

MARKET WATCH

Choosing The Best Hospital: The Limitations Of Public Quality Reporting

Despite multiple and often conflicting ratings, public reporting on hospital quality seems to have gained a permanent foothold.

by Michael B. Rothberg, Elizabeth Morsi, Evan M. Benjamin, Penelope S. Pekow, and Peter K. Lindenauer

ABSTRACT: The call for accountability in health care quality has fueled the development of consumer-oriented Web sites that provide hospital ratings. Taking the consumer perspective, we compared five Web sites to assess the level of agreement in their rankings of local hospitals for four diagnoses. The sites assessed different measures of structure, process, and outcomes and did not use consistent patient definitions or reporting periods. Consequently, they failed to agree on hospital rankings within any diagnosis, even when using the same metric (such as mortality). In their current state, rating services appear likely to confuse, rather than inform, consumers. [*Health Affairs* 27, no. 6 (2008): 1680–1687; 10.1377/hlthaff.27.6.1680]

DRIVEN BY A DESIRE TO increase the transparency and accountability of the U.S. health care system, hospitals face growing requirements to release performance data to the public. For more than a century, however, public reporting has been controversial, frequently criticized for being unreliable and for failing to capture the true complexity of patients' conditions.¹

Despite these concerns, in the face of rapidly rising costs and striking gaps in the quality of health care, consumers are being asked to make better-informed health care choices, with the expectation that market forces will drive down prices and improve quality. High-deductible insurance plans and health savings

accounts (HSAs) are predicated on this idea, despite a lack of evidence that health care consumers make choices based on performance data.² Even if consumers do not use the performance data, their public release appears to stimulate quality improvement initiatives by hospitals, perhaps by appealing to the professional ethos of clinicians, and out of fear that poor performance would threaten the reputation of the organization.³

Over the past decade, a number of hospital rating services have emerged, offering free Web sites aimed at health care consumers. The services report on multiple aspects of health care quality, including those reflecting structural aspects of care, processes, and outcomes.

.....
Michael Rothberg (Michael.Rothberg@bhs.org) is an assistant professor of medicine at Tufts University School of Medicine, Baystate Medical Center, in Springfield, Massachusetts. Elizabeth Morsi is a research assistant in the medical center, and Evan Benjamin is vice president for quality. Penelope Pekow is director of the Center for Research and Education in Women's Health at the University of Massachusetts School of Public Health in Amherst. Peter Lindenauer is an associate professor at the Tufts University School of Medicine.

For consumers to make good decisions, however, publicly reported information must be accessible, interpretable, and consistent.⁴ In this context, it remains unclear whether consumers can correctly interpret the available data or whether such data would lead to consistent health care decisions. Moreover, the spectrum of performance measurement has been limited to few diseases, so most patients will not find quality information about their condition. Information about related conditions might be helpful, but only if performance in one disease area predicts performance in another. Taking the perspective of a patient seeking to choose the best hospital, we examined five reporting services and evaluated their level of agreement across hospitals for four conditions.

Study Data And Methods

In August 2007 we identified all hospitals within a thirty-mile radius of Boston, Massachusetts. We focused our analysis on hospitals competing for consumers who would be willing to travel up to one hour to receive high-quality care. To create a manageable sample, we limited our analysis to nonspecialty hospitals with at least 250 beds, because we did not believe that consumers would travel to receive care in a community hospital.⁵

We identified five well-known public reporting services: Hospital Compare, HealthGrades, Leapfrog Group, *U.S. News and World Report*, and Massachusetts Healthcare Quality and Cost (Mass QC, a state-run service sponsored by the state's executive office of health and human services). First, we compared the services based on the number of diagnoses or procedures reported; the use of structure, process, and outcome measures; the reporting time period; the patient population used to derive the ratings; and the method of presentation.

We then selected four nonemergent conditions to compare across the databases: community-acquired pneumonia, total hip replacement, percutaneous coronary intervention (PCI), and coronary artery bypass grafting (CABG). In a separate analysis, we com-

pared adjusted mortality for acute myocardial infarction, because three of the five services reported this measure. For each condition, we ranked all hospitals within each database, using the measures provided by that database. We evaluated association among pairs of raters for the same diagnosis via Spearman rank correlations of hospital scores; Spearman correlations were also used to evaluate the association of ratings for pairs of diagnoses within each hospital.

