A Critique of Published Hospital Rankings that are Based on Medicare Administrative Data

Lawrence I. Bonchek, M.D., F.A.C.C., F.A.C.S.
Editor in Chief

INTRODUCTION

The public availability of administrative data from Medicare has resulted in the periodic release of hospital "rankings" by various media groups. It is vital to realize that these differ in critically important ways from reports by professional organizations, such as the Society of Thoracic Surgeons' (STS) Adult Cardiac Surgery National Database, and by government agencies, such as the reports of the Pennsylvania Health Care Cost Containment Council (PHC4), which are based on clinical data.

The fundamental challenge of all hospital "ranking" systems is to develop a statistical model that compensates for the different risks of patients treated at different institutions. In comparison with for-profit specialty hospitals, for example, large non-profit tertiary and quaternary acute care hospitals are likely to treat a disproportionate share of high risk patients who are sicker, older, and have more (and more serious) co-morbidities. Thankfully we are beyond the days when raw mortality data were released; all "ranking" systems now attempt to construct statistical models that adjust for differences in risk. But to do so, even imprecisely, requires clinical data about large numbers of patients, not merely administrative billing data from Medicare. For if one hospital's patients average 75 years of age, and another's average 65 years, exactly how much should that age difference prolong the average length of stay? Exactly how many deaths per 100 admissions can that age difference account for? Perhaps those numbers could be calculated as *isolated variables* by objective analysis of *clinical data* from a very large population of patients, but as the number of interacting co-morbidities increases they become increasingly imprecise.

Since media "rankings" based on publicly available Medicare data don't have access to enough relevant clinical information, they use whatever information they can get. For example, length-of-stay may be used as a surrogate for complications on the assumption that complications prolong the LOS, even though many non-morbid factors can influence LOS. That's why it is nearly impossible to develop an accurate risk-adjustment model from the Medicare administrative data that are available to the public.

RATINGS BASED ON CLINICAL DATA

In contrast, the STS Adult Cardiac Surgery National Database, which first enrolled patients in 1990, is the outstanding example of how a professional organization can develop an accurate risk adjustment model by objective analysis of clinical "big data."[1] That database now contains more than 4.5 million surgical records, representing an estimated 94 percent of all adult cardiac surgery centers across the U.S.* These centers submit clinical data *voluntarily* as a means to quality improvement—a commitment that requires considerable time and expense for data collection, since there are more than 500 data points for each patient. Data are sent regularly to the Duke University Clinical Research Institute for analysis and ongoing refinement of the risk-adjustment statistical model. And even though gaming of data has never been an overt problem in this database, years ago the STS took the proactive step of instituting random audits, in which at least 10 programs per year are randomly selected for actual chart review by an independent auditing firm to verify accurate and complete data collection.

In the interests of transparency, the Society has also established STS Public Reporting Online which publishes (on www.sts.org) CABG composite quality data from more than 400 Database participants who have volunteered thus far to participate. Since 2009, STS has collaborated with *Consumer Reports* (CR), which now presents the STS CABG composite star ratings on the health section of its website.[2]

---

* Lancaster General Hospital had a clinical database for cardiac surgery from the first open heart operation here in 1983. In 1992 we became early participants in the STS Database, and one of the largest members at the time, since we transferred the clinical records of our first 3,500 patients into the STS Database.

Taking a cue from the STS, in 1997 the American College of Cardiology initiated the voluntary ACC National Cardiovascular Data Base for catheterization laboratory procedures, and has since expanded to other aspects of cardiovascular care.

These sterling examples of voluntary clinical databases run by professional societies make it all the more disappointing when government agencies or the media initiate "ranking" systems with poorly conceived and statistically invalid models. Still, when government agencies initiate reporting systems, at least they are a sincere effort to improve the quality of health care. (For example, though PHC4's statistical models were severely criticized at first, they have been steadily improved.)

Media "rankings," on the other hand, though offered with a professed concern for the public interest and a desire to disseminate useful information, are often presented in a provocative style that—rather than enlightening and informing—seems intended to stimulate controversy (and presumably subscriptions). A recent *Consumer Reports* article that purports to "rank" surgical results at U.S. hospitals begins with the sentences "Surgery is scary. It usually involves having your body cut open, and sometimes things go wrong."[3] That approach seems more likely to frighten people than to inform them.

Of course when a media "ranking" such as that of *U.S. News*[4] gives LGH high marks that corroborate what our own extensive system of internal quality controls tells us, we cannot avoid taking credit for what we feel is an accurate assessment, even if the risk-adjustment model seems to lack some risk factors and overemphasize other parameters. But when the results of a contrived "ranking" system deviate strikingly from our own assessments of our services, as well as our high ranking in the U.S. News report, we are obliged to see whether there might be internal reasons for the discrepancy—which would imply that we need to make some internal changes—or whether the ranking system is too flawed to be meaningful.

