Original Article

Understanding the Relationship Between the Centers for Medicare and Medicaid Services' Hospital Compare Star Rating, Surgical Case Volume, and Short-Term Outcomes After Major Cancer Surgery

Deborah R. Kaye, MD [1]; Edward C. Norton, PhD[2,3,4]; Chad Ellimoottil, MD, MSc[1]; Zaojun Ye, MSc[1]; James M. Dupree, MD, MPH[1]; Lindsey A. Herrel, MD, MSc[1]; and David C. Miller, MD, MPH[1]

BACKGROUND: Both the Centers for Medicare and Medicaid Services' (CMS) Hospital Compare star rating and surgical case volume have been publicized as metrics that can help patients to identify high-quality hospitals for complex care such as cancer surgery. The current study evaluates the relationship between the CMS' star rating, surgical volume, and short-term outcomes after major cancer surgery. **METHODS:** National Medicare data were used to evaluate the relationship between hospital star ratings and cancer surgery volume quintiles. Then, multilevel logistic regression models were fit to examine the association between cancer surgery outcomes and both star rankings and surgical volumes. Lastly, a graphical approach was used to compare how well star ratings and surgical volume predicted cancer surgery outcomes. **RESULTS:** This study identified 365,752 patients undergoing major cancer surgery for 1 of 9 cancer types at 2,550 hospitals. Star rating was not associated with surgical volume ($P < .001$). However, both the star rating and surgical volume were correlated with 4 short-term cancer surgery outcomes (mortality, complication rate, readmissions, and prolonged length of stay). The adjusted predicted probabilities for 5- and 1-star hospitals were 2.3% and 4.5% for mortality, 39% and 48% for complications, 10% and 15% for readmissions, and 8% and 16% for a prolonged length of stay, respectively. The adjusted predicted probabilities for hospitals with the highest and lowest quintile cancer surgery volumes were 2.7% and 5.8% for mortality, 41% and 55% for complications, 12.2% and 11.6% for readmissions, and 9.4% and 13% for a prolonged length of stay, respectively. Furthermore, surgical volume and the star rating were similarly associated with mortality and complications, whereas the star rating was more highly associated with readmissions and prolonged length of stay. **CONCLUSIONS:** In the absence of other information, these findings suggest that the star rating may be useful to patients when they are selecting a hospital for major cancer surgery. However, more research is needed before these ratings can supplant surgical volume as a measure of surgical quality. ***Cancer* 2017;123:4259-67.**

KEYWORDS: cancer surgery, outcomes, patient decision making, publicly available hospital ratings, quality.

INTRODUCTION

Selecting a hospital for cancer surgery is challenging. Likewise, the best way for patients to determine where to obtain hospital-based health care remains a topic of debate in both the lay press and the scientific literature.[1-4] In an attempt to help patients make such decisions, several private and public organizations have released rating guides that rank hospitals according to various measures of quality and safety.[5-7]

The newest measure in this area is the Centers for Medicare and Medicaid Services' (CMS) Hospital Compare star rating system. This program uses a complex methodology, based on a hospital's performance with mortality, safety, readmissions, patient experience, care effectiveness, care timeliness, and efficient use of medical imaging, to assign each hospital a star rating ranging from 1 (lowest score) to 5 (highest score). The stated goal for this program is "to help millions of patients and their families learn about the quality of hospitals, compare facilities in their area side-by-side, and ask important questions about care quality when visiting a hospital or other health care provider."[8] Although the star ratings have the benefit of being publically available, the system has been criticized by some as being inaccurate, and there is little empirical information that validates the relationship between star ratings and important patient outcomes, including those of major cancer surgery.[9]

Corresponding author: Deborah R. Kaye, MD, Dow Division for Urologic Health Services Research, University of Michigan, 2800 Plymouth Road, NCRC Building 16, Room 108E, Ann Arbor, MI 48109-2800; kayed@med.umich.edu

[1]Dow Division of Health Services Research, Department of Urology, University of Michigan, Ann Arbor, Michigan; [2]Department of Health Management and Policy, University of Michigan, Ann Arbor, Michigan; [3]Department of Economics, University of Michigan, Ann Arbor, Michigan; [4]National Bureau of Economic Research, Cambridge, Massachusetts.

Additional supporting information may be found in the online version of this article

DOI: 10.1002/cncr.30866, **Received:** April 20, 2017; **Revised:** May 23, 2017; **Accepted:** June 6, 2017, **Published online** June 30, 2017 in Wiley Online Library (wileyonlinelibrary.com)

In contrast, arguably the best structural measure of quality for major cancer surgery is surgical case volume. Illustrating this point, several prominent health systems have issued a volume pledge for select surgical procedures, including some major cancer operations, as a way of demonstrating their commitment to transparency and high-quality care.[10] For the most part, however, the impact of this metric remains limited by the fact that hospital surgical volumes are not routinely available to the public.

