When Stars Do Not Align: Overall Hospital Quality Star Ratings and the Volume-Outcome Association

Mark Alan Fontana, PhD, Stephen Lyman, PhD, Wasif Islam, BS, and Catherine H. MacLean, MD, PhD

Investigation performed at the Hospital for Special Surgery, New York, NY

Background: Volume-outcome relationships are well established for coronary artery bypass grafting and total joint arthroplasty surgery. Although the U.S. Centers for Medicare & Medicaid Services (CMS) Overall Hospital Quality Star Ratings program includes outcome quality measures for these procedures, these outcome quality measures are not counted toward the star ratings for low-volume hospitals. We sought to assess whether excluding low-volume hospitals from surgical quality measures with known volume-outcome relationships affects the star ratings.

Methods: We identified quality measures used in CMS’s star ratings that are related to surgical procedures with a known volume-outcome relationship and tested for the presence of the volume-outcome association for each of these measures. We then imputed missing values for low-volume hospitals for each measure and otherwise identically repeated the CMS calculations in order to assess the percentages of hospitals with the same, better, or worse ratings.

Results: Among the measures used to calculate star ratings, we identified 4 quality measures (2 related to coronary artery bypass grafting and 2 related to total joint arthroplasty) with known volume-outcome relationships that were excluded from the calculations of the star ratings for low-volume hospitals. We confirmed a volume-outcome association in the CMS data for all 4 measures. When total joint arthroplasty complications were imputed for low-volume hospitals and then included in the calculation of the star ratings, over one-third of hospitals received a different rating; both low-volume and other hospitals were more often hurt than helped. Imputing the other 3 quality measures among low-volume hospitals left the ratings unchanged.

Conclusions: The CMS star ratings do not fully represent the risks of undergoing procedures at low-volume hospitals, potentially misrepresent quality across facilities, and hence are of uncertain utility to consumers.

Health-care spending accounted for 17.9% of United States Gross Domestic Product (GDP) in 2016—a total expected to rise to 19.7% by 2026. The Affordable Care Act has catalyzed a shift from fee-for-service purchasing to value-based purchasing in the hope of facilitating quality per dollar spent. Tracking this effort involves accurately measuring and transparently reporting measures of quality to policymakers, clinicians, researchers, and the public.

In an effort to support consumer decision-making, the Centers for Medicare & Medicaid Services (CMS) created its Overall Hospital Quality Star Ratings ("stars"). First released in July 2016, the aim was to make quality information more actionable and accessible by consolidating dozens of hospital quality measures into a familiar 5-star rating system. The December 2017 star ratings are based on up to 57 quality measures, each categorized into 1 of 7 domains: mortality, safety of care, readmission, patient experience, effectiveness of care, timeliness of care, and efficient use of medical imaging. Some quality measures span the entire spectrum of care in a hospital, whereas others are specific to a particular procedure. All quality measures available from each hospital in each domain are entered into a statistical model called a latent variable model, which calculates weights (also known as loadings) for each quality measure. These loadings are then applied to calculate up to 7 domain scores per hospital, which are then averaged according to another, predetermined weighting scheme created by a panel of experts, resulting in a single summary score per hospital. Finally, hospital summary scores are clustered into 1 to 5 stars.
Hospital star ratings are based on whichever of the 57 quality measures are applicable to each hospital and for which CMS does not deem the number of cases too few to report. Extensive literature on both surgeon and hospital volume has demonstrated that higher volume is associated with better outcomes, although the magnitude of this association varies across conditions, procedures, and years. With regard to hospital star ratings, volume-outcome associations have been demonstrated for total knee arthroplasty as well as for total hip arthroplasty and coronary artery bypass grafting. We investigated the effect of excluding individual hospitals' surgical quality measures because of low volume and its impact on the star ratings.

We hypothesized that if some quality measures for certain hospitals are excluded from CMS's calculations because of low volume, and if low volume is associated with worse outcomes, then including those omitted quality measures would negatively impact associated hospitals’ ratings, leaving other hospitals relatively unaffected or positively impacted.