THE PROBLEMS OF MEDIA RANKINGS

In the case of the aforementioned "ranking" of surgical services by *Consumer Reports*,[5] our suspicion of flawed methodology is aroused not only by our own below average "ranking," but by the fact that many of the finest acute care hospitals in the country received similar rankings: Massachusetts General Hospital, Brigham and Women's Hospital, the Lahey Clinic, Johns Hopkins, etc. Indeed, if we are below average, we are in very good company!

How could a ranking system that purports to adjust for differences in risk produce such clearly erroneous results? It seems that the CR ranking system, which was developed for CR by a private consulting group, has both the general shortcomings inherent in "rankings" based entirely on Medicare administrative (rather than clinical) data, along with flaws that are uniquely its own. To understand how so many of America's finest institutions could be ranked below average requires a detailed explanation of the CR methodology.

THE CONSUMER REPORTS RATING OF SURGICAL RESULTS

Based on Medicare data for 2009-2011, CR looked at how "hospitals nationwide compare in avoiding adverse events in Medicare patients during their hospital stay for surgery." They evaluated "27 categories of scheduled surgeries" by looking at death and length of stay (LOS) as outcome measures, and they assumed that these were adequate surrogates for complications. They also attempted to adjust for risk based on a few administrative parameters: age, gender, and a few easily recorded health conditions such as high blood pressure and diabetes.

The CR methodology immediately raises many red flags:

1. They did no direct assessment of complication rates.

2. They did not assess readmission rates, which Medicare now uses as a principal indicator of quality and determinant of reimbursement. One cause of readmission is discharging a patient too soon in order to artificially lower LOS. Also, most deaths after discharge would have been undetected.

3. They apparently looked only at inpatient surgery, which fails to allow for the very large number of outpatient procedures done at hospitals like LGH. Outpatients almost never die as a result of the procedure, and their LOS is zero days; both of these statistics, if included, would markedly lower the mortality and average LOS.

4. They lumped together all types of institutions: multi-specialty referral centers like St. Mary's, the principal hospital of the Mayo Clinic; acute care, full service hospitals like LGH; and short-stay, often physician-owned, specialty hospitals that pre-select their patients since they cannot provide full-service care to patients with multiple, severe co-morbidities. The crude CR risk-adjustment model cannot compensate for such extremely different patient populations.

5. Many of the smaller hospitals they rated did not do all the procedures that were evaluated. Even though these smaller or more specialized institutions lacked many (or any) complex, riskier procedures, these hospitals were ranked solely on the basis of the results of the less complex procedures they did perform. (No risk-adjustment model can compensate for missing data.)

6. The CR risk-adjustment model only included a rough assessment of a few co-morbidities. It matters greatly whether "high blood pressure" and "diabetes" are mild or severe, but the distinctions are not available in Medicare administrative data.

7. CR included angioplasty as "surgery to remove blockages in arteries in the heart," but angioplasty is not surgery and it doesn't ordinarily "remove" anything. The inclusion of angioplasty in a review of surgical results makes one wonder about the design of this entire project.

8. They used outdated information from 2009-2011. There has been considerable attention focused on LOS since then, and many routines surrounding hospital discharge have changed dramatically. Since 2009, LGH has reduced its LOS by 20%, not because of improvements in surgical results (despite what CR may think), but because of a concerted effort to improve procedural efficiency within the hospital's clinical units and laboratories.

CONCLUSIONS

In defending flawed models, publishers often assert that "poor data are better than no data at all," and "we have to start somewhere." Actually, bad data are worse than no data at all for at least two reasons. One, bad data besmirches the reputations of the most capable institutions that care for the most complicated patients. As a result, some patients with complex problems may go to less capable institutions with higher "rankings," and may not receive the most advanced care available for their problem. Furthermore, in states with highly publicized "rankings," some programs may turn away some high risk patients if they fear that the risk-adjustment model is too imperfect, and poor results with risky patients would adversely affect their ranking. This phenomenon was disturbingly apparent when the NY State ratings for cardiac surgery were first published, because some institutions reported fewer high risk patients in the second year.

A second problem with imperfect risk-adjustment models is that they encourage some programs to rationalize gaming the system by up-coding the severity of illness in their patients, thus achieving a higher risk rating for their population. For example, in NY state the incidence of COPD in some hospitals increased strikingly from year 1 to year 2 of the rankings. Fortunately, this was quickly noted and just as quickly suppressed.

Though these phenomena were noted in NY, it could happen anywhere. Fortunately, as the risk-adjustment model improved in NY, these problems also dissipated.