In this context, an important question is whether a hospital's annual surgical volume correlates strongly with its Hospital Compare star rating for major cancer surgery. Furthermore, it is unknown whether one of these measures more strongly predicts important patient outcomes. The availability of such data would not only clarify the relevance of the star rating system for patients in need of major cancer surgery but also provide a better sense of its potential value as a quality metric for a broader range of conditions. Accordingly, we used data from the Hospital Compare program and national Medicare claims to evaluate the relationship between the CMS star rating and surgical volume for hospitals performing major cancer surgery. We also examined the frequency of short-term adverse outcomes with respect to star ratings and surgical volumes, and we assessed the relative predictive value of these measures for short-term adverse outcomes after major cancer surgery.

MATERIALS AND METHODS

Data Sources

We used 3 data sets to perform this analysis. We used the 100% Medicare Provider Analysis and Review File from 2011 to 2013 to identify the patient cohort, clinical data, and outcomes of interest. We also used publically available 2016 Hospital Compare data to identify hospital star rankings and the American Hospital Association annual survey to evaluate hospital characteristics.

Study Population

Our study population included Medicare beneficiaries aged 66 to 99 years who underwent a major extirpative surgery for colorectal, prostate, bladder, esophageal, kidney, liver, lung, ovarian, or pancreatic cancer from January 1, 2011 through November 30, 2013. Because we wanted to answer the question of whether a typical patient could use the Hospital Compare star rating system to choose a hospital for major cancer surgery, we excluded patients on the basis of 2 criteria. First, because they represent more complex operations with inherently different outcomes, we excluded patients who had 2 or more different oncologic procedures on the same day. Second, because patients with synchronous malignancies or staged procedures are also often more complex cases and may be served only by select centers with the appropriate resources, we excluded those who had more than 1 procedure $\leq$ 180 days apart. Finally, to improve statistical reliability, we also excluded hospitals with fewer than 10 oncologic procedures during the period of interest. These exclusions accounted for only 1.7% of the entire patient cohort ($n = 6504$). Finally, we excluded patients who received surgery at a hospital that lacked CMS star ratings.

Exposure Variables

For each patient, we first identified the hospital at which the patient had the extirpative cancer surgery. We then determined the CMS star rating (ie, 1, 2, 3, 4, or 5 stars) and average annual cancer surgery volume for the hospital where the surgery occurred. The star rating is a composite score comprising 64 possible measures that are currently part of the CMS Hospital Compare program. The individual measures are assigned to 7 different categories: mortality, safety, readmissions, patient experience, care effectiveness, care timeliness, and efficient use of medical imaging. To receive a star rating, a hospital has to report a minimum of 3 measures in at least 3 categories, including 1 of the outcome categories (mortality, safety, and readmissions). The CMS calculates a hospital's star rating with only the measures that the hospital chooses to report. The rating is publically available as part of the CMS Hospital Compare program.[11]

For each hospital, we assigned cancer surgery volume on the basis of the cumulative number of cases for all 9 cancers included in our analysis (bladder, colon, esophageal, kidney, liver, lung, ovarian, pancreatic, and prostate cancer) averaged over the 3 years of data in our cohort. We classified hospitals into quintiles on the basis of this volume measure (with 1 representing the lowest volume and 5 representing the highest volume), and we assigned patients to the volume quintile of their treating hospital. We used the overall cancer surgery volume for these analyses rather than individual procedure volumes because we believe that this composite measure reflects the most policy-relevant volume metric and one that could be considered reasonably analogous to the composite star ratings available from Hospital Compare.

Outcome Measures

We measured 4 outcomes occurring within 30 days of the index cancer surgery: mortality, complications, prolonged length of stay, and hospital readmissions. Complications were defined with established methods[12-14] and included infections, bleeding, gastrointestinal complications, neurologic complications, pulmonary complications, renal complications, cardiac complications, and other complications. A prolonged length of stay was defined as a hospital stay exceeding the 90th percentile for an individual procedure.[15]

Statistical Analysis

We first performed univariate statistical analyses to compare patient and hospital characteristics across star rating categories and surgical volume strata. We also used univariate statistical tests to evaluate the relationship between hospital star ratings and cancer surgery volume quintiles.

Next, we examined the association between cancer surgery outcomes and both star ratings and surgical volumes. To do this, we fit multilevel logistic regression models and controlled for both patient characteristics (age, sex, race, and Elixhauser comorbidities) and hospital characteristics (hospital bed number, urban location vs rural location, region, and teaching status).

We then used a graphical approach to compare how well star ratings and surgical volumes predict cancer surgery outcomes. We show how the outcomes compare not only across star ratings and volume quintiles but also across hospitals ranked into quintiles by the 4 cancer surgery outcomes. If the star ratings or surgical volumes predict cancer surgery outcomes, they will graphically show patterns similar to those for cancer surgery outcomes.