Materials and Methods

We focused attention on quality measures related to surgical procedures with a demonstrated volume-outcome relationship. Among the 57 quality measures included in the December 2017 star ratings, we identified 6 such measures: 2 for total joint arthroplasty (namely, complication and readmission rates), 2 for coronary artery bypass grafting surgery (namely, mortality and readmission rates), 1 for colon surgery, and 1 for abdominal hysterectomy. We next evaluated whether these measures were ever not included for low-volume hospitals in the calculation of the star ratings. CMS's public flat files contain footnotes indicating why a particular quality measure for a hospital is not included; for the colon and hysterectomy measures, the number of cases or patients was never too few to report. For the other 4 measures, a footnote for some hospitals indicated that the number of cases was too few to report, meaning that there were <25 cases (but not 0 cases) over the 3 years covered by each measure. For total joint arthroplasty complications and readmissions, 601 and 599 of 4,579 hospitals, respectively, had low volume, and thus these measures were excluded from the calculation of their star ratings (Table I). Similarly, for coronary artery bypass grafting mortality and readmissions, 146 and 142 of 4,579 hospitals, respectively, had low volume.

We then verified the volume-outcome relationship for these 4 measures in the Hospital Compare December 2017 data release. We performed univariate linear regressions of each hospital quality measure on associated volumes. As a robustness check, we also performed multivariable regressions, controlling for state, hospital type, hospital ownership, emergency services, and meaningful use of electronic health records.

TABLE I Hospital Surgical Quality Measure Summary Statistics*

| Quality Measure                                      | Abbreviation                        | N     | Mean (SD) | Median | 90th Percentile | No. of Low-Volume Hospitals |
|------------------------------------------------------|-------------------------------------|-------|-----------|--------|-----------------|-----------------------------|
| Hospital-level 30-day all-cause risk-standardized complication rate following elective primary total hip arthroplasty and total knee arthroplasty | Total joint arthroplasty complications | 2,755 | 0.028 (0.005) | 0.027 | 0.035 | 601 |
| Hospital-level 30-day all-cause risk-standardized readmission rate following elective total hip arthroplasty and total knee arthroplasty | Total joint arthroplasty readmissions | 2,765 | 0.044 (0.006) | 0.046 | 0.053 | 599 |
| Coronary artery bypass graft 30-day mortality rate   | Coronary artery bypass grafting mortality | 1,034 | 0.033 (0.009) | 0.031 | 0.045 | 146 |
| Coronary artery bypass graft 30-day readmission rate | Coronary artery bypass grafting readmissions | 1,029 | 0.138 (0.013) | 0.137 | 0.154 | 142 |

*N = number of hospitals with each quality measure included in CMS’s December 2017 Overall Hospital Quality Star Ratings, and No. of Low-Volume Hospitals = number of hospitals at which the relevant procedure was performed but the associated quality measure was not included by CMS in the star ratings.
For our main analysis, we imputed values of missing quality measures among low-volume facilities with use of 3 methods that varied in terms of how the missing values were filled using data from other hospitals’ non-missing quality measures: (1) using the predicted value from a best-fit multivariable regression of quality measures on state, hospital type, hospital ownership, emergency services, and meaningful use of electronic health records (henceforth referred to as the regression-based imputation method), (2) using the 90th percentile (the bottom 10% of performance) of quality measure values (henceforth referred to as the bottom-performance imputation method) (Table I), and (3) using the mean quality measure value (henceforth referred to as the mean-based imputation method) (Table I). The mean-based imputation method was generous to low-volume hospitals given the volume-outcome association (i.e., it represented a sort of best-case scenario for these facilities). The regression-based imputation method was also generous in practice, predicting distributions of quality for low-volume hospitals centered on the overall mean.

Independently for each quality measure and imputation method, we then otherwise identically repeated Hospital Compare’s methodology with use of CMS’s publicly released code (assuming a volume of 12 for imputed hospitals). Finally, we compared the results after imputation with the results from our own replication of Hospital Compare’s unaltered methodology, examining what happened to hospitals’ 1 to 5-star ratings, hospitals’ composite scores, hospitals’ domain scores, and the underlying latent variable models.