In regard to LGH, any public report, even if it is as flawed as the Consumer Report ranking, should and will prompt a careful review of our own practices, procedures, and outcomes; if improvements are needed, they will be initiated.** But at the same time, we should all understand that this intentionally provocative report from CR was not a scientific analysis of outcomes. In the introduction to their report, CR says that "Up to 30% of patients suffer infections, heart attacks, strokes, or other complications after surgery..." Yet, their own analysis only used surrogate measures of these complications! Clearly, their risk-adjustment model was seriously deficient, and—even as we review our outcomes—we should not consider their ranking an indicator of systemic problems.

ALSO IN THIS ISSUE

The articles in this issue are characterized by a remarkable degree of timeliness.

Dr. Jon Bentz delves deeply into the science behind the headlines about concussion and offers a rational perspective that contrasts with all the media hype and distortions. His article is especially timely because in the past few weeks alone the NFL has agreed to a $765 million settlement with more than 4,500 former players who had accused the league of hiding the dangers of brain injury while profiting from the sport's violence. In addition to compensation for the players, the

---

**LGH is now participating in the American College of Surgeons National Surgical Quality Improvement Program (NSQUIP) database. LGH will submit complication data for general surgery and vascular surgery based on clinical chart review (not administrative billing data) and the results will be benchmarked nationally. We will also submit specific clinical data on each patient (similar to the STS database) so that the results are adjusted for risk and case-mix to provide more accurate national benchmarking.

league agreed to fund medical exams and a program of medical research.

Of course this does not end the story from a legal perspective since additional lawsuits have already been filed by players who did not participate in the original lawsuit. But much more importantly, this does not end risk to players, especially since at levels below the NFL there has been much less discussion of what happens behind the scenes. From grade school football to big time Division I College football, concussions continue to occur, and decisions are made by coaches and trainers about when and whether to put key players back into a game. There are horrendous conflicts of interest involved, particularly when the trainers responsible for on-field decisions about the condition of players report to coaches responsible for winning games. The Chronicle of Higher Education website reports on a survey[7] sent to hundreds of athletic trainers and training-staff members from the NCAA's 120 largest football programs.

The respondents included 101 head athletic trainers, head football trainers, and other sports-medicine professionals from the highest rung of college football, the NCAA's Football Bowl Subdivision. Out of the 101 who responded:

11 said they reported directly to the football coach or a member of the coaching staff;

32 said a member of the football coaching staff had influence over hiring and firing decisions for their position;

53 said they had felt pressure from football coaches to return a student to play faster than they thought was in his best interest medically;

42 said they had felt pressure from football coaches to return an athlete to the field even after he suffered a concussion.

According to the *Chronicle*, "When trainers push back too hard, they often face repercussions. More than a dozen Division I athletic trainers have been fired or demoted in recent years, often over questionable return-to-play calls."

Also in this issue, and equally timely, is a discussion of LGH's rationale for not hiring smokers. This is not quite as simple an issue as it seems, and there is a school of thought that opposes such a policy. Mary Miskey, VP of Human Resources Operations, provides a comprehensive analysis of this subject.

Dr. Joseph Kontra discusses a concern that is continuously in the headlines: the looming specter of "super-bugs" that are resistant to all antibiotics, and the dearth of research to develop new ones. In addition to exploring the reasons for this dangerous situation, he offers considerable encouragement and a pathway toward resolution of this dilemma.

Tina Davis, MSN, CRNP and Drs. Rolf Andersen and Jose Ibarra from The Heart Group of Lancaster General Health, describe a research study that is now enrolling patients in a trial of a new pharmacological approach to patients with hyperlipidemia that is difficult to manage with conventional therapy.

Dr. Alan Peterson continues his illuminating discussion of guidelines and offers some additional Top Tips.

And finally, in an effort to complement these extraordinarily timely articles, I added a historical article that we published some time ago, and refers to events that were timely in the 19th century!

I hope you find this balance of timely articles both educational and entertaining.

REFERENCES

1. Wimer, PE. Evolution of the society of thoracic surgeons national cardiac surgery database. J Lanc Gen Hosp. 2009; 4:113-118. (www.jlgh.org/Past-Issues/Volume-4--Issue-3/Evolution-of-the-Society-of-Thoracic-Surgeons.aspx)

2. www.sts.org/news/sts-partners-consumer-reports-national-heart-surgery-ratings

3. www.consumerreports.org/content/cro/en/consumer-reports-magazine/z2013/September/yourSaferSurgerySurvivalGuide.print.html

4. health.usnews.com/best-hospitals/rankings

5. health.usnews.com/best-hospitals/area/pa/lancaster-general-hospital-6231120

6. www.consumerreports.org/health/resources/pdf/surgery-ratings-national-list/SurgeryRatingsRev.pdf

7. chronicle.com/article/Trainers-Butt-Heads-With/141333/