To do this, we first divided hospitals into quintiles based on their actual performance with each outcome measure. For example, a new hospital mortality rate quintile variable was created from the actual mortality rates for individual hospitals. By definition, this new measure represents the greatest possible relationship that could be attained between the outcome and any measurement rating system restricted to 5 quintiles and a linear line (eg, star rating or surgical volume); necessarily, therefore, this relationship represents a gold standard for comparing other hospital performance measures. Graphically, the slope of the line that is closest to the slope for the outcome measurement represents the better measure for predicting the outcome. A horizontal line would represent no association between the measurement strata and outcome ($R^2 = 0$).

We repeated this analysis for the star rating and surgical volume measurement categories for each cancer surgery outcome. That is, we calculated the mean hospital outcome by the rating quintile, fit the best line through the means, and calculated the slope of the line. The measurement (surgical volume or star rating) with the steeper relative slope represents the measurement that, on average, better predicts the outcome.

Finally, we performed several sensitivity analyses to examine the robustness of our findings. To analyze case-mix differences, we calculated the share of high-risk patients (>2 comorbidities) and low-risk patients (≤2 comorbidities) in hospitals across star ratings; we computed the percent share of each cancer type that contributed to a hospital's total cancer surgery volume and compared this percent share across star ratings; and we performed our primary analyses at the hospital level, which weighted each hospital's cancer specific outcome measures by the representative share of that cancer in our national sample. Second, to evaluate whether small numbers of procedures altered our findings, we repeated our primary analyses with hospitals limited to those with the highest cancer surgery volume (the top 50% and 75%). Third, we examined outcome stability over time. Fourth, to evaluate the impact of reporting patterns on star ratings, we compared the number of measurement categories that each hospital reported to the Hospital Compare program across star ratings, and we stratified outcomes by the number of reported measures among 5-star hospitals. Fifth, to evaluate whether our findings were clinically relevant in addition to policy-relevant, we measured the correlation between the overall cancer surgery volume quintile and the cancer-specific surgical volume quintile.

All analyses were performed with Stata 14 (StataCorp LP, College Station, Texas) at the 5% significance level. The University of Michigan institutional review board deemed this study exempt from review.

RESULTS

We identified 384,519 patients who underwent major cancer surgery for 1 of 9 cancers at 2667 hospitals in the United States. In this group, 365,752 patients were treated at 2550 hospitals that were assigned a star rating in Hospital Compare. Overall, 5% of the hospitals were assigned a 1-star rating, 24% were assigned 2 stars, 44% were assigned 3 stars, 25% were assigned 4 stars, and 3% were assigned 5 stars. The median annual cancer surgery surgical volume for each volume quintile was 5.0, 11.3, 22.1, 43.7, and 98.3, respectively.

Table 1 presents differences in patient characteristics according to star rankings and volume quintiles. As illustrated in Figure 1, the star rating had little to no association with surgical volume. Illustrating this point, 24% of the 5-star hospitals were high-volume hospitals, whereas

TABLE 1. Patient and Hospital Characteristics by the Centers for Medicare and Medicaid Services' Hospital Compare Star Rating and Surgical Volume Quintile

	Star Rating					
Characteristic	1 (Lowest)	2	3	4	5 (Highest)	*P*
Patients						
Age, mean, y	74.4	74.5	74.5	74.5	74.5	0.033
Male sex, %	56.5	57.9	57.5	58.4	62.2	<.001
Race, %						<.001
White	75.0	84.5	87.2	88.9	90.0	
African American	17.5	10.6	7.4	5.8	5.1	
Other	7.5	4.9	5.5	5.3	4.8	
No. of comorbidities, %						<.001
0	9.9	9.6	10.0	10.5	12.4	
1	20.1	20.9	21.0	21.5	22.0	
2	23.0	22.9	22.6	22.4	22.8	
3	18.6	18.8	18.4	17.8	17.4	
≥4	28.4	27.9	28.0	27.8	25.4	
Hospitals						
Geographic region, %						<.001
Northeast	39.7	23.3	15.4	13.1	6.4	
Midwest	11.6	15.5	24.8	40.0	42.9	
South	31.4	43.0	38.0	28.2	33.3	
West	17.4	18.1	21.8	18.7	17.5	
No. of beds, %						.302
<200	81.0	75.6	80.3	76.4	71.4	
200-399	13.2	14.8	11.7	14.2	15.9	
400-599	3.3	4.9	4.8	6.1	4.8	
≥600	2.5	4.7	3.3	3.4	7.9	
Hospital profit status, %						<.001
For profit	18.2	21.1	18.5	10.5	14.3	
Nonprofit	50.4	64.3	69.4	78.2	77.8	
Public	31.4	14.6	12.2	11.3	7.9	
Teaching hospital, %	59.5	41.8	30.7	30.0	33.3	<.001
Urban location, %	87.6	79.1	67.1	73.2	81.0	<.001