**Results**

We confirmed that the volume-outcome relationship held for each of the 4 quality measures assessed (see Appendix Figure E-1). A 100-case interval increase was associated with a

| Total Joint Arthroplasty Complications | ImputationStars | Comparison | Total |
|---------------------------------------|----------------|------------|-------|
| Replication Stars                      | 1   | 2   | 3   | 4   | 5   | 2283 (61.29%) |
| 1                                     | 178 | 70  | 12  | 0   | 0   | 261 Better 477 (12.92%) |
| 2                                     | 145 | 448 | 106 | 24  | 0   | 752 Worse 952 (25.79%) |
| 3                                     | 28  | 269 | 145 | 14  | 10  | 1,169 Total 3692 (100.00%) |
| 4                                     | 1   | 53  | 328 | 688 | 82  | 1,153 |
| 5                                     | 0   | 0   | 20  | 107 | 210 | 337 Newly rated 7 |
| Total                                 | 354 | 834 | 1,242 | 960 | 302 | 3,692 Never rated 860 |

| Total Joint Arthroplasty Readmissions  | ImputationStars | Comparison | Total |
|---------------------------------------|----------------|------------|-------|
| Replication Stars                      | 1   | 2   | 3   | 4   | 5   | 3686 (99.84%) |
| 1                                     | 256 | 0   | 0   | 0   | 0   | 261 Better 2 (0.05%) |
| 2                                     | 0   | 752 | 0   | 0   | 0   | 752 Worse 4 (0.11%) |
| 3                                     | 0   | 0   | 1,189 | 0   | 0   | 1,189 Total 3692 (100.00%) |
| 4                                     | 0   | 0   | 0   | 1,148 | 0   | 1,153 |
| 5                                     | 0   | 0   | 0   | 0   | 337 | 337 Newly rated 18 |
| Total                                 | 256 | 754 | 1,193 | 1,149 | 337 | 3,692 Never rated 869 |

| Coronary Artery Bypass Grafting Mortality | ImputationStars | Comparison | Total |
|------------------------------------------|----------------|------------|-------|
| Replication Stars                        | 1   | 2   | 3   | 4   | 5   | 3689 (99.92%) |
| 1                                       | 261 | 0   | 0   | 0   | 0   | 261 Better 2 (0.05%) |
| 2                                       | 0   | 752 | 0   | 0   | 0   | 752 Worse 2 (0.05%) |
| 3                                       | 0   | 0   | 1,189 | 0   | 0   | 1,189 Total 3692 (100.00%) |
| 4                                       | 0   | 0   | 0   | 1,153 | 0   | 1,153 |
| 5                                       | 0   | 0   | 0   | 0   | 337 | 337 Newly rated 0 |
| Total                                   | 261 | 752 | 1,189 | 1,154 | 337 | 3,692 Never rated 867 |

| Coronary Artery Bypass Grafting Readmissions | ImputationStars | Comparison | Total |
|---------------------------------------------|----------------|------------|-------|
| Replication Stars                           | 1   | 2   | 3   | 4   | 5   | 3691 (99.97%) |
| 1                                          | 261 | 0   | 0   | 0   | 0   | 261 Better 1 (0.03%) |
| 2                                          | 0   | 752 | 0   | 0   | 0   | 752 Worse 0 (0.00%) |
| 3                                          | 0   | 0   | 1,189 | 0   | 0   | 1,189 Total 3692 (100.00%) |
| 4                                          | 0   | 0   | 0   | 1,153 | 0   | 1,153 |
| 5                                          | 0   | 0   | 0   | 0   | 337 | 337 Newly rated 0 |
| Total                                      | 261 | 752 | 1,189 | 1,154 | 337 | 3,692 Never rated 867 |

