Published in final edited form as:

Surgery. 2018 September ; 164(3): 379–386. doi:10.1016/j.surg.2018.04.010.

Commercial Quality “Awards” Are Not a Strong Indicator of Quality Surgical Care

Adrienne N. Cobb, MD^{1,2}, Taylor R. Erickson, BS¹, Anai N. Kothari, MD^{1,2}, Emanuel Eguia, MD^{1,2}, Sarah A. Brownlee, BS¹, Weiwei Yao, BS³, Hyunyou Choi, BS³, Victoria Greenberg, BS³, Joy Mboya, BS³, Michael Voss, BS³, Daniela Stan Raicu, PhD³, Raffaella Settimi-Woods, PhD³, and Paul C. Kuo, MD^{2,3}

¹Loyola University Medical Center, Department of Surgery, Maywood, IL

²One:MAP Section of Surgical Analytics, Department of Surgery, Loyola University Chicago, Maywood, IL

³University of South Florida, Department of Surgery, Tampa, FL

⁴DePaul University, College of Computing and Digital Media, Chicago, IL

Abstract

BACKGROUND—This study aimed to determine if publicized hospital rankings can be used to predict surgical outcomes.

METHODS—Patients undergoing one of nine surgical procedures were identified using the Healthcare Cost and Utilization Project State Inpatient Database for Florida and New York 2011–2013 and merged with hospital data from the AHA Annual Survey. Nine quality designations were analyzed as possible predictors of inpatient mortality and postoperative complications using logistic regression, decision trees and support vector machines.

RESULTS—We identified 229,657 patients within 177 hospitals. Decision trees were the highest performing machine learning algorithm for predicting inpatient mortality and postoperative complications (accuracy 0.83, $P < .001$). The top 3 variables associated with low surgical mortality (relative impact) were Hospital Compare (42), total procedure volume (16) and, Joint Commission (12). When analyzed separately for each individual procedure, hospital quality awards were not predictors of postoperative complications for seven of the nine studied procedures. However, when

---

Corresponding Author: Paul C. Kuo, MD, MS, MBA, Professor and Chair, USF Dept. of Surgery, Professor, USF Muma College of Business, Richard G. Connar Endowed Chair in Surgery, Surgeon in Chief, Tampa General Hospital, 2 Tampa General Circle, Rm 7015, Tampa, FL 33606, Phone: (813) 250-2572.

Disclosure

The authors have no conflicts of interest or financial disclosures.

Meeting Information

This study was presented during the Central Surgical Association Annual Meeting held on July 30–August 1, 2017 in Chicago, Illinois.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

grouping together procedures with a volume-outcome relationship, hospital ranking becomes a significant predictor of postoperative complications.

CONCLUSIONS—Hospital quality rankings are not a reliable indicator of quality for all surgical procedures. Hospital and provider quality must be evaluated with an emphasis on creating consistent, reliable, and accurate measures of quality that translate to improved patient outcomes.

Keywords

quality improvement; hospital quality; machine learning; surgical outcomes; quality awards

INTRODUCTION

Hospital ranking systems such as U.S. News & World Report, Health Grades, and the Center for Medicare Services attempt to measure healthcare quality using a variety of hospital structure, process, and outcome data.¹ These efforts have developed out of a demand for increased transparency and accountability. Ultimately, the intent of each organization that publishes ranking systems centers upon identifying hospitals that provide the highest quality care and inform patients, clinicians, and hospitals.¹

The best way to use these rankings remains elusive.² Despite their goals to aid healthcare consumers in choosing the best place to receive their care, there are often discrepancies across ranking systems in their assessment of quality, methods for creating rankings, and data used to assign ranks.^{3–4} The challenges inherent to ranking systems often lead to confusion for hospitals, insurance providers, and patients on how best to interpret and apply their rankings.⁵ One particularly enticing application of hospital ranking is to use the information to predict patient-level outcomes. Given the ease of identifying hospitals that earn public rankings, they could serve as a powerful tool for assisting patients in finding the best place to receive care. Currently, it remains unclear if public ranking systems can be used to forecast clinical outcomes in both medical and surgical populations successfully.⁶ It is also unknown if some hospital ranking systems are better than others at predicting clinical outcomes.^{7–8} The objective of this study was to use machine learning techniques to determine if publically available hospital rankings can be used to predict surgical morbidity and mortality across a broad range of inpatient operations.

