the other hand, large tertiary care centers often treat the sickest patients and therefore may have worse outcomes than smaller hospitals.

However, the greater the difference between service settings being compared, the more difficult it is to be sure that patients were similar, or to isolate

Table 3.—Determining Whether Differences in Prognosis, Rather Than Differences in the Intervention, Explain Differences in Outcomes

Were all important prognostic factors measured?
Were measures of patients' prognostic factors reproducible and accurate?
To what extent were patients similar with respect to these factors?
Was multivariate analysis used to adjust for imbalances in prognostic factors?
Did additional analyses (particularly in low-risk subgroups) demonstrate the same results as the primary analysis?

which aspects, if any, of the process of care relate to the outcomes observed. This is especially true when comparisons are made on a broad geographic footing between regions or countries in which populations and processes of care differ in many ways. One recent study compared outcomes of Canadian and American patients enrolled in a major trial of thrombolytic therapy for acute myocardial infarction.[10] Rates of revascularization and use of specialist services were much higher in the United States. The investigators used an appropriately broad range of outcomes measures and observed that in terms of symptoms, functional status, psychological well-being, and health-related quality of life, Canadian patients fared somewhat worse than their American counterparts—a finding of obvious concern to Canadian practitioners. However, some of the difference may be because the types of patients recruited by Canadian investigators were destined for worse outcomes irrespective of management. Canadians may also have a different cultural threshold for reporting symptoms or functional impairment.

A third source of variations in outcomes that may occur within similar health systems is the type of treatment provided. This is the sort of comparison that was done in the outcomes studies of TURP vs open prostatectomy described in this article's opening scenario. Such comparisons may avoid some of the broad health system effects and sociocultural or even genetic differences that threaten the validity of outcomes comparisons made across widely disparate populations. However, it is still possible that differences in outcomes may have been due to differences in patients receiving the alternative management strategies, for without randomization, patients will inevitably differ in ways other than the treatment being provided to them. This phenomenon is called "selection bias." When two alternative procedures are being compared in research, selection bias arises from the exercise of good clinical judgment in routine practice. For example, urologists may choose younger, healthier patients to undergo the more extensive open prostatectomy, and older, sicker patients for TURP. Patients then end up differing in obvious or subtle ways that affect their likelihood of having a good or bad outcome. Epidemiologists use the term "confounding" to describe this problem. The validity of any form of observational research is threatened by case selection biases that create noncomparable groups of patients and confound any outcomes comparisons.

Researchers must therefore somehow adjust for differences between groups of patients. The sophistication of these so-called risk adjustment methods is growing rapidly.[11] However, researchers and quality-of-care evaluators are unlikely to know all the prognostic factors that interact with treatments to affect outcomes. Randomization is important precisely because it distributes these unknown factors in an unbiased manner. The problem worsens when one considers that all known prognostic features may not have been measured, and if they have been measured, they may not have been measured or recorded accurately. Inaccurate measurement or recording is a particular concern when information comes from administrative databases. For instance, Jollis et al[12] compared information about cardiac risk factors in an administrative database in patients undergoing angiography with information collected prospectively for a clinical database by a cardiac fellow who actually saw the patients. A chance-corrected measure of agreement (κ statistic) showed good agreement only for diabetes (83% agreement) and whether patients had an acute myocardial infarction (76%); agreement was moderate for hypertension (56%), poor for the presence of heart failure (39%), and no better than chance (9%) for unstable angina. Hannan et al[13] found similar discrepancies in comparing a cardiac surgery registry with an administrative database in New York State. These inaccuracies mattered: the ability of evaluators to predict mortality was clearly higher with the detailed clinical data as opposed to the administrative database.[13] Thus, the accuracy, reproducibility, and fairness of adjustments for differences in patients can be undermined by poor data quality.

The problem of limited or inaccurate data in insurance databases or computerized hospital discharge abstracts may be partly ameliorated by supplementing the information with chart audits.[14] This is time-consuming and expensive, but may be the only way to reduce the chances of missing or misconstruing important differences among groups of patients. A more efficient mechanism may be to establish specific registry mechanisms geared to measuring key patient characteristics, process of care elements, and relevant outcomes.