Fig. 1

Star comparison matrices: replication versus imputation. Entries highlighted in blue indicate the number of hospitals that have identical star ratings between replication (without imputation) and imputation with use of the regression-based imputation method. Other entries not highlighted indicate the number of hospitals that have different star ratings between replication and imputation. Comparison and total columns indicate the number (and percentage) of hospitals with the same, better, or worse star ratings after imputation. “Newly rated” indicates the number of hospitals that did not have a star rating before imputation but did after imputation. “Never rated” indicates the number of hospitals that did not have a star rating before or after imputation.
0.032 percentage point decrease in total joint arthroplasty complications (p < 0.001). For total joint arthroplasty readmissions, coronary artery bypass grafting mortality, and coronary artery bypass grafting readmissions, a 100-case interval increase was associated with percentage point decreases of 0.027 (p < 0.001), 0.150 (p < 0.001), and 0.157 (p < 0.001), respectively. For example, for a hospital that performs 25 cases over 3 years, the total joint arthroplasty complication rate is expected to be 2.90%; for a hospital that performs 3,025 cases over 3 years, the rate is expected to be 1.94% (i.e., 0.029 + 3,000/100 × [−0.00032] = 0.0194). After controlling for state, hospital type, hospital ownership, emergency services, and meaningful use of electronic health records, a 100-case interval increase was associated with 0.034 (p < 0.001), 0.030 (p < 0.001), 0.158 (p < 0.001), and 0.159 (p < 0.001) percentage point decreases, respectively.

For total joint arthroplasty readmissions, coronary artery bypass grafting mortality, and coronary artery bypass grafting readmissions, >99% of hospitals retained the same star rating after regression-based imputation (Fig. 1). Hospitals’ associated continuous composite scores were also nearly identical, as were their underlying domain scores and latent variable models. The other 2 imputation methods yielded nearly identical results.

However, for total joint arthroplasty complications, across all 3 imputation methods, over one-third of hospitals received a different star rating. After regression-based imputation, 12.9% of hospitals received a better rating and 25.8% received a worse rating (Fig. 1). For low-volume hospitals for which the total joint arthroplasty complication measure was imputed, 3.1% of hospitals received a better rating and 23.9% received a worse rating, whereas for hospitals for which the total joint arthroplasty complications measure was not imputed (i.e., those for which CMS-reported rates were used), 14.4% of hospitals received a better rating and 26.1% received a worse rating (see Appendix Table E-1). Results were similar for the other 2 imputation methods. The only exception was that, for the bottom-performance imputation method, among imputed hospitals, only 0.2% received a better rating and 74.4% received a worse rating (see Appendix Table E-1). Overall, although low-volume hospitals were more often hurt than helped after imputation (especially for the bottom-performance method), higher-volume hospitals were also more often hurt than helped.

Hospitals’ continuous composite scores (after taking weighted averages of the 7 domain scores) also differed after regression-based imputation of total joint arthroplasty complications (Fig. 2). The safety domain scores drove these changes: imputed hospitals’ safety scores collapsed toward the all-hospital average after imputation, and the safety scores of hospitals without total joint arthroplasty complications imputed bore little resemblance to their original values (Fig. 3). Similar trends held across imputation methods. These changes were driven by changes in latent variable model “loadings” (weights assigned to the quality measures in the safety domain). Without imputation of total joint arthroplasty complications, the Complication/Patient Safety for Selected Indicators (PSI-90) measure received the highest weight (0.944); this value was reduced to 0.189 after regression-based imputation (see Appendix Figure E-2). Conversely, the weight for total joint arthroplasty complications moved from 0.211 to 0.955. The
The December 2017 CMS stars measure set includes 4 surgical quality measures with known volume-outcome relationships and also missing values for some hospitals because of low volume. Imputing 3 of these measures (total joint arthroplasty readmissions, coronary artery bypass grafting mortality, and coronary artery bypass grafting readmissions) for low-volume hospitals resulted in virtually unchanged star ratings. Given the known and demonstrated volume-outcome relationships for these 3 measures, we expected that imputing quality measures from low-volume facilities would harm their relative ratings and help the relative ratings of other hospitals; however, this did not occur. This finding speaks to the stability of the underlying mortality and readmission latent variable models, although the scores that they produce omit important quality information; the hospitals excluded from these 3 measures performed fewer than 25 procedures over 3 years (about 1 procedure every 6 weeks). Yet, consumers viewing their ratings would have no indication that they should expect higher complication rates. Although these results are consistent with CMS’s assertion that changing 1 measure should make little difference for star ratings, they point out limitations of the method in fully describing hospital quality.