METHODS

Data Sources and Patient Selection

This analysis used data from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Database (SID) for Florida and New York from 2011 to 2013. Patients undergoing pulmonary lobectomy, mastectomy, pancreatectomy, total knee/hip replacement, laparoscopic cholecystectomy, prostatectomy, bariatric surgery, and abdominal aortic aneurysm (AAA) repair were identified using *International Classification of Diseases, 9th Revision, Clinical Modification*, codes. (Supp. Table 1) These procedures were selected to cover a range of high-, medium-, and low-risk procedures based on the American College of Cardiology/American Heart Association guidelines of cardiac risk of non-cardiac surgery and a modified Johns Hopkins surgical severity criteria.^{9–11} The SID is an administrative,

all-payer dataset aggregated by the Agency for Healthcare Research and Quality (AHRQ) to inform health-related decisions.¹² The Institutional Review Board at our institution deemed the study exempt from review as the data are de-identified, protected and publicly available.

Ranking information was collected via the Informed Patient Institute and Google® searches.^{13–20} Organizations included for study are displayed in Table 1.²¹ Because rankings are based on hospital data from the previous year, all discharges from 2013 and all rankings from 2011 were removed, as there were no 2014 rankings or 2010 discharge data available. An aggregate accolade score for each ranked hospital was created by adding the number of designations received. A binary variable representing hospital ranking was created for analysis, with 0 indicating low ranked hospitals (2 or fewer rankings in total) and 1 indicating high ranked hospitals (3 or more quality rankings).

Standard data pre-processing was completed to address missing values, taking the aggregated inpatient dataset from 347,858 total discharges to the final dataset which yielded 229,657 patients. The dataset contained hospital rankings, patient demographics, mortality information and comorbidities. Patient comorbidities included were as defined by AHRQ and pre-existent in the SID.¹² Mortality was defined as death prior to discharge from the hospital and was included as a binary variable. The use of inpatient mortality to assess hospital quality is debated, but it is often one of the primary benchmarks used by the organizations responsible for hospital rankings, and therefore felt to be appropriate here. For three of the nine procedures (AAA repair, hip replacement, pancreatectomy) included here, mortality has actually been advocated as a quality indicator by the AHRQ.²² Postoperative complications were assessed using an aggregate complication variable which included death and inpatient outcomes. Measured inpatient complications were abstracted using ICD-9-CM codes and are included in Supplemental Table 2.²³

Machine Learning Techniques

Data were analyzed to determine if quality rankings were predictors of inpatient mortality and postoperative complications among patients undergoing one of nine surgeries at 177 hospitals with various hospital quality designations. The original dataset was split into two randomly selected samples with 70% of the data used to train the model and the remaining 30% used for model testing. Data analysis included several classification models such as decision trees, logistic regression, and support vector machines. Each model's performance was assessed using sensitivity, specificity, and accuracy.

Exploratory analysis of the data revealed high variability in complication rates by procedure. As a result, the surgeries were divided into high and low-risk groups using a heuristic 5% cutoff value for complication rates derived by the observed distribution of the data. High-risk surgeries, defined as surgeries with complication rates greater than 5%, included open AAA repair, pancreatectomy, and pulmonary lobectomy. Since the original dataset had an imbalanced class distribution for complications with only a minority of patients experiencing complications, models were trained using balanced datasets with a 50/50 split of complications. The balanced datasets were created by selecting all the patients with complications and a random sample of patients without complications for both low and high-risk surgery patients.

A second analysis utilized procedure-specific decision trees as well as logistic regression models but focused on the complication rates by hospital and procedure rather than inpatient mortality. Its goal was to investigate the utility of quality rankings by predicting the occurrence of inpatient complications across procedures. This was performed on each individual procedure as well as two subsets of the population, those patients that had strict inpatient only procedures (prostatectomy, AAA repair, and pulmonary lobectomy) and those that had procedures where a strong volume-outcome relationship is noted (bariatric surgery, pancreatectomy, total hip and knee replacements) with the goal of focusing on procedures that may be more representative of hospital quality. While some of the volume-outcome procedures are performed on an inpatient only basis as well, we did not want the two groups to overlap, for concern that the volume-outcome relationship would overpower the analysis and skew the results of the inpatient procedure subgroup. Only hospitals with 10 or greater procedures for each surgery were considered for analysis to ensure adequate sample size. The complication rate for each surgery and hospital was computed by dividing the number of patients that had complications after receiving the surgery in an individual hospital by the total number of patients receiving that surgery. The mean complication rates for hospitals with different rankings were compared using the non-parametric Mann-Whitney U two-sample test. Statistical analyses were computed using the R statistical software (Version 3.3.2 Vienna, Austria) and SPSS (IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.)