Study Results

■ Overview of hospital rating services.

The number of conditions reported on by the rating services ranged from four to thirty-two, with the mean and median both fourteen (Exhibit 1). Of the three types of quality measures—structure, process, and outcomes—most of the rating services focused on only one, although some reported measures in two spheres. Only Leapfrog reported measures in all three. Three of the five rating services reported outcome measures, all of which included risk-adjusted mortality. Although we examined all services on the same date, the time period on which the rankings were based varied in both duration and start year. There was even variation in time periods within the same rating system: *U.S. News and World Report* based its 2007 ratings on mortality data for fiscal years 2003–2005 and physician survey data for calendar years 2005–2007. Four of the services used at least some data that were two to four years old. Three services limited their data to Medicare or Medicaid patients. Finally, three services incorporated reports from other services: HealthGrades reported some Leapfrog measures, Mass QC reported Hospital Compare measures, and Leapfrog linked to Hospital Compare.

■ **Hospital rankings.** Of twenty-five general hospitals within thirty miles of Boston, nine had at least 250 beds. For any given diagnosis, there was little overall agreement among the rating services.⁶ For example, the two hospitals ranked first for CABG by at least one service were also ranked fourth and last by another (Exhibit 2). Conversely, the two hospi-

EXHIBIT 1 Comparisons Of Five Hospital Quality Rating Services: Summary

Rating service	No. of conditions scored	Structure measures	Process measures	Outcome measures	Time period	Patient population
HealthGrades	32	Presence of newborn ICU	Leapfrog safe-practices score	Mortality, major complications, volume of vaginal and c-section deliveries, newborn mortality, women's health rating	2003–2005	Medicaid, Medicare
Leapfrog	8	CPOE, ICU staffing	Safe-practices score, procedure volumes	PCI, CABG, and AAA mortality	2007 (mortality: 2006)	All patients
Hospital Compare	4	None	Antibiotic use in surgery, 5 measures for pneumonia, 8 measures for heart attack, 4 measures for heart failure	Adjusted mortality	2005–2006	Medicaid, Medicare
Mass QC	14	None	Provided recommended care	Cost, risk-adjusted mortality, length-of-stay, total c-sections and VBAC utilization	FY 2004 and FY 2005	All patients
U.S. News	16	Advanced services, trauma center, magnet hospital, epilepsy center certification, FACT accreditation, nursing ratio, patient service index, NCI cancer center, total volume	Reputation based on physicians surveyed	Risk-adjusted mortality, total volume	2005–2007 survey (mortality: FY 2003–2005)	Medicare

SOURCE: Results derived by authors from providers' Web sites.

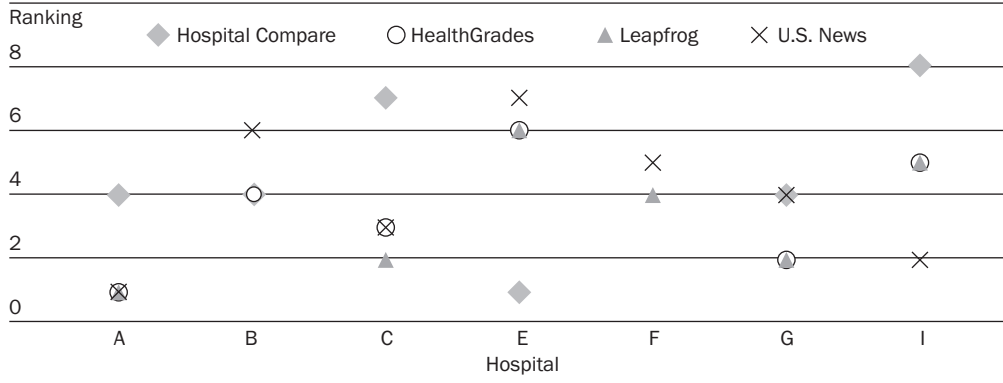
NOTES: ICU is intensive care unit. CPOE is computerized physician order entry. PCI is percutaneous coronary intervention. CABG is coronary artery bypass graft. AAA is abdominal aortic aneurysm. VBAC is vaginal birth after caesarean. FACT is Foundation for the Accreditation of Cellular Therapy. NCI is National Cancer Institute. Additional information and more detailed definitions are available in Appendix Exhibit 1, online at <http://content.healthaffairs.org/cgi/content/full/27/6/1680/DC1>.