In contrast, imputing missing values for the fourth measure, total joint arthroplasty complications, resulted in significantly different star ratings. The fact that low-volume hospitals were more often hurt than helped after imputation is consistent with the volume-outcome association. However, the fact that the much larger number of hospitals without the measure were also more often hurt than helped, even when low-volume hospitals are assigned quality measures in the worst decile, is inconsistent with the volume-outcome relationship. We can trace these star rating changes to changes in the safety domain, driven by a substantially lower loading on PSI-90 in the latent variable model, and a substantially higher loading on total joint arthroplasty complications. That is, the model underlying the safety domain is currently heavily weighted toward 1 measure, PSI-90, calling into question CMS’s assertion that poor performance on 1 safety measure does not preclude high performance in safety overall. The model generating this weighting is unstable, and changes to it have substantial impact on the stars. Our results call into question the safety latent variable model as well as the list of quality measures composing that domain, particularly PSI-90; these results are consistent with past reports of problems with PSI-90. Although some of these issues have been addressed, the PSI-90 component measures still may not accurately measure certain adverse events and can be biased by differential surveillance across hospitals.

We appreciate that CMS faces a tradeoff in how it handles quality measures with few observations, particularly for quality measures such as total joint arthroplasty complications, which denote relatively rare outcomes for which risk adjustment is likely incomplete. With regard to risk adjustment, Ayers et al. noted that CMS’s risk adjustment uses billing data, not clinician or patient-reported health information. Using the FORCE-TJR (Function and Outcomes...
Lack of adjustment for socioeconomic status could be another possible source of incomplete risk adjustment. CMS’s reliance on claims-based measures missing these details, especially in light of our analyses, is cause for hesitation. On the one hand, with regard to the rarity of outcomes, the volume-outcome relationship predicts that omitting measures from low-volume hospitals can impact star ratings (as we have demonstrated for total joint arthroplasty complications). On the other hand, lower volume indicates a greater degree of uncertainty around the usefulness of a quality measure for predicting performance. The calculated quality measure for some low-volume hospitals therefore will put them in a more positive light than might be warranted and may put other hospitals in a more negative light. Perhaps a reasonable compromise would be to use the omitted quality data to calculate the star ratings and to include those low-volume measures in the released data files but not elsewhere on the CMS website. Another approach could be to add a structural measure that requires a volume threshold.

Although it might seem counterintuitive that imputing one hospital’s quality measure can impact the domain scores and ratings of other, non-imputed hospitals, this interconnectedness is by design. The loadings assigned to the quality measures are estimated jointly from all hospitals’ available data, producing domain scores based on relative performance. As health-care incentives move toward value-based programs in which higher quality is rewarded financially, defining quality in relative terms may very well prevent some providers with high absolute quality scores, but lower relative quality scores, from achieving these rewards. Relative ratings will tend to overemphasize the differences between hospitals, even if true differences are minimal. This overemphasizing differences, in turn, could engender counterproductive competition by discouraging hospitals from sharing best practices and collaborating.

The star ratings are intended to reflect the type of care provided most often at any given hospital. A related limitation of the star ratings is that they are not comparable across hospitals because the specific quality measures used to calculate them differ from hospital to hospital. Existing work has focused on measuring systematic differences in the star ratings and the number of measures used to calculate them across hospital types. Hospitals that report fewer measures (smaller, non-teaching, specialty hospitals) tend to have more stars. Not including low-volume surgical quality measures could exacerbate this bias beyond the implications of the volume-outcome relationship by mechanically reducing the number of quality measures included in the star ratings, rendering the measure set even less uniform.

In April 2018, CMS announced it would only update the star ratings in its July and December releases.14-26 The July 2018 release, however, was cancelled after CMS published its preview reports; there was concern regarding updates to the measures in the safety domain and its latent variable model: the loading on PSI-90 fell (from 0.94 to 0.17), whereas the loading on total joint arthroplasty complications commensurately rose (from 0.21 to 0.96).25,26 This “flip” is nearly identical to what we found, although our findings resulted from imputing low-volume hospitals, whereas CMS’s findings resulted from updating the underlying measures and their collection period. Importantly, the updated PSI-90 included substantial methodological changes (i.e., the components that were included in PSI-90, their weighting, and an underlying switch from International Classification of Diseases (ICD)-9 to ICD-10). It seems clear that the safety domain, whether from imputing low-volume hospitals or changing an underlying quality measure, is unstable; therefore, applying a latent variable model to it is problematic.