How, then, can you best assure yourself that, short of randomization, investigators have made the fairest possible outcomes comparison possible? We summarize the steps in Table 3. First, did the researchers convince you, through their review of the literature and on the basis of what you know about the determinants of prognosis, that they measured all of the important prognostic factors? This is more likely to occur if the analysis involves chart audits or, better still, a specific clinical registry, as opposed to reliance on available administrative data. Second, since these measurements are only as good as the data that go into them, you should consider whether these measures of patients' prognostic factors are reproducible and accurate. Third, did the researchers show the extent to which the groups being compared differed on the prognostic factors that they measured? Fourth, did they use some form of multivariate analysis wherein they tried to adjust simultaneously not only for the obvious prognostic factors, but also for other more subtle differences that may have confounded the comparisons?

Localio and colleagues[15] have recently reported on the consequences of not taking into account all possible prognostic factors. A large corporation's managed care program sought to determine which of the hospitals serving the corporation's employees delivered better quality of care as reflected in part by fewer in-hospital deaths. A consultant concluded that the hospitals differed, and this conclusion influenced the company's choices about hospital selection. As it turned out, an appropriate analysis conducted by a group of academic investigators concluded that the difference between even the hospital with the worst record and the rest could be easily attributable to the play of chance. Furthermore, when the investigators included an adjustment for age, a prognostic factor that had been left out of the consultant's initial analysis, the rank order of the hospitals changed.[16]

Because observational data are so susceptible to selection biases that may confound the outcome comparisons, the researchers should determine whether their results persist when they analyze the data in different ways. For example, if there is a severe imbalance in allocation of patients with a particularly important prognostic factor, it may make sense to eliminate all patients with that factor and repeat the analyses. Unfortunately, even relative balance on a prognostic factor does not guarantee comparability. One reason is that administrative data and

registries tend to use fairly simple categories, such as whether a disease is or is not present. Yet, the category "disease present" may be associated with a wide range of underlying dysfunction, and therefore equally variable prognosis. Patients with chronic lung disease or chronic heart failure, for instance, can vary from mild to severe, with very different prognostic implications. Thus, apparent balance on the proportion of patients with these diagnoses can mask a situation in which one group has many more severely affected patients than the other. This is even true for advanced age as a prognostic factor, since elderly persons may vary considerably in their overall robustness.

Because of this problem, a useful double-check in any outcomes comparison is to ensure that the findings are replicable within a relatively low-risk subgroup of the patients being examined. By eliminating patients in categories associated with widely varying physiological states, we increase the likelihood of a "level playing field" for comparisons.

How do our two studies of prostate surgery measure up in this regard? Andersen et al[1] considered patients' ages at surgery, but relied only on diagnoses coded in the computerized hospital records as indicating compromised health status. Even with these limited data, fewer open prostatectomy patients had high-risk diagnoses. They were also younger and had less heart disease and cancer. In a multivariate analysis to try to adjust for these differences, it did appear that TURP continued to confer a 30% to 40% relative increase in the risk of death over several years of follow-up. Extensive sensitivity analyses were performed, including a specific examination of low-risk patients (described as "healthiest men"). Although low-risk patients also showed an excess risk with TURP, the relative magnitude of the increased risk of death was smaller for low-risk patients than for high-risk patients. As Andersen et al[1] stated: "The extent to which this difference is attributable to the surgical intervention itself remains an open question. The two groups of patients are quite different with regard to age and preoperative health status, and available data may not be sufficient to control such differences through statistical analysis."

Concato et al[2] used chart review methods with a detailed and systematic abstraction of information related to health status based on inpatient and ambulatory care records. They carefully confirmed that two reviewers independently agreed on patients' health status assessments. Patients in the TURP group were again found to be older and sicker. However, in a multivariate analysis, the adjusted excess risk of TURP diminished as the degree of detail on comorbidity was increased. Their best estimate was that TURP actually conferred no increased risk relative to open prostatectomy. Unfortunately, owing to the small sample size, their results were very imprecise, with 95% confidence limits ranging from much increased to much reduced risk with TURP (eg, from 0.57 to 1.87). Thus, the Yale study highlights the issue of noncomparability and selection biases, but does not rule out harms of the magnitude demonstrated by the Danish investigators. Moreover, the study provides data on outcomes for only a single city; the results may not be generalizable.

CONCLUSIONS AND RESOLUTION

Given the limitations of observational studies of large databases, can we better define the role of this sort of health services research? Observational studies do remain important in the generation of hypotheses about causal pathways from a pathophysiological standpoint. Moreover, once randomized trials have helped define what treatments are likely to work best for your patients, observational outcomes studies generate information about what happens when these practices are used in the real world as opposed to the selected populations of patients and practitioners participating in randomized trials. This information deepens our understanding of practical effectiveness as opposed to theoretical efficacy, and may add new insights since trials do not always measure all the outcomes of interest to patients and physicians.