tals that were ranked last for CABG by at least one service were ranked first or second by another. Spearman rank correlation among pairs of services within a diagnosis was lowest for Mass QC and Leapfrog rating of PCI ($r = -0.67$, $p = 0.14$), and best for HealthGrades and Leapfrog ratings of both PCI and CABG ($r = 0.97$, $p < 0.005$). The strong agreement between Leapfrog and HealthGrades was not seen for total hip replacement ($r = -0.41$, $p = 0.42$).

Within a given rating service, there was little

agreement among the ratings for a single hospital across pairs of diagnoses, except for Leapfrog and Hospital Compare. Leapfrog's mostly structural measures tended to be the same for each condition ($r = 0.97$ to $r = 1.00$, $p < 0.001$); Hospital Compare offered a single measure—preoperative antibiotic timing—for all surgical procedures, so its surgical rankings were perfectly correlated ($r = 1.00$, $p < 0.0001$), but none of these was correlated with community-acquired pneumonia ($r = 0.14$, $p = 0.73$).

EXHIBIT 2
Seven Hospitals' Rankings For Coronary Artery Bypass Grafting (CABG) Based On Publicly Reported Ratings Using Four Different Hospital Quality Rating Systems



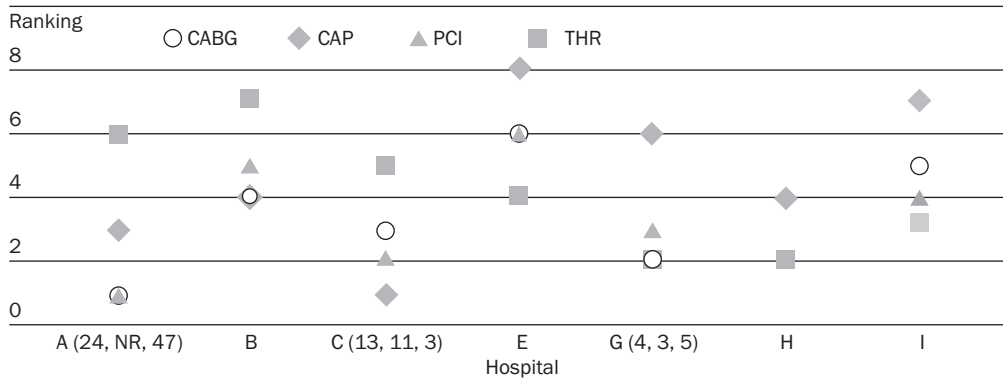
SOURCE: Authors' computations based on providers' Web sites.

HealthGrades showed a strong mortality correlation between CABG and PCI ($r = 0.89, p = 0.019$) but not between CABG and total hip replacement ($r = -0.14, p = 0.79$). No hospital scored in either the top or bottom half on all four diagnoses (Exhibit 3).

■ **Mortality rates.** Two services provided mortality rates for all of the diagnoses, and three services provided mortality rates for AMI (Exhibit 4). Neither the observed mortality rates nor the observed/predicted rates were

consistent across services. The hospital ranked first by HealthGrades had the second-highest observed mortality; it ranked seventh according to Mass QC. Conversely, HealthGrades' seventh-ranked hospital (the only hospital statistically worse than average) was ranked first by Mass QC. This same hospital was ranked fifth in the nation by *U.S. News and World Report* for cardiology.

EXHIBIT 3
Seven Hospitals' Rankings For Four Different Conditions Based On Publicly Reported Mortality Ratings From HealthGrades



SOURCE: Authors' computations based on providers' Web sites.

NOTES: Hip replacement rankings are based on complication rates instead of mortality. Numbers indicated the hospital rank for respiratory disorders, orthopedics, and heart by *US News and World Report*. CABG is coronary artery bypass grafting. CAP is community-acquired pneumonia. PCI is percutaneous coronary intervention. THR is total hip replacement. NR is not ranked.