	Surgical Volume Quintile					
Characteristic	1 (Lowest)	2	3	4	5 (Highest)	*P*
Patients						
Age, mean, y	77.1	76.3	75.5	74.8	74.0	<.001
Male sex, %	48.3	51.1	53.9	56.5	60.1	<.001
Race, %						<.001
White	86.5	85.7	84.9	86.5	86.9	
African American	7.9	9.1	9.2	8.0	7.9	
Other	5.6	5.2	5.9	5.5	5.2	
No. of comorbidities, %						<.001
0	6.7	7.3	7.6	9.0	11.3	
1	16.2	16.6	17.7	19.8	22.6	
2	21.5	21.1	21.6	22.2	23.2	
3	19.6	19.3	19.3	18.5	17.9	
≥4	35.9	35.6	33.8	30.5	25.0	
Hospitals						
Geographic region, %						.025
Northeast	12.4	19.1	19.0	19.7	17.5	
Midwest	32.2	26.0	25.2	23.8	24.2	
South	36.1	36.9	36.4	43.2	38.0	
West	19.4	18.0	19.4	22.4	20.3	
No. of beds, %						<.001
<200	100.0	100.0	100.0	83.9	6.1	
200-399	0.0	0.0	0.0	16.1	50.0	
400-599	0.0	0.0	0.0	0.0	25.4	
≥600	0.0	0.0	0.0	0.0	18.5	
Hospital profit status, %						<.001
For profit	22.0	18.6	22.4	13.2	8.9	
Nonprofit	58.0	66.4	64.8	76.0	83.3	

TABLE 1. *Continued*

Characteristic	Surgical Volume Quintile					
	1 (Lowest)	2	3	4	5 (Highest)	*P*
Public	20.0	15.1	12.8	10.8	7.9	
Teaching hospital, %	15.1	18.0	28.8	39.9	71.1	<.001
Urban location, %	37.3	58.7	78.2	91.9	97.6	<.001

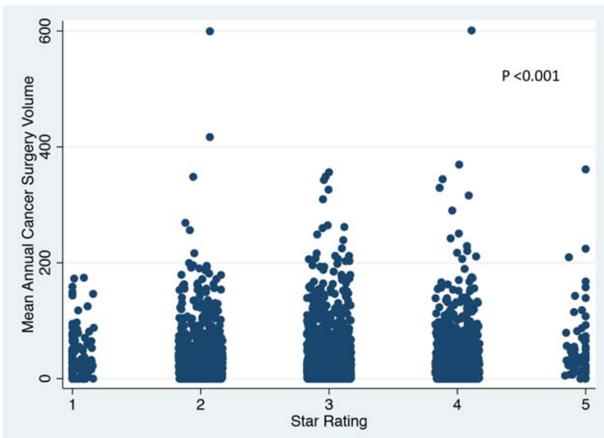

Figure 1. Relationship between the Centers for Medicare and Medicaid Services' Hospital Compare star rating and the average annual major cancer surgery volume among hospitals in the United States.

16% of the 1-star hospitals were in the highest quintile for overall cancer surgery volume ($P < .001$).

Across all hospitals and all procedures, the average frequencies of our measured 30-day outcomes were as follows: 3.4% for mortality, 43.7% for complications, 12.3% for readmissions, and 10.3% for a prolonged length of stay. In univariate and multivariate analyses, both the Hospital Compare star ratings and the surgical volume quintiles were inversely associated with the occurrence of each short-term cancer surgery outcome (Table 2 and Fig. 2). After we controlled for patient and hospital characteristics, 5-star hospitals had a 2.3% mortality rate, whereas 1-star hospitals had a 4.5% rate; the complication rates were 39% and 48% ($P < .001$), the readmission rates were 10% and 15% ($P < .001$), and the rates for a prolonged length of stay were 8% and 16% ($P < .001$) for 5- and 1-star hospitals, respectively. With respect to the surgical volume, the average mortality rates were 2.7% and 5.8% for the highest and lowest volume hospitals, respectively ($P < .001$); the rates of complications were 41% and 55% ($P < .001$), the readmission rates were 12.2% and 11.6% ($P = .195$), and the rates for a prolonged length of stay were 9.4% and 13% ($P < .001$) for hospitals in the highest and lowest quintiles, respectively (Table 2).

TABLE 2. Unadjusted Overall Cancer Surgery Outcomes by the Centers for Medicare and Medicaid Services' Hospital Compare Star Rating and Surgical Volume Quintile

Outcome	Star Rating					
	1 (Lowest)	2	3	4	5 (Highest)	*P*
30-day mortality rate	4.4	3.7	3.5	3.0	2.3	<.001
30-day complication rate	47.1	44.5	43.6	43.1	39.3	<.001
30-day readmissions	14.5	13.2	12.1	11.6	10.3	<.001
Prolonged length of stay	14.7	11.7	10.0	8.9	7.7	<.001
	Surgical Volume Quintile					
Outcome	**1 (Lowest)**	**2**	**3**	**4**	**5 (Highest)**	***P***
30-day mortality rate	5.8	5.6	4.8	4.0	2.7	<.001
30-day complication rate	55.0	52.8	49.7	45.9	40.7	<.001
30-day readmissions	11.6	13.0	12.4	12.2	12.3	.018
Prolonged length of stay	13.1	13.6	12.6	11.1	9.2	<.001

Figure 3 presents a comparison of the performance of star ratings and surgical volumes for predicting each of the cancer surgery outcomes. For 30-day mortality, the line fit to actual mortality quintiles has a slope of –2.48. This contrasts with slopes of –0.58 for the star rating and –0.73 for surgical volume. This, therefore, suggests that, on average, surgical volume and the star rating are similarly associated with the mortality rate. We observed a comparable relationship for the complication rate (Fig. 3). This is in contrast to the slopes for the readmission rate and a prolonged length of stay, which suggest a stronger relationship with the star rating than the surgical volume.

