VIEWPOINT

The New CMS Hospital Quality Star Ratings
The Stars Are Not Aligned

Karl Y. Bilimoria, MD, MS
Surgical Outcomes and Quality Improvement Center (SOQIC), Department of Surgery and Center for Healthcare Studies, Feinberg School of Medicine, Northwestern University, Chicago, Illinois.

Cynthia Barnard, MBA, PhD
Quality Division, Northwestern Medicine, Chicago, Illinois.

Author Audio Interview

Corresponding Author: Karl Y. Bilimoria, MD, MS, Surgical Outcomes and Quality Improvement Center (SOQIC), Department of Surgery and Center for Healthcare Studies, Feinberg School of Medicine, Northwestern University, 633 N St Clair St, 20th Floor, Chicago, IL 60611 (k-bilimoria@northwestern.edu).

On July 27, 2016, the Centers for Medicare & Medicaid Services (CMS) released the controversial Overall Hospital Quality Star Ratings,[1] despite concerns raised by hospitals and Congress that CMS had not provided enough details regarding the methodology, shared the underlying data, or responded adequately to comments from various stakeholders. Federal legislation has now been introduced to have CMS take down the Star Ratings.[2]

A composite measure of quality is a laudable objective, as it is challenging for patients to make decisions about where to seek care when presented with hundreds of individual metrics that have varying relevance and validity and that offer conflicting ratings (eg, a hospital is rated good for infections but poor for myocardial infarction). Public reporting initiatives are important, as hospital ratings can successfully drive hospital quality improvement efforts and payers increasingly seek to use hospital ratings to support contracting.

Many hospital quality rating systems are now available, but there is virtually no agreement among the rating systems in identifying "better" or "worse" hospitals.[3] The CMS Star Ratings system is drawing particular scrutiny because, as the single largest payer in the country and as a federal agency, CMS has considerable legitimacy and credibility. It is especially important that CMS' rating system demonstrate merit, given its potential influence. However, the CMS Star Ratings have several shortcomings, with the potential risk of doing more harm than good unless careful revision is undertaken.

CMS Star Ratings and Stakeholder Objections

The CMS Overall Hospital Quality Star Ratings are a composite of 64 diverse Hospital Compare measures grouped into 7 domains of quality and safety.[1] These domains are weighted based on input from patients, advocacy groups, and a technical expert panel. Mortality, safety of care, readmission, and patient experience each contribute 22% of the score. Effectiveness of care, timeliness of care, and efficient use of medical imaging each contribute 4%. Hospitals receive a rating from 1 to 5 stars that is displayed on the main page for each hospital on Hospital Compare. Of the 3647 hospitals receiving a star rating, 2.4% received 5 stars; 22.5%, 4 stars; 51.6%, 3 stars; 19.6%, 2 stars; and 4.0%, 1 star.[1]

Hospitals and hospital associations objected to the release of the star ratings for several reasons. First, there were concerns about transparency: the national data and detailed methodology were not made available to hospitals prior to the release to allow data and methodology validation. Second, the weighting of the composite components was criticized, as it appears to treat certain metrics equally that have dissimilar clinical significance (eg, mortality and readmissions).[4]

Third, the Star Ratings composite includes and heavily weights several measures that have been widely criticized because of poor validity. For example, a composite measure of complications, Patient Safety Indicator-90 (PSI-90), has been criticized for its biased component measures (eg, venous thromboembolism outcome measure is limited by surveillance bias),[5] inaccurate and clinically flawed component measures (eg, accidental puncture and laceration), statistical methodology, weighting, and risk adjustment.[6,7] A flawed, but improved, revision to PSI-90 has been endorsed by the National Quality Forum, but it may be some time before this will be incorporated into CMS public reporting and pay-for-performance programs.

Fourth, there is a lack of adequate comorbidity and case mix risk adjustment, as well as a lack of consideration of sociodemographic factors (eg, for readmission measures). Fifth, the rationale and clustering methodology to force the data into 5 star-rating groups is debatable. Do only 2% of hospitals merit 5 stars? Sixth, the accuracy of the underlying data and potential gaming of the component measures remain a major concern. Seventh, since the release of the star ratings, preliminary analyses have demonstrated that decisions about how to treat missing data appear to have resulted in systematic bias leading to a higher star rating for hospitals that reported fewer metrics—typically smaller, nonteaching and specialty hospitals.[1]

Why Hospital Ratings Can Mislead Stakeholders

Patients rarely use public reports of hospital quality when making decisions about where to seek care. They are far more likely to take recommendations from their physicians, family, and friends. This may be because the ratings have not met patients' needs and provided the type of high-quality, specific data that they need for decision making.

Most health care ratings are not sensitive to patients' health literacy and numeracy differences. Moreover, these ratings typically report measures of convenience that were collected for other reasons such as billing, resulting in measures that are often not important to patients. For example, although they may be interesting measures in some respects, it is doubtful that many patients would choose a hospital based on readmissions, influenza immunization, or imaging measures.

Adding to the confusion, hospital ratings systems offer conflicting information, as they rarely agree on which hospitals are top performers and which are poor performers.[3] In fact, the same hospital may be rated "best" on some rankings and "worst" on others. For example, in a study comparing 4 leading ratings, only 10% of hospitals rated as high performers by 1 of the 4 ratings

systems were high performers on even 1 of the others.[3] Additionally, only 1 hospital of 20 on the US News Honor Roll received 5 stars on the CMS Star Ratings.

Moreover, the conditions and procedures covered in most of these measures reflect only a fraction of inpatient care. There are no measures for most elective surgeries, and the medical conditions highlighted, such as acute myocardial infarction, pneumonia, and congestive heart failure, make up less than 15% of hospital admissions and are among the least likely measures for which patients may compare hospitals, given the urgency of these conditions.