EXHIBIT 4 Acute Myocardial Infarction Mortality And Ratings From Three Hospital Quality Rating Services

Rating service									
Hospital	HealthGrades				Mass QC				Hospital Compare
	Obs.	Pred.	Ranking ^a	Rating ^b	Obs.	Pred.	Ranking	Rating	Rating
A	6.61	7.16	2	**	5.45	8.11	2	**	**
B	7.95	7.14	5	**	6.64	9.54	3	**	**
C	6.53	6.35	3	**	5.57	7.93	4	**	**
D	20.51	18.85	4	**	15.56	14.65	9	**	**
E	9.26	7.81	8	**	10.47	12.06	5	**	**
F	- ^c	- ^c	- ^d	- ^c	9.5	8.99	8	**	**
G	7.83	6.82	7	*	5.57	9.18	1	**	**
H	19.46	22.27	1	**	17.65	20.25	7	**	**
I	7.56	6.79	5	**	6.54	7.53	6	**	**

SOURCE: Providers' Web sites.

^aRanking is based on observed/predicted mortality.

^bRating compares the adjusted mortality to the mean. Two stars mean not statistically different from average, and one star means worse than average. Three stars (not pictured) means better than average.

^cNot available.

^dNot applicable.

Discussion

■ **Lack of consistent agreement.** In an environment of increasing transparency, in which patients are taking an active role in choosing health care providers and assuming more financial responsibility, accurate ratings of hospitals are essential. In this study of five leading health care rating services, we found that these services failed to consistently agree on either top- or bottom-performing hospitals in a single metropolitan area. Among the nine institutions studied, hospitals ranked first or second by one system were often ranked seventh or eighth by another. These findings can be explained by variations in the methods used by the rating systems, which measured different processes and outcomes, examined widely different time periods, and included different patient populations. Even when two services simultaneously measured the same outcome (mortality), agreement among the services was poor. Most rating systems did not perform statistical tests, but when they did, all nine hospitals were indistinguishable. In fact, only one hospital (out of seventy-one) in the state

had cardiac mortality that was statistically better than the mean.

■ **Previous studies of rating services.** Studies of rating services have generally been limited to two-way comparisons and have included hospitals nationwide. *U.S. News and World Report's* "Best Hospitals" has been compared to Hospital Compare, the Cooperative Cardiovascular Project, and risk-standardized mortality rates.⁷ "Best Hospitals" as a group usually performed better on both process measures and mortality, although some individual "Best Hospitals" performed poorly, so a "Best Hospital" rating does not ensure quality. Our sample included several "Best Hospitals," none of which performed consistently well. Comparing HealthGrades to the Cooperative Cardiovascular Project produced similar conclusions.⁸ The process measures from Hospital Compare are inversely related, albeit weakly, to standardized mortality rates.⁹ We found that not only were mortality rankings poorly correlated with process measures, but they were poorly correlated with other rating services' mortality rankings.

■ **Study limitations.** Our study has a

number of limitations. First, data were collected from large hospitals located in a single metropolitan area with a high concentration of hospitals. Most health care consumers do not have as much choice, although today patients are encouraged to move beyond traditional geographic boundaries in search of better care. In this context, ratings are even more essential. Second, our data are limited to a single point in time, while health care rating systems are in rapid continuous evolution.

However, our data are representative of the information available to the public in 2007. Third, we used unweighted rankings instead of absolute scores. Adding weights might produce different rankings, but the rating services themselves did not identify any measure as more important than another, nor does the current science support using any particular measure over another. Finally, we did not account for the magnitude of differences between hospitals, and we do not know whether patients would take this into account. For this reason, we also examined mortality measures for which there was statistical testing and found that the systems generally could not discriminate among hospitals, demonstrating the limitations of such measures.

■ **Barriers patients face.** Public reporting has the potential to be a powerful agent for quality improvement. Patients maintain that they would use mortality data in making decisions; yet despite the wide availability of mortality reports, few patients actually use them, and even the best public reports do not seem to affect market share or consumers' choices.¹⁰ This may change as consumers become more aware of the rating services and as high-deductible plans drive patients to seek care beyond their local hospital. Nevertheless, the existence of multiple and often conflicting reports can only further complicate the relationship between public reporting and consumer choice.