Our sensitivity analyses identified no substantive changes to our principal findings. In our examination of hospital reporting and star ratings, 95% of the hospitals reported either 6 or 7 measures, and there was a similar share of hospitals across star categories that reported on

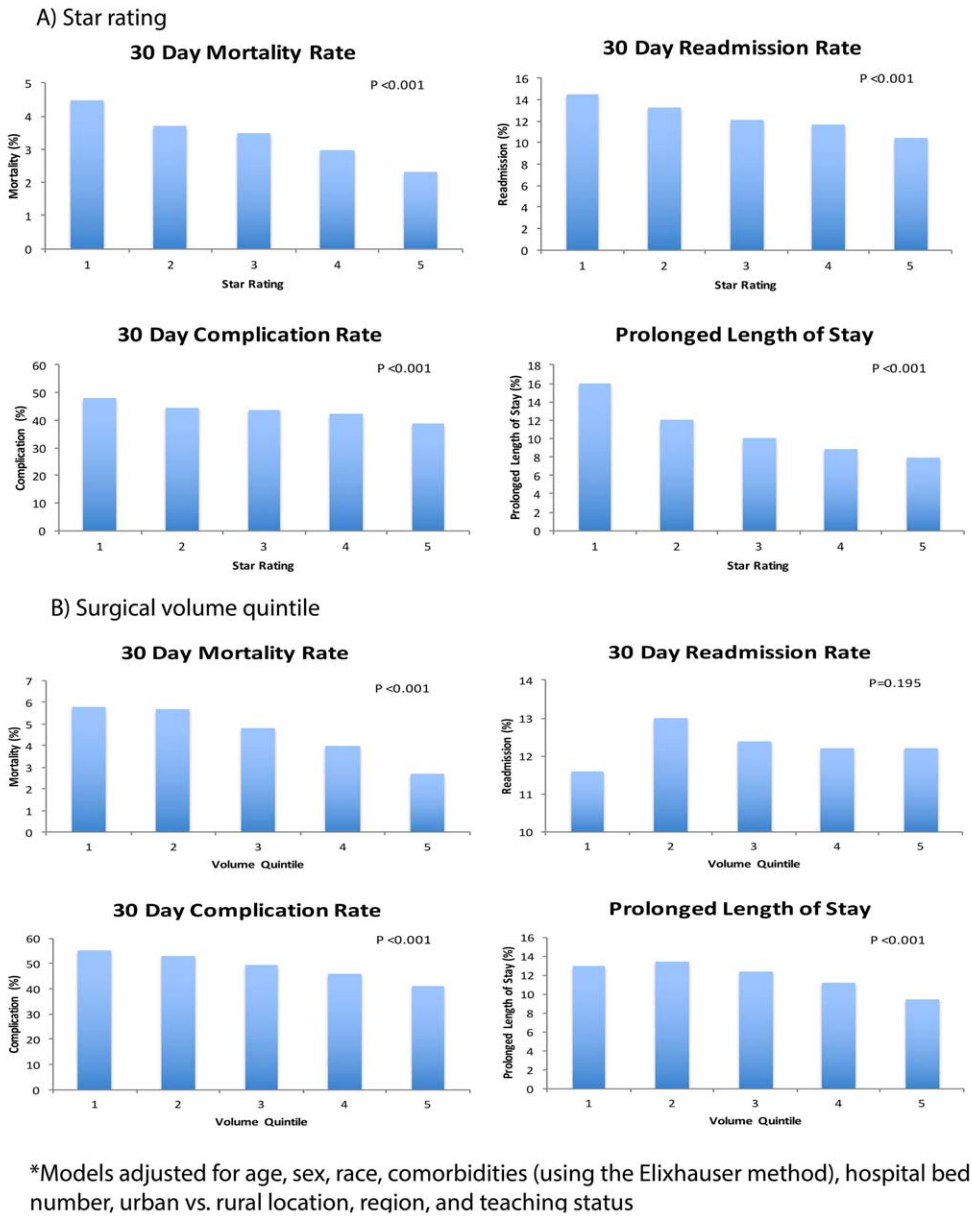

Figure 2. Adjusted overall cancer surgery outcomes by (A) the Centers for Medicare and Medicaid Services' Hospital Compare star rating and (B) surgical volume quintile.

the full 7 measures (89% for 1-star hospitals vs 84% for 5-star hospitals). Among 5-star hospitals, the mean mortality rate, complication rate, and prolonged-length-of-stay rate were highest for those reporting 3 outcome measures (the maximum allowable). Readmission rates were highest among hospitals reporting 2 outcome measures. Lastly, our analysis of the correlation between the overall cancer surgery volume and the cancer-specific volume demonstrated that the cancer-specific volume was at least moderately correlated with the overall cancer surgery volume for 7 of 9 cancers (Supporting Table 1 [see online supporting information]).

DISCUSSION

Across a national sample of hospitals in the United States, we observed no association between CMS' Hospital Compare star ratings and annual cancer surgery volumes. However, both higher star ratings and larger surgical volumes were associated with better short-term cancer surgery outcomes, including lower rates of mortality, complications,

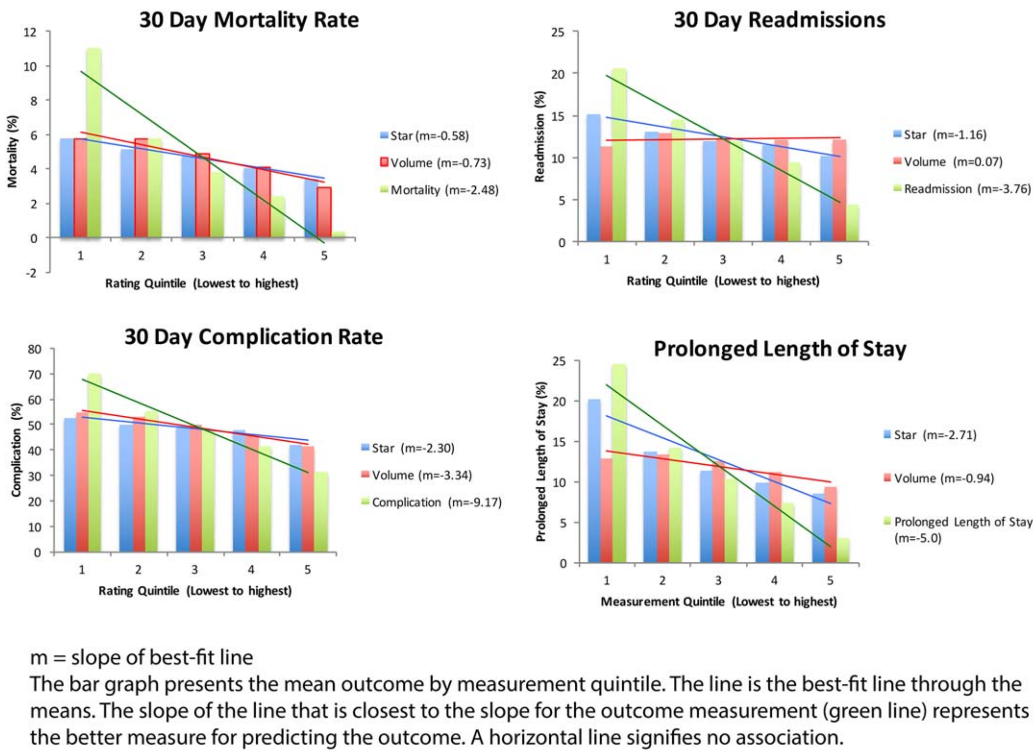

Figure 3. Comparison of goodness-of-fit lines for the Centers for Medicare and Medicaid Services' Hospital Compare star rating and cancer surgery volume with respect to actual measured outcomes.

readmissions, and prolonged length of stay. On average, the surgical volume and the star rating were similarly associated with mortality and complication rates, whereas the star rating had a stronger relationship with readmissions and a prolonged length of stay. Collectively, these findings suggest that in the absence of other information, the star rating may be useful in helping patients to select a hospital for major cancer surgery, but more research is needed before these ratings can supplant surgical volume as a measure of surgical quality.

Our findings are consistent with previous work that has convincingly demonstrated a strong association between higher surgical volumes and better short- and long-term outcomes after major cancer surgery.[16-19] Importantly, however, our observation that there is little, if any, correlation between the star rating and surgical volume for cancer surgery indicates that the Hospital Compare measure cannot be used as a publically available proxy for case volume. In other words, choosing a hospital with a high star rating does not mean that a patient is selecting a high-volume facility.

At the same time, we also identified a significant inverse association between hospital star ratings and short-term cancer surgery outcomes. This is a new finding and, in many ways, is discordant with previous literature demonstrating inconsistencies among differing public reporting systems.[9] Moreover, despite the potentially important implications of this finding for patients seeking cancer care, many have criticized the methodology used to create the star ranking as both flawed and excessively opaque.[2,20-22]

In this context, there are several potential reasons for the observed association between star ratings and short-term cancer surgery outcomes that need to be addressed before star ratings can supplant surgical volume as a measure of surgical quality. Included among these is the possibility that star rankings will prove to be an accurate measure of hospital performance with major cancer surgery. However, alternative explanations must also be considered. Although we controlled for measurable patient comorbidities and carefully assessed case-mix differences between the hospitals, it remains possible that the observed association reflects residual unmeasured differences in cancer severity or comorbidity among patients in the different star rating categories. Our finding could also reflect a tautological relation (ie, a finding that is true because of the way it was conceived) to the extent that some of the outcomes that we measured (eg, readmissions) are also used to assign star ratings in the first place. Although it is difficult to empirically examine this concern, it is worth noting that the Hospital Compare methodology does not include any cancer-specific quality measures. Instead, the Hospital Compare methodology predominantly reflects outcomes, processes, and satisfaction with care for cardiac and pulmonary disease and stroke. In fact, only 9 of 64 possible measures contributing to the star ratings could directly apply to patients undergoing major cancer surgery, and the majority would be demonstrated in our complication outcome.

Finally, the observed association could also reflect the effect of hospitals with small case numbers or hospitals that gamed the star rating system by reporting fewer measures.[2] Although valid, these concerns are mitigated by findings from our sensitivity analyses demonstrating no differences in our findings after we had excluded hospitals with low surgical volumes as well as the findings that 95% of hospitals reported 6 or 7 measures and that the proportion of hospitals reporting on all 7 measures was similar across rating categories (89% for 1-star hospitals vs 84% for 5-star hospitals). Moreover, among hospitals with a 5-star rating, those reporting the maximum number of outcome measures[3] had higher mortality rates, higher complication rates, and increased lengths of stay in comparison with those reporting fewer measures, and this is counter to the argument that gaming is occurring.

Our study has several limitations. First, we examined only short-term cancer surgery outcomes for 9 types of cancer. As a result, our findings may not apply for other outcomes and tumors. That being said, the cancer sites included in our analysis compose approximately 46% of the estimated newly diagnosed invasive cancers in 2016 in the United States.[23] Second, although we are equating the outcomes measured in this analysis with quality of care for cancer surgery, there are obviously many other measures that could be considered, including longer term or patient-reported outcome measures. It is also true, however, that the outcomes assessed herein can be defined and measured accurately with claims data, and they are widely used to measure and compare surgical quality. Third, the care delivered occurred before the star rating. However, the CMS devised the star rating system on the basis of previous years' measures, and our sensitivity analyses demonstrated that outcomes did not vary over time. Fourth, although we excluded some patients undergoing cancer surgery to increase generalizability and statistical reliability, this step may have resulted in a biased sample. However, because of the very small proportion of cases (1.7% of the total sample) affected by our exclusion criteria, the implications for our overall findings are likely limited. Last, we used overall surgical volume when one could argue that the more clinically relevant measure is cancer-specific volume. Nonetheless, we selected overall volume because it is more consistent with evaluating the hospital as a whole and is comparable to the Hospital Compare 5-star system. Furthermore, our sensitivity analysis indicated that cancer-specific volumes correlated fairly well with overall cancer surgery volumes.

These limitations notwithstanding, our findings have important implications for patients, policymakers, payers, and hospital administrators. For patients, our analyses support using surgical volume, when available, as a guide for choosing where to have major cancer surgery. Our findings also suggest a potential role for the Hospital Compare star rankings; however, because of the preliminary nature of this measure, surgical volume currently remains a more acceptable metric of a hospital's average short- and long-term cancer surgery outcomes.

For CMS policymakers, our findings do suggest that star ratings may capture real differences between hospitals for cancer surgery outcomes. Nevertheless, important questions still remain about how hospitals get into one category versus another, about the opacity of the measurement methodology, and about the extent to which reporting differences influence a hospital's rating. Thus, although our findings do provide support for additional assessment of these measures, several concerns must be addressed before the star rating can be viewed as an actionable and reliable measure of quality. An important next step would be to analyze how the measure performs for disease states (eg, gastrointestinal bleeds) that are far removed from the conditions contributing to the measure. Finally, because our findings suggest that star rankings may have some validity as a hospital quality measure, hospital administrators may view these results as an additional motivation to measure and improve performance on the various processes and outcomes that collectively yield the Hospital Compare star measure.

In the future, additional research needs to be performed on case-mix differences, tautology, reporting bias, and other procedures and diagnoses. It will also be important to examine the stability of this measure over time and its relationship with other measures of cancer care quality. Thus, although the Hospital Compare star rating appears to be associated with short-term cancer surgery outcomes, more evaluation is required before it can be used to make selecting a hospital for cancer surgery any easier.

FUNDING SUPPORT

This project was supported by the National Cancer Institute (1-R01-CA-174768-01-A1 to David C. Miller and 5-T32-CA-180984-03 to Deborah R. Kaye).

CONFLICT OF INTEREST DISCLOSURES

James M. Dupree reports a grant from Blue Cross Blue Shield of Michigan outside the submitted work. David C. Miller reports a contract for serving as the director of the Michigan Urological Surgery Improvement Collaborative; part of his salary at the University of Michigan is paid through this contract.