RESULTS

We identified 101,876 complete patient records in New York, and 127,781 complete patient records in Florida in the years 2011 and 2012. Baseline characteristics of the study population are shown in Table 2. The most commonly performed procedure was total knee replacement with 87,032 patients undergoing this procedure, followed closely by total hip replacement (53,866 patients). Open AAA repair was the least commonly performed procedure in the cohort (795 patients). The overall mortality rate for the study population was 0.4% (824/229,657), ranging from 0.1% (9/12,826, mastectomy) to 4.8% (38/800, open AAA repair) (Table 3a). The overall inpatient complication rate was 2.7% (6,273/229,657), ranging from 1.0% (128/12,826, mastectomy) to 20.0% (162/795, open AAA repair). The mean complication rate for each procedure is shown in Table 3b.

A total of 177 hospitals were included in the analytic cohort. Most hospitals received either zero or one quality ranking, 87% of hospitals receiving at most one quality ranking in 2012 and 81% in 2013 (Table 4). The median number of hospital quality rankings awarded to hospitals was 1 (IQR 0 – 1) for both years. No hospitals were awarded greater than 5 quality rankings; 43% (N=77) hospitals in 2012 and 22% (N=39) hospitals in 2013 were awarded no quality rankings, and 3.4% (N=6) received either 4 or 5 quality rankings in both 2012 and 2013. After grouping procedures into risk categories, an increasing number of quality rankings was associated with a decreased incidence of complications for low and high-risk procedures (Figure 1). No association between the number of quality rankings and inpatient mortality was noted.

Machine learning algorithms were used to determine if quality rankings could predict perioperative outcomes. Decision trees computed using the CRT-GINI growing method, performed the best with the highest accuracy in predicting inpatient mortality and postoperative complications for the testing set. A decision tree is a powerful classification algorithm that can be represented by a tree-like structure, where the nodes represent the most critical variables to predict the target variable.²³ A representative decision tree schematic is shown in Supplemental Figure 1. Decision tree analysis for inpatient mortality had a sensitivity of 78%, specificity of 92% and overall accuracy of 83% (95% CI 0.68–0.92) (Table 5a). Hospital Compare, Joint Commission, and Becker's Review were the top predictors of low inpatient mortality (Table 5b). U.S. News & World Report label had no predictive power. Decision tree was also the highest performing machine learning algorithm for predicting surgical complications with a sensitivity of 100%, specificity of 90%, and accuracy of 93%. Hospital Compare, Becker's Hospital Review, and Joint Commission were predictive of complications (Table 5b). Of note, Magnet designation was a negative predictor of outcomes indicating that those hospitals with Magnet designation were more likely to have perioperative morbidity and mortality.

Procedure-specific decision tree analyses were performed to determine if hospital quality rankings could predict outcomes for specific surgeries. The accuracy of individual trees ranged from 64.1% – 74.7% (Table 6). Hospital ranking was a top predictive factor for only two of the studied procedures: total hip replacement and mastectomy (Supp. Fig 2). Factors such as electrolyte imbalance and comorbidities were stronger predictors of perioperative complications. When examining inpatient only procedures such as prostatectomy, open AAA repair, and pulmonary lobectomy, we found that hospital ranking was not a predictor of postoperative complications. Logistic regression outperformed logistic regression with an accuracy of 79%, sensitivity of 68% and specificity of 80%. (Table 7) Rather than hospital ranking, the most important predictors of complications were comorbidities similar to those in the overall analysis including electrolyte disorders, chronic lung disease, weight loss, coagulopathy, and pulmonary circulation disorders.

Conversely, when evaluating patients who had procedures associated with a volume-outcome relationship (bariatric surgery, pancreatectomy, total hip and knee replacements), hospital ranking was important. We found a significant association with individual high volume procedures. The top 5 predictors of postoperative complication in order of importance, via decision tree analysis were electrolyte disturbance, quality ranking, chronic lung disease, CHF and pulmonary circulation disorders. The logistic regression analysis showed that the number of rankings received by a hospital is negatively correlated to the probability of a complication. (Supplemental Table 3) Patients who got their procedures at hospitals with less than 4 rankings increased the odds of a complication when compared with patients who had their procedures at hospitals with 4 or 5 rankings. (Figure 2)

DISCUSSION

In this study, we used novel machine learning models to examine a set of publically available quality rankings/designations and their ability to predict hospital quality using outcomes following a range of low to high-risk procedures. Of the nine rankings included,

we found that only two, Hospital Compare and Becker's Review, were essential factors in predicting inpatient mortality and postoperative complications. