EDITORIAL VIEWS

Anesthesiology 2010; 113:1001–3

Copyright © 2010, the American Society of Anesthesiologists, Inc. Lippincott Williams & Wilkins

Risk Stratification, Risk Adjustment, and Other Risks

CME This article has been selected for the ANESTHESIOLOGY CME Program. Learning objectives and disclosure and ordering information can be found in the CME section at the front of this issue.

"BUT our patients are sicker than yours!" is the refrain when physicians are faced with adverse comparisons of their patients' outcomes with those of other groups, hospitals, or health systems. Meaningful and fair comparison of clinical performance requires statistical adjustment of outcomes for differences in clinical acuity of patients treated and complexity of procedures performed, among other characteristics, comprising case mix. Without such risk adjustment, comparisons are biased, physicians understandably avoid sicker patients and/or more challenging procedures, and sicker patients face barriers to access for necessary care. In this issue of ANESTHESIOLOGY, Sessler *et al.*[1] describe development and validation of a "broadly applicable," robust tool for risk-adjusting mortality and length-of-stay outcomes of U.S. hospital care. To understand their achievement and its benefits and limitations, we must first appreciate the challenges.

Risks are ubiquitous but not evenly distributed in time or space. A hierarchy exists in need for health services, and health status generally worsens (acuity of illness increases) sequentially from the unselected general population to outpatient settings to community hospitals and finally to academic medical centers (fig. 1A). Surgical care has a similar spatial distribution (fig. 1B). Thus, complication and death rates for ostensibly similar care are likely to differ across settings and physicians, and they should not be compared without meaningful effort to adjust outcomes for case mix.

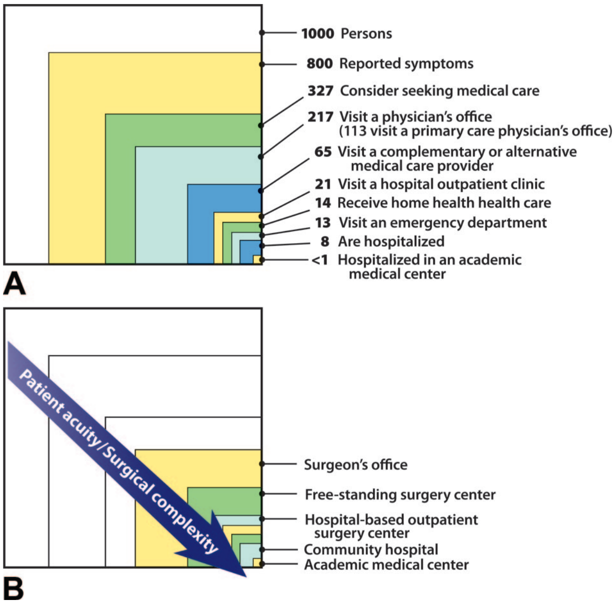
1000 Persons
800 Reported symptoms
327 Consider seeking medical care
217 Visit a physician's office
(113 visit a primary care physician's office)
65 Visit a complementary or alternative medical care provider
21 Visit a hospital outpatient clinic
14 Receive home health care
13 Visit an emergency department
8 Are hospitalized
<1 Hospitalized in an academic medical center
A
Patient acuity/Surgical complexity
Surgeon's office
Free-standing surgery center
Hospital-based outpatient surgery center
Community hospital
Academic medical center
B

Fig. 1. (*A*) Ecology of care, with the site of U.S. healthcare utilization for each 1,000 persons in the year 2000. Reprinted with permission from Green LA, Fryer GE Jr, Yawn BP, Lanier D, Dovey SM: The ecology of medical care revisited. N Engl J Med 2001; 344:2021–5. (*B*) Analogous spatial distribution of U.S. surgical care, with different types of surgical sites reflecting different strata (not to scale) of patient acuity and surgical complexity. Reprinted with permission from Orkin FK, Longnecker DE. Anesthesia risk, Anesthesiology (edited by Longnecker DE, Brown DL, Newman MF, Zapol WM). New York, McGraw-Hill, 2008.