AUTHOR CONTRIBUTIONS

Deborah R. Kaye: Conception, design, data analysis, interpretation, writing, and review. **Edward C. Norton**: Conception, design, data analysis and interpretation, and review. **Chad Ellimoottil**: Conception, design, data analysis, interpretation, writing, and review. **Zaojun Ye**: Data analysis, interpretation, and review. **James M. Dupree**: Design, data interpretation, and review. **Lindsey A. Herrel**: Design, data interpretation, and review. **David C. Miller**: Conception, design, data analysis, interpretation, writing, and review.

REFERENCES

1. Pronovost PJ. Why surgical volumes should be public. http://health.usnews.com/health-news/blogs/second-opinion/articles/2016-11-18/why-surgical-volumes-should-be-public?src=usn_tw. Accessed November 2016.
2. Xu S, Grover A. CMS' hospital quality star ratings fail to pass the common sense test. http://healthaffairs.org/blog/2016/11/14/cms-hospital-quality-star-ratings-fail-to-pass-the-common-sense-test/. Accessed November 2016.
3. Chhabra, K. Reporting surgical volumes would be more useful & less misleading than "outcomes"/star ratings. https://twitter.com/krchhabra/status/802938217666408448. Accessed Dec 2016.
4. Dudley RA. More evidence of assoc bw social factors & hosp ratings: hosps in cities w stressed out populations get fewer stars. https://twitter.com/search?q=%40RAdamsDudleyMD.%20%22More%20evidence%20of%20assoc%20bw%20social%20factors%20%26%20hosp%20ratings%3A%20hosps%20in%20cities%20w%20stressed%20out%20populations%20get%20fewer%20stars&src=typd. Accessed Dec 2016.
5. Goodrich K, Conway P. CMS hospital star ratings first step in effort to improve quality measures. https://news.aamc.org/patient-care/article/cms-star-ratings-effort-improve-quality-measures. Accessed October 2016.
6. Comarow A. FAQ: how and why we rank and rate hospitals. http://health.usnews.com/health-care/best-hospitals/articles/faq-how-and-why-we-rank-and-rate-hospitals. Accessed October 2016.
7. Merrill AL, Jha AK, Dimick JB. Clinical effect of surgical volume. *N Engl J Med.* 2016;374:1380-1382.
8. Centers for Medicare and Medicaid Services. First release of the overall Hospital Compare quality star rating on Hospital Compare. https://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2016-Fact-sheets-items/2016-07-27.html. Published July 27, 2016. Accessed August 2016.
9. 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;34:423-430.
10. Sternberg S. Hospitals move to limit low-volume surgeries. https://www.usnews.com/news/articles/2015/05/19/hospitals-move-to-limit-low-volume-surgeries. Accessed August 2016.
11. Data.Medicare.gov. Hospital Compare datasets. https://data.medicare.gov/data/hospital-compare. Accessed July 25, 2016.
12. Iezzoni LI, Mackiernan YD, Cahalane MJ, et al. Screening inpatient quality using post-discharge events. *Med Care.* 1999;37:384-398.
13. Osborne NH, Nicholas LH, Ryan AM, et al. Association of hospital participation in a quality reporting program with surgical outcomes and expenditures for Medicare beneficiaries. *JAMA.* 2015;313:496-504.
14. Tan HJ, Norton EC, Ye Z, et al. Long-term survival following partial vs radical nephrectomy among older patients with early-stage kidney cancer. *JAMA.* 2012;307:1629-1635.
15. Herrel LA, Norton EC, Hawken SR, et al. Early impact of Medicare accountable care organizations on cancer surgery outcomes. *Cancer.* 2016;122:2739-2746.
16. Finks JF, Osborne NH, Birkmeyer JD. Trends in hospital volume and operative mortality for high-risk surgery. *N Engl J Med.* 2011;364:2128-2137.
17. Morche J, Mathes T, Pieper D. Relationship between surgeon volume and outcomes: a systematic review of systematic reviews. *Syst Rev.* 2016;5:204.
18. Trinh QD, Bjartell A, Freedland SJ, et al. A systematic review of the volume-outcome relationship for radical prostatectomy. *Eur Urol.* 2013;64:786-798.
19. Birkmeyer JD, Stukel TA, Siewers AE, et al. Surgeon volume and operative mortality in the United States. *N Engl J Med.* 2003;349:2117-2127.
20. Medicare Payment Advisory Commission. RE: hospital star rating system. http://www.medpac.gov/docs/default-source/comment-letters/20160922_hospitalstarratings_medpac_comment_sec.pdf?sfvrsn=0. Published September 22, 2016. Accessed October 2016.
21. MacLean C, Shapiro L. Does the Hospital Compare 5-star rating promote public health? http://healthaffairs.org/blog/2016/09/08/does-the-hospital-compare-5-star-rating-promote-public-health/. Published September 8, 2016. Accessed October 2016.
22. Bilimoria KY, Barnard C. The new CMS hospital quality star ratings: the stars are not aligned. *JAMA.* 2016;316:1761-1762.
23. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin.* 2016;66:7-30.