However, this complementary or supplementary role of large-scale observational studies departs sharply from using administrative data or clinical registries to decide which specific management strategies will yield better outcomes: eg, surgery vs medical, invasive vs noninvasive, different surgical procedures, and so on. To determine the relative merits of treatments, randomized trials are usually possible and preferable given the unavoidable biases of observational studies.

Do observational studies have any role at all in choosing best practices? Randomized trials are expensive and difficult to conduct and cannot be undertaken for all the clinical questions in which practitioners are interested. Observational studies may identify situations in which one therapy appears so much better than an alternative that bias would be a very unlikely explanation for the difference. As well, the hypothesis-generating role of observational studies is illustrated by the example of open prostatectomy. (Unfortunately, the convenience of transurethral surgery, together with deeply held beliefs about its safety, probably precludes ever mounting a large-scale trial comparing transurethral and open prostatectomy.) Finally, if the outcomes of interest are very rare, such as unusual idiosyncratic side effects of a drug, researchers can only obtain adequate sample sizes through use of administrative databases.

There are other situations in which randomization is not feasible, such as looking for systematic variations in outcomes of similar procedures provided by different practitioners or institutions ("who" or "where" rather than "what"; see Table 2). It is untenable to assume that all hospitals or providers practice equally well and observational outcomes comparisons have a role in assessing quality of care. This is especially applicable for some well-defined services (eg, coronary artery bypass grafting) where there are validated risk-adjustment algorithms[17-20] and dedicated registries to measure risk factors and outcomes, so that these comparisons are probably meaningful. In general, however, potential harm to patients from poor quality care must be weighed against the harm to skilled health workers and fine institutions caused by poorly founded inferences about inferior outcomes.

Given the relatively weak inferences possible from most observational studies of outcomes, alternative strategies for ensuring the quality of medical care should always be considered. For some processes of care (though certainly not all, as we caution in the next article in this series), we can accurately document what went on and make confident judgments about its appropriateness. For example, randomized trials show that preoperative antibiotic and antithrombotic prophylaxis improves patients' outcomes after various surgical procedures. Systematically omitting these treatments puts patients at risk and indicates a need for practitioners and institutions to improve their quality of care. We suggest that in most instances it is most efficient to use randomized trials or meta-analyses of trials to establish optimal management strategies, and then assess if quality of care is maintained by monitoring the process of care to ensure that well-proven practices are consistently applied to eligible patients.

What, then, of your patient? Perhaps predictably, given what we know about the limitations of observational studies, your exploration has been inconclusive. Indeed, had you used MEDLINE on CD-ROM for the years prior to 1990, the relevant literature would not have

moved you much further. Related work[21,22] on increased mortality after TURP as opposed to open prostatectomy has incorporated extra detail on differences among patients drawn from chart reviews and failed to eliminate the excess mortality seen with TURP; however, the adjustments were arguably less detailed than those used by Concato et al.[2] One very small randomized trial has also shown a trend to excess mortality with TURP.[23] On the other hand, there has been no definitive trial comparing the two forms of surgery and TURP remains the predominant procedure for benign prostatic hyperplasia.

The retired internist returns in 4 weeks as planned. "Was I right about the risks of the keyhole method?" he asks. You admit that the abandonment of open prostatectomy may have been premature, but caution that his age and medical status make him a poor candidate for the more extensive procedure, even if you could find a urologist competent to do it. Hearing your own advice, you again appreciate that similar selection biases may be the real reasons for the apparently higher mortality after TURP. Fortunately, your patient has had an excellent response to the α-blocker and the issue of prostatectomy can be set aside for some time. As you usher him from the office, he grumbles: "By the way, did you see that the operative mortalities for all the local heart surgeons are on the front page of the newspaper? Thank heavens I retired."