The present study had several limitations. First, from the public dataset, we did not know the actual volume of the relevant procedure among the low-volume hospitals, other than that it was <25 (and not zero); we assumed a value of 12. Second, we did not know the actual values of omitted low-volume quality measures. Given that our results did not vary among our best-case and worst-case imputation scenarios and that the true values of these missing variables likely lie somewhere between these scenarios, we believe that our findings would hold even if the actual values were known. Regarding both points, in order to facilitate replication and transparency, we felt that it was important to avoid using data that were not publicly posted for an analysis of a public ratings system. Third, 2 of our imputation methods were fairly simple: assigning all low-volume hospitals either the mean or 90th percentile (worst decile) among non-missing hospitals. It is unlikely that, if we had been able to observe these missing quality measures, they would have been the same. To address this concern, we fitted a multivariable regression based on a number of covariates to assign low-volume hospitals quality measures. Fifth, although our focus was on volume, many other factors are important for achieving good outcomes (good teams, standardization of protocols, management of risk, etc.); most volume research investigates the impact of volume, holding constant these sorts of factors to the extent they are measured (or at least correlated with measured factors).

In conclusion, the CMS Overall Hospital Quality Star Ratings do not represent the risks of undergoing procedures at low-volume hospitals, potentially misrepresent quality across facilities, and hence are of uncertain utility to consumers. The fact that excluding scores of low-volume hospitals for 3 of 4 quality measures with a demonstrated volume-outcome relationship did not impact the overall star ratings raises questions as to the meaningfulness of the rating system, especially for patients seeking quality information for these procedures. Moreover, the fact that the total joint arthroplasty complication
measure impacts ratings for both low-volume and other hospitals raises questions regarding the safety domain.

Appendix

Supporting material provided by the authors is posted with the online version of this article as a data supplement at jbjs.org (http://links.lww.com/JBJSOA/A90).

Mark Alan Fontana, PhD1,2
Stephen Lyman, PhD1,2

Wasif Islam, BS1,2
Catherine H. MacLean, MD, PhD1

1Center for the Advancement of Value in Musculoskeletal Care, Hospital for Special Surgery, New York, NY
2Department of Healthcare Policy & Research (M.A.F. and S.L.), Weill Cornell Medical College (W.I.), New York, NY

E-mail address for M.A. Fontana: fontanaam@hss.edu

ORCID iD for M.A. Fontana: 0000-0002-0529-9035
ORCID iD for S. Lyman: 0000-0003-4549-3817
ORCID iD for W. Islam: 0000-0002-3241-1571
ORCID iD for C.H. MacLean: 0000-0003-2581-189X