We identified several additional barriers patients face. First, to rank the hospitals, it is

necessary to interpret the data, but the services offer no guidance as to the relative benefit of specific measures. For example, when considering pneumonia, is it more important to receive antibiotics within four hours or to receive the correct antibiotic? In cardiac surgery, is it more important to choose a high-volume hospital or one that has an electronic health record and a high safe-practices score? Second, even seemingly straightforward measures such as mortality rates vary by rating

service, because there is no standardized risk-adjustment methodology.¹¹

■ **Recommendations.**

Despite these limitations, public reporting seems to have gained a permanent foothold; it stimulates quality improvement activities, al-

though its effects on outcomes are mixed.¹² Unfortunately, proliferation of conflicting hospital rating services may dilute the impact of low scores, because poor performers on one service's measures can still claim to be rated as top performers by another.¹³

How, then, might hospitals and reporting sites join forces to offer information that is useful, accurate, and consistent? First, instead of passively accepting reporting mandates, hospitals must embrace public reporting and help design measures that can be efficiently collected and that represent the true quality of care offered. The Hospital Quality Alliance (HQA) represents a step in this direction, but its early measures focus on processes that, although laudable, are not of particular interest to patients and have relatively weak links to outcomes.¹⁴ The National Surgical Quality Improvement Program (NSQIP) is another example of hospitals' taking the initiative to measure and share outcome data for the purposes of improving quality.¹⁵ Although these measures are not currently available to the public, in the future they could be.

Patients, however, may want different information. Qualitative studies suggest that patients want to know about other patients' experiences and that they are willing to choose

“Even the best public reports do not seem to affect market share or consumers’ choices.”

physicians based on that information.¹⁶ Patients also care about patient satisfaction, being cared for, and having their physical and informational needs met.¹⁷ Some of these concerns are addressed though the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey, which was recently added to Hospital Compare. Alternatively, rapidly growing social networking Web sites may eclipse traditional public reporting for both physician and hospital rating.¹⁸ More research is needed, on both what patients want and what they can understand. In the future, reporting mandates should incorporate new quality assessments that are meaningful to patients; these should focus on both disease-specific measures (for example, functional status following stroke) and hospital-specific measures (for example, rates of hospital-acquired infections). Such efforts would benefit from the participation of providers and consumer advocacy groups.

Finally, rating services must solve the problems of risk adjustment and random variation. Standardizing risk-adjustment methodology and reporting periods would aid in transparency and level the playing field for providers; also, allowing consumers to compare two hospitals directly might reveal statistical differences that are obscured when comparing each hospital to the national or state mean.

OUR STUDY demonstrates that sometimes more is less. Providing more measures may actually obscure comparisons, because consumers may be overwhelmed by confusing and conflicting information. If public reporting is to succeed in improving health care, hospitals should join with raters to provide stakeholders with data that are consistent and easily interpretable. As a first step, many of the rating services now include other services' data on their Web sites, providing a range of measures of structure, process, and outcomes. Still missing is the means to prioritize and synthesize these measures. Until this is achieved, public reporting may remain of limited value to those whom it was designed to help.

.....
Michael Rothberg is the recipient of a Doris Duke Clinical Scientist Development Award.