Imposition of Medicare's Prospective Payment System for hospital care in 1983 was the stimulus, augmented later beyond Medicare, for developing risk-adjustment methods when comparing clinical outcomes, costs of care, and physician performance. Such outcomes are viewed as products of complex functions of patient-related clinical and nonclinical factors, treatment effectiveness, and random chance (fig. 2). In a particular application (*e.g.*, comparison of hospitals), data for one outcome (*e.g.*, postoperative myocardial infarction) is modeled in a regression analysis, with data for all relevant factors (*e.g.*, patient demographics, comorbidities, and type of surgery) entered as candidate predictor variables for each patient. The analysis computes an expected rate of

◆ This Editorial View accompanies the following article: Sessler DI, Sigl JC, Manberg PJ, Kelley SD, Schubert A, Chamoun NG: Broadly applicable risk stratification system for predicting duration of hospitalization and mortality. ANESTHESIOLOGY 2010; 113:1026–37.

Accepted for publication August 13, 2010. The author is not supported by, nor maintains any financial interest in, any commercial activity that may be associated with the topic of this article. The figures in this article were prepared by Annemarie B. Johnson, C.M.I., Medical Illustrator, Wake Forest University School of Medicine Creative Communications, Wake Forest University Medical Center, Winston-Salem, North Carolina.

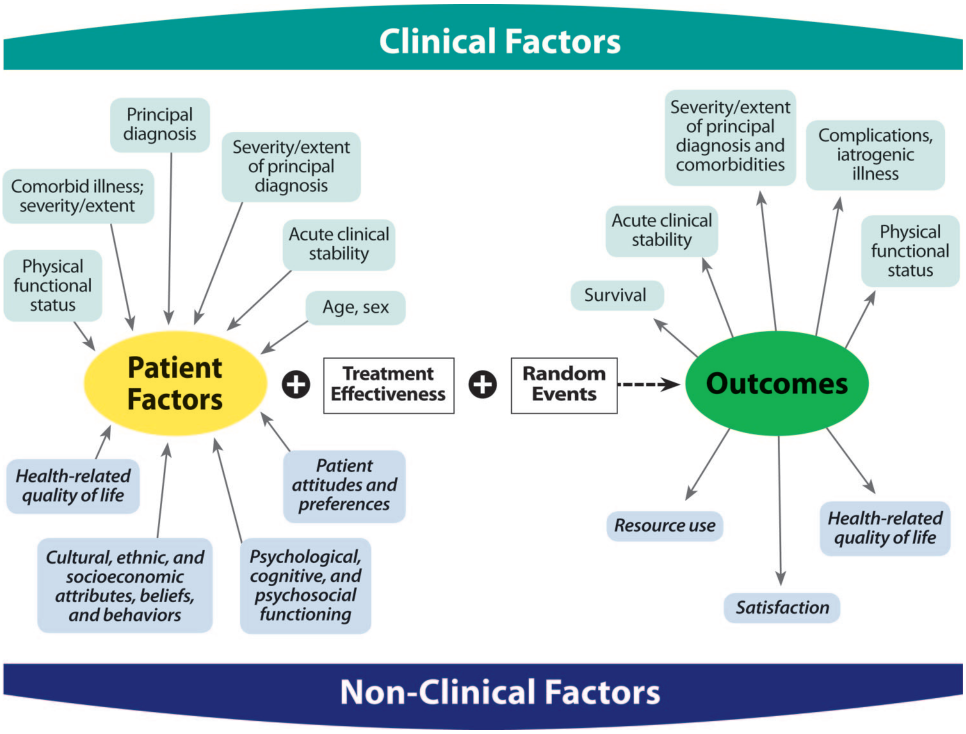
Clinical Factors
Principal diagnosis
Severity/extent of principal diagnosis
Comorbid illness; severity/extent
Acute clinical stability
Physical functional status
Age, sex
Patient Factors
Treatment Effectiveness
Random Events
Outcomes
Health-related quality of life
Patient attitudes and preferences
Cultural, ethnic, and socioeconomic attributes, beliefs, and behaviors
Psychological, cognitive, and psychosocial functioning
Severity/extent of principal diagnosis and comorbidities
Complications, iatrogenic illness
Acute clinical stability
Physical functional status
Survival
Resource use
Health-related quality of life
Satisfaction
Non-Clinical Factors

Fig. 2. A conceptual model in which outcomes of care are a function of patient-related factors, treatment effectiveness, and random chance. Modified, with permission, from Iezzoni LI, ed. Reasons for risk adjustment, Risk Adjustment for Measuring Health Care Outcomes, 3rd ed. Chicago, Health Administration Press, 2003.

myocardial infarction for each hospital's group of patients; if that expected rate is higher than a hospital's observed rate, we might infer that the hospital provides higher quality care. However, the devil is in the details, principally data quality.[2] As in any analysis of observational data, omission of important variables is an ever-present source of bias, as famously demonstrated when Medicare released its initial risk-adjusted, hospital-mortality rankings in 1986: the facility with the most aberrant death rate was a hospice!

In selecting their data source, Sessler *et al.*[1] faced a choice between clinical data and administrative (billing claims) data; the former are rich in clinical detail, yet often plagued by issues of reliability (*e.g.*, terminology not uniformly defined across and even within settings), objectivity (*e.g.*, subjective and even biased assessments), completeness (*e.g.*, missing data), and generally not present in or conducive to electronic format. Administrative data consist of thousands of arcane diagnosis and procedure codes never designed to capture clinical nuances (*e.g.*, important clinical information is absent or variably captured[3]); sometimes coding is temporally imprecise in important clinical distinctions (*e.g.*, was the myocardial infarction a comorbidity or a complication?). Nonetheless, administrative data offer many benefits: very large numbers of patients and facilities (potentially all), uniform data content and format across diverse settings set forth by regulations, capture of care provided throughout a community rather than just specific facilities, potential for tracking patients over time and across settings, generalizeability of risk estimates to most clinical settings, and relatively low cost because the data exist in digital format.

Ideally, Sessler *et al.*[1] might have opted for the high-quality clinical data collected in the National Surgical Quality Improvement Program—begun in Veterans Affairs' hospitals in the early 1990s and extended to academic medical centers and large community hospitals in 2004 by the American College of Surgeons—in which dedicated, trained nurses in each hospital collect preoperative and postoperative, prospectively-defined data for each patient.[4,5] Even though its meager anesthesia-related data might be enhanced by collaboration with the American College of Surgeons, participation entails substantial local costs, less than 10% of U.S. hospitals participate, and risk estimates derived from participating large hospitals may not be generalizeable to nonparticipating small and rural facilities.

Faced with this trade-off between data quality and volume of cases, Sessler *et al.*[1] opted for administrative data, because their goal was a risk-adjustment method "broadly applicable" (generalizable) to all hospitals. Their data were drawn from the 2001–2006 Medicare Provider Analysis and Review database of almost 80 million medical and surgical cases, each with demographic data, length of stay, days from admission to death, and up to 10 diagnosis and 6 procedure codes based on *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM). After appropriate exclusions, their dataset of more than 35 million cases was divided randomly into development and validation datasets. In a Herculean effort, they identified sets of the 5,000 nested ICD-9-CM codes, which optimally predict each of four outcomes: length of stay (1,096 codes), in-patient mortality (184), 30-day mortality (240), and 1-yr mortality (503) Each of the four regression analyses comprises a Risk Stratification Index and is highly predictive of its outcome in the development dataset and, as also might be expected, in the validation dataset. In a more rigorous validation, each Risk Stratification Index performed well—and better than the Charlson Comorbidity Index,[6] a commonly used clinical prediction metric—when applied to a dataset of more than 101,000 of Cleveland Clinic surgical patients aged 18 and older; the addition of patient demographics did not materially enhance prediction of any outcome. The performance of each Risk Stratification Index was also comparable with or better than that of proprietary risk-adjustment methods.[7]

As an extension of their validations, we should expect that their method would be very useful in risk-adjusting outcomes in clinical research, particularly multicentered studies, and other outcome comparisons in which the number of cases at each site exceeds several thousand. With fewer cases, risk estimates are likely to be less reliable and confidence intervals balloon (see their fig. 5). Thus, prospective risk prediction for individual patients would be highly unreliable; this may reflect the trade-off in not using arguably more precise clinical data, because National Surgical Quality Improvement Program data afford individual risk prediction.[8]

Although Sessler *et al.*[1] propose using their method for public reporting of hospital-level outcomes, the notion of report cards is problematic: consumers pay more attention to ratings when buying a toaster than selecting hospitals,[9] possibly as a result of restrictions imposed by their health insurance plans. Moreover, although use of risk-adjusted rankings has been an important advance in quality-improvement projects,[9,10] risk-adjusted mortality is a marginal quality metric across hospitals,[7] possibly because the hospital rescue function is ignored. Patient characteristics predict complications better than they predict death; whether a complication turns into a death reflects the capability of the facility to rescue the patient.[11]

One final caveat: We must remain focused on what risks we wish to adjust for and for what purpose.[2] Almost 25 yr ago, when case-based physician payment was broached,[12,13] the American Society of Anesthesiologists commissioned a simulation of payment-reform proposals using data from disparate anesthesiology practices. Without risk adjustment and with loss of time-based payment, systematic variations in payments were predicted, with anesthesiologists in rural and nonteaching facilities gaining and those in urban or suburban sites losing. After adjustment for surgical complexity, payment was inversely related to duration of surgery.[14] Ability to risk adjust one outcome may have no influence on other outcomes.

Fredrick K. Orkin, M.D., M.B.A., S.M., Department of Anesthesiology, Yale University School of Medicine, New Haven, Connecticut. fred.orkin68@post.harvard.edu

References

1. Sessler DI, Sigl JC, Manberg PJ, Kelley SD, Schubert A, Chamoun NG: A broadly applicable risk stratification system for predicting duration of hospitalization and mortality. ANESTHESIOLOGY 2010; 113:1026–37
2. Iezzoni LI, ed. *Risk Adjustment for Measuring Health Care Outcomes*, 3rd ed. Chicago, Health Administration Press, 2003
3. Lee DS, Donovan L, Austin PC, Gong Y, Liu PP, Rouleau JL, Tu JV: Comparison of coding of heart failure and comorbidities in administrative and clinical data for use in outcomes research. Med Care 2005; 43:182–8
4. Khuri SF, Daley J, Henderson W, Hur K, Demakis J, Aust JB, Chong V, Fabri PJ, Gibbs JO, Grover F, Hammermeister K, Irvin G 3rd, McDonald G, Passaro E Jr, Phillips L, Scamman F, Spencer J, Stremple JF: The Department of Veterans Affairs' NSQIP: The first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. Ann Surg 1998; 228:491–507
5. Khuri SF, Henderson WG, Daley J, Jonasson O, Jones RS, Campbell DA Jr, Fink AS, Mentzer RM Jr, Neumayer L, Hammermeister K, Mosca C, Healey N, Principal Investigators of the Patient Safety in Surgery Study: Successful implementation of the Department of Veterans Affairs' National Surgical Quality Improvement Program in the private sector: The Patient Safety in Surgery study. Ann Surg 2008; 248:329–36
6. Charlson ME, Pompei P, Ales KL, MacKenzie CR: A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. J Chron Dis 1987; 40:373–83
7. Iezzoni LI: The risks of risk adjustment. JAMA 1997; 278:1600–7
8. Cohen ME, Bilimoria KY, Ko CY, Hall BL: Development of an American College of Surgeons National Surgery Quality Improvement Program: Morbidity and mortality risk calculator for colorectal surgery. J Am Coll Surg 2009; 208:1009–16
9. Jha AK, Epstein AM: The predictive accuracy of the New York State coronary artery bypass surgery report-card system. Health Aff (Millwood) 2006; 25:844–55
10. Hall BL, Hamilton BH, Richards K, Bilimoria KY, Cohen ME, Ko CY: Does surgical quality improve in the American College of Surgeons National Surgical Quality Improvement Program: An evaluation of all participating hospitals. Ann Surg 2009; 250:363–76
11. Silber JH, Williams SV, Krakauer H, Schwartz JS: Hospital and patient characteristics associated with death after surgery. A study of adverse occurrence and failure to rescue. Med Care 1992; 30:615–29
12. Mitchell JB: Physician DRGs. N Engl J Med 1985; 313:670–5
13. Mitchell JB, Rosenbach ML: Feasibility of case-based payment for inpatient radiology, anesthesia, and pathology services. Inquiry 1989; 26:458–67
14. Revicki DA, Orkin FK, Luce BR, McMenamin P, Weschler JM: Physician payment reform: Anesthesiology as a case study. ANESTHESIOLOGY 1990; 73:760–9