References

1. Hartman M, Martin AB, Espinosa N, Catlin A; The National Health Expenditure Accounts Team. National health care spending in 2016: spending and enrollment growth slow after initial coverage expansions. Health Aff (Millwood). 2018 Jan;37(1):150-60. Epub 2017 Dec 6.
2. Centers for Medicare & Medicaid Services. National health expenditures fact sheet. 2018 Dec 6. https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nhe-fact-sheet.html. Accessed 2018 Dec 14.
3. Centers for Medicare & Medicaid Services. CMS updates website to compare hospital quality. 2017 Dec. https://www.cms.gov/newsroom/press-releases/cms-updates-website-compare-hospital-quality. Accessed 2019 Feb 14.
4. Venkatesh AK, Bernheim SM, Qin L, Bao H, Simoes J, Norton E, Wing M, Glennon G, Shah R, Herrin J, Lin H, Lin Z, Krumholz HM. Overall hospital quality star rating on Hospital Compare December 2017 updates and specifications report. 2017 Dec. https://s3.amazonaws.com/assets.qualitynet.org/dcs/ContentServer?c=FORCE-TJR to improve the accuracy of risk-adjustment prediction models for thirty-day readmission after total hip replacement and total knee replacement. J Bone Joint Surg Am. 2015 Apr 15;97(8):668-71.
12. Fehring TK. AHKS risk adjustment initiative: why is it important? J Arthroplasty. 2016 Jun;31(6):1148-50. Epub 2016 Mar 22.
13. DeLancey JO, Softcheck J, Chung JW, Barnard C, Dahlke AR, Bilimoria KY. Associations between hospital characteristics, measure reporting, and the centers for Medicare and Medicaid services overall hospital quality star ratings. JAMA. 2017 May 16;317(19):2016-7.
14. Centers for Medicare & Medicaid Services. April 2018 Hospital Compare release frequently asked questions. 2018 April. https://www.qualitynet.org/dcs/ContentServer?c=Page&pageName=QnetPublic%2FPage%2FQnetTier3&cid=1228775959020. Accessed 2019 Feb 15.
15. Bilimoria KY, Barnard C. The new CMS hospital quality star ratings the stars are not aligned. JAMA. 2016 Nov 1;316(17):1761-2.
16. Agency for Healthcare Research and Quality. AHRQ quality indicators PSI 90 fact sheet. 2016 Aug 31. https://www.qualityindicators.ahrq.gov/News/PSI90_FactSheet_FAQ_v1.pdf. Accessed 2018 Dec 14.
17. Utter GH, Zrelik PA, Baron R, Tancredi DJ, Sadeghi B, Geppert JI, Romano PS. Positive predictive value of the AHRQ accidental puncture or laceration patient safety indicator. Ann Surg. 2009 Dec;250(6):1041-5.
18. Bilimoria KY, Chung J, Ju MH, Haut ER, Bentrem DJ, Ko CY, Baker DW. Evaluation of surveillance bias and the validity of the venous thromboembolism quality measure. JAMA. 2013 Oct 9;310(14):1482-9.
19. Meddings JA, Reichert H, Hofer T, McMahon LF Jr. Hospital report cards for hospital-acquired pressure ulcers: how good are the grades? Ann Intern Med. 2013 Oct 15;159(8):505-13.
20. Ayers DC, Fehring TK, Odum SM, Franklin PD. Using joint registry data from FORCE-TJR to improve the accuracy of risk-adjustment prediction models for thirty-day readmission after total hip replacement and total knee replacement. J Bone Joint Surg Am. 2015 Apr 15;97(8):668-71.
21. Fehring TK. AHKS risk adjustment initiative: why is it important? J Arthroplasty. 2016 Jun;31(6):1148-50. Epub 2016 Mar 22.
22. MacLean C, Shapiro L. Does the Hospital Compare 5-star rating promote public health? Ann Surg. 2016 Sep 8. https://www.healthaffairs.org/do/10.1377/hblog20160908.096393/full/. Accessed 2019 Jan 11.
23. DeLancey JO, Softcheck J, Chung JW, Barnard C, Dahlke AR, Bilimoria KY. Associations between hospital characteristics, measure reporting, and the centers for Medicare & Medicaid services overall hospital quality star ratings. JAMA. 2017 May 16;317(19):2016-7.
24. Centers for Medicare and Medicaid Services. Overall hospital quality star rating FAQs for hospitals April 2018 preview reports. 2018 Apr.
25. QualityNet. CMS will not update Hospital Compare overall hospital quality star ratings. 2018 Jun 12. https://www.qualitynet.org/dcs/ContentServer?c=Page&pageName=QnetPublic%2FPage%2FQnetBasic&cid=1228776614455. Accessed 2018 Jul.
26. Centers for Medicare & Medicaid Services. Overall hospital quality star rating on Hospital Compare July 2018 updates and specifications report. 2018 Jul. https://www.qualitynet.org/dcs/ContentServer?c=Page&pageName=QnetPublic%2FPage%2FQnetTier3&cid=1228775959066. Accessed 2019 Jan 11.
27. Centers for Medicare & Medicaid Services. CMS recalibrated patient safety indicators fact sheet.