NOTES

1. F. Nightingale, *Notes on Hospitals*, 3d ed. (London: Longman Green, 1863); L.K. Altman, "Report on Mortality: Guarded Praise," *New York Times*, 18 December 1987; and E.C. Schneider and A.M. Epstein, "Use of Public Performance Reports: A Survey of Patients Undergoing Cardiac Surgery," *Journal of the American Medical Association* 279, no. 20 (1998): 1638–1642.
2. J.H. Hibbard, J. Stockard, and M. Tusler, "Hospital Performance Reports: Impact on Quality, Market Share, and Reputation," *Health Affairs* 24, no. 4 (2005): 1150–1160; and M.N. Marshall et al., "The Public Release of Performance Data: What Do We Expect to Gain? A Review of the Evidence," *Journal of the American Medical Association* 283, no. 14 (2000): 1866–1874.
3. *Ibid.*
4. Hibbard et al., "Hospital Performance Reports"; and L.M. Schwartz, S. Woloshin, and J.D. Birkmeyer, "How Do Elderly Patients Decide Where to Go for Major Surgery? Telephone Interview Survey," *British Medical Journal* 331, no. 7520 (2005): 821.
5. A more detailed description of the methods appears in the online Technical Appendix, <http://content.healthaffairs.org/cgi/content/full/27/6/1680/DCL>.
6. The rankings of the hospitals for each of the four conditions studied appear in Appendix Exhibit 2; *ibid.*
7. L.K. Halasyamani and M.M. Davis, "Conflicting Measures of Hospital Quality: Ratings from 'Hospital Compare' versus 'Best Hospitals,'" *Journal of Hospital Medicine* 2, no. 3 (2007): 128–134; J. Chen et al., "Do 'America's Best Hospitals' Perform Better for Acute Myocardial Infarction?" *New England Journal of Medicine* 340, no. 4 (1999): 286–292; and O.J. Wang et al., "America's Best Hospitals' in the Treatment of Acute Myocardial Infarction," *Archives of Internal Medicine* 167, no. 13 (2007): 1345–1351.
8. H.M. Krumholz et al., "Evaluation of a Consumer-Oriented Internet Health Care Report Card: The Risk of Quality Ratings Based on Mortality Data," *Journal of the American Medical Association* 287, no. 10 (2002): 1277–1287.
9. R.M. Werner and E.T. Bradlow, "Relationship between Medicare's Hospital Compare Performance Measures and Mortality Rates," *Journal of the American Medical Association* 296, no. 22 (2006): 2694–2702; and A.K. Jha et al., "The Inverse Rela-

- tionship between Mortality Rates and Performance in the Hospital Quality Alliance Measures," *Health Affairs* 26, no. 4 (2007): 1104–1110.
10. Schneider et al., "Use of Public Performance Reports"; Hibbard et al., "Hospital Performance Reports"; Romano et al., "Do Well-Publicized Risk-Adjusted Outcomes Reports Affect Hospital Volume?"; Schwartz et al., "How Do Elderly Patients Decide Where to Go?"; and DW. Baker et al., "The Effect of Publicly Reporting Hospital Performance on Market Share and Risk-Adjusted Mortality at High-Mortality Hospitals," *Medical Care* 41, no. 6 (2003): 729–740.
 11. L.I. Iezzoni, "The Risks of Risk Adjustment," *Journal of the American Medical Association* 278, no. 19 (1997): 1600–1607; and R. Behal, "The Lake Wobegon Effect: When All the Patients Are Sicker," *American Journal of Medical Quality* 21, no. 6 (2006): 365–366.
 12. C.H. Fung et al., "Systematic Review: The Evidence That Publishing Patient Care Performance Data Improves Quality of Care," *Annals of Internal Medicine* 148, no. 2 (2008): 111–123.
 13. P.J. Pronovost, M. Miller, and R.M. Wachter, "The GAAP in Quality Measurement and Reporting," *Journal of the American Medical Association* 298, no. 15 (2007): 1800–1802.
 14. Werner and Bradlow, "Relationship between Medicare's Hospital Compare Performance Measures and Mortality Rates"; and E.H. Bradley et al., "Hospital Quality for Acute Myocardial Infarction: Correlation among Process Measures and Relationship with Short-Term Mortality," *Journal of the American Medical Association* 296, no. 1 (2006): 72–78.
 15. K.S. Rowell et al., "Use of National Surgical Quality Improvement Data as a Catalyst for Quality Improvement," *Journal of the American College of Surgeons* 204, no. 6 (2007): 1293–1300.
 16. S.A. Richard, S. Rawal, and D.K. Martin, "Patients' Views about Cardiac Report Cards: A Qualitative Study," *Canadian Journal of Cardiology* 21, no. 11 (2005): 943–947; and G. Fanjiang et al., "Providing Patients Web-Based Data to Inform Physician Choice: If You Build It, Will They Come?" *Journal of General Internal Medicine* 22, no. 10 (2007): 1463–1466.
 17. L.V. Doering, A.W. McGuire, and D. Rourke, "Recovering from Cardiac Surgery: What Patients Want You to Know," *American Journal of Critical Care* 11, no. 4 (2002): 333–343.
 18. L. Landro, "Social Networking Comes to Health Care," *Wall Street Journal*, 27 December 2006.