References

1. Andersen TF, Bronnum-Hansen H, Sejr T, Roepstroff C. Elevated mortality following transurethral resection of the prostate for benign hypertrophy! but why? *Med Care.* 1990;28:870-881.
2. Concato J, Horwitz RI, Feinstein AR, Elmore JG, Schiff SF. Problems of comorbidity in mortality after prostatectomy. *JAMA.* 1992;267:1077-1082.
3. Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, II: how to use an article about therapy or prevention, A: are the results valid? *JAMA.* 1993;270:2598-2601.
4. Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V, for the Evidence-Based Medicine Working Group. Users' guide to the medical literature, IV: how to use an article about harm. *JAMA.* 1994; 271:1615-1619.
5. White K. Improved medical statistics and health services systems. *Pub Health Rep.* 1967;82:847-854.
6. Hannan EL, Kilburn H Jr, O'Donnell JF, et al. A longitudinal analysis of the relationship between in-hospital mortality in New York State and the volume of abdominal aortic aneurysm surgeries performed. *Health Serv Res.* 1992;27:517-542.
7. Jollis JG, Peterson ED, DeLong ER, et al. The relation between the volume of coronary angioplasty procedures at hospitals treating Medicare beneficiaries and short-term mortality. *N Engl J Med.* 1994;331:1625-1629.
8. Showstack JA, Rosenfeld KE, Garnick DW, Luft HS, Schaffarzick RW, Fowles J. Association of volume with outcome of coronary artery bypass graft surgery: scheduled vs nonscheduled operations. *JAMA.* 1987;257:785-789.
9. Hannan EL, Kilburn H Jr, Bernard H, O'Donnell JF, Lukacik G, Shields EP. Coronary artery bypass graft surgery: the relationship between in-hospital mortality rate and surgical volume after controlling for clinical risk factors. *Med Care.* 1991;29:1094-1107.
10. Mark DB, Naylor CD, Hlatky MA, et al. Use of medical resources and quality of life after acute myocardial infarction in Canada and the United States. *N Engl J Med.* 1994;331:1130-1135.
11. Daley J, Shwartz M. Developing risk-adjustment methods. In: Iezzoni LI, ed. *Risk Adjustment for Measuring Health Care Outcomes.* Ann Arbor, Mich: Health Administration Press; 1994:199-238.
12. Jollis JG, Ancukiewicz M, DeLong ER, Pryor DB, Muhlbaier LH, Mark DB. Discordance of databases designed for claims payment versus clinical information systems: implications for outcomes research. *Ann Intern Med.* 1993;119:844-850.
13. Hannan EL, Kilburn H Jr, Lindsey ML, Lewis R. Clinical versus administrative data bases for CABG surgery: does it matter? *Med Care.* 1992; 30:892-907.
14. Malenka DJ, McLerran D, Roos N, Fisher ES, Wennberg JE. Using administrative data to describe casemix: a comparison with the medical record. *J Clin Epidemiol.* 1994;47:1027-1032.
15. Localio AR, Hamory BH, Sharp TJ, Weaver SL, TenHave TR, Landis JR. Comparing hospital mortality in adult patients with pneumonia. *Ann Intern Med.* 1995;122:125-132.
16. Wu AW. The measure and mismeasure of hospital quality: appropriate risk-adjustment methods in comparing hospitals. *Ann Intern Med.* 1995;122: 149-150.
17. Tu JV, Jaglal SB, Naylor CD, and the Steering Committee of the Provincial Adult Cardiac Care Network. Multicenter validation of a risk index for mortality, intensive care unit stay, and overall hospital length of stay after cardiac surgery. *Circulation.* 1995;91:677-684.
18. O'Connor GT, Plume SK, Olmstead EM, et al, for the Northern New England Cardiovascular Disease Study Group. Multivariate predictors of in-hospital mortality associated with coronary artery bypass graft surgery. *Circulation.* 1992;85:2110-2118.
19. Higgins TL, Estafanous FG, Loop FD, Beck GJ, Blum JM, Paranandi L. Stratification of morbidity and mortality outcome by preoperative risk factors in coronary artery bypass patients: a clinical severity score. *JAMA.* 1992;267:2344-2348.
20. Edwards FH, Clark RE, Schwartz M. Coronary artery bypass grafting: the Society of Thoracic Surgeons National Database experience. *Ann Thorac Surg.* 1994;57:12-19.
21. 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.
22. Malenka DJ, Roos N, Fisher ES, et al. Further study of the increased mortality following transurethral prostatectomy: a chart-based analysis. *J Urol.* 1990;144:224-228.
23. Meyhoff H-H. Transurethral versus transvesical prostatectomy: clinical, urodynamic, renographic and economic aspects: a randomized study. *Scand J Urol Nephrol.* 1987;4(suppl 102):1-26.