Hospitals are appropriately skeptical of these conflicting rating systems, and quality improvement planning should not focus unduly on them. The ratings conflict because of different metrics, statistical methods, and weights. Some hospitals seem to refocus their quality goals every couple months whenever one of the many ratings systems releases updated rankings. Hospitals risk spending their scarce resources improving flawed publicly reported metrics rather than improving actual care processes that are important and show evidence of defects.

Advancing Toward a Meaningful Star Rating System

Several changes could be helpful in making hospital rating systems more meaningful. First, the composite should be based on a smaller number of relevant measures that patients actually find beneficial in decision making (eg, mortality, serious complications, patient-reported outcomes).[8] The measures selected must be the best available—valid, reliable, and well tested, designed to minimize ascertainment bias and gaming. Second, the measures should be collected by all hospitals being compared. However, this often leads to reliance on administrative data that are often not valid for assessing many risk-adjusted outcomes. Clinical registries offer much better data for assessing outcomes, but only a subset of hospitals participate in these registries, limiting the ability to compare all hospitals. Third, it may be possible that the wrong outcomes are being measured. Avoiding in-hospital mortality is undoubtedly important to patients, but most patients usually have little interest in using readmission, surgical site infection, or venous thromboembolism rates in decisions about where to seek care. Significant national investment is needed to obtain, measure, and report risk-adjusted patient-reported outcomes for a wide variety of conditions and procedures. Most patients generally are more interested in their ability to walk long distances without pain after knee replacement than whether they develop a catheter-associated urinary tract infection after surgery. Fourth, the hospital star ratings are a one-size-fits-all approach: All patients view the same ratings irrespective of their preferences, comorbidities, age, and ethnicity. Patients should be able to customize a hospital rating composite's components based on personal preferences.

Another common concern has been that the CMS Star Ratings do not have face validity: many major national referral centers did not receive 5 stars, and some received only 1 or 2 stars. This disconnect demonstrates the relevance of reputation as a reflection of clinicians' recognition of experts and advanced services (eg, procedures, trials, technology) that are not captured in the CMS rating and most other hospital rating systems but are known by referring physicians. There may be some merit in measuring facility and clinician reputation to approximate some of these factors, if performed in a methodologically rigorous way, until better metrics are available.

To reconcile the numerous conflicting and methodologically weak rating systems, an approach to "rate the raters" would be beneficial, while continuing vigorous work to improve the availability of meaningful measures. There are too many hospital report cards, each using different methods, measures, weights, and displays. A system of assessing public hospital ratings would help patients decide which ratings to use and trust. If a hospital rating system used metrics known to be flawed or untested, it would receive a lower rating compared with a report card that used well-vetted, established metrics and methods. Most importantly, rating the raters should not stifle innovation in measurement and transparency but rather needs to motivate development, use, and testing of valid metrics and measures that are meaningful to the intended consumers.

Conclusions

Patients need and want composite measures of hospital and clinician quality, and public reporting can be effective in stimulating quality improvement. However, current hospital rating systems do not provide comprehensive, relevant, accessible information, and current composites may mislead patients, payers, and hospitals. But opportunities to provide meaningful hospital quality report cards are within reach.

ARTICLE INFORMATION

Conflict of Interest Disclosures: The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Bilimoria reported receiving support from the National Institutes of Health, Agency for Healthcare Research and Quality (AHRQ R21HS021857 and R01HS024516), American Board of Surgery, American College of Surgeons, Accreditation Council for Graduate Medical Education, Health Care Services Corporation, Mallinkrodt, Northwestern University, the Robert H. Lurie Comprehensive Cancer Center, and Northwestern Memorial Hospital and receiving honoraria from hospitals, professional societies, and continuing medical education companies for clinical care and quality improvement research presentations. Dr Barnard reported no disclosures.

Additional Contributions: We acknowledge Jeffrey Softcheck (Northwestern Memorial Hospital), Jeanette Chung, PhD (Northwestern University), John Oliver Delancey, MD, MPH (Northwestern University), Scott Wetzel, MPP (Association of American Medical Colleges), and Atul Grover, MD, PhD (Association of American Medical Colleges), for their assistance. None of these persons received compensation for their contributions.

REFERENCES

1. Centers for Medicare & Medicaid Services (CMS). *Overall Hospital Quality Star Ratings on Hospital Compare Methodology Report (v2.0)*. Baltimore, MD: Centers for Medicare & Medicaid Services; May 2016.

2. Hospital Quality Rating Transparency Act of 2016, HR 5927, 114th Cong, 2nd Sess (2016).

3. Austin JM, Jha AK, Romano PS, et al. National hospital ratings systems share few common scores and may generate confusion instead of clarity. *Health Aff (Millwood)*. 2015;34(3):423-430.

4. Jha AK. The stars of hospital care: useful or a distraction? *JAMA*. 2016;315(21):2265-2266.

5. Bilimoria KY, Chung J, Ju MH, et al. Evaluation of surveillance bias and the validity of the venous thromboembolism quality measure. *JAMA*. 2013;310(14):1482-1489.

6. Rajaram R, Chung JW, Kinnier CV, et al. Hospital characteristics associated with penalties in the Centers for Medicare & Medicaid Services Hospital-Acquired Condition Reduction Program. *JAMA*. 2015;314(4):375-383.

7. Rajaram R, Barnard C, Bilimoria KY. Concerns about using the patient safety indicator-90 composite in pay-for-performance programs. *JAMA*. 2015;313(9):897-898.

8. Berwick DM. Era 3 for medicine and health care. *JAMA*. 2016;315(13):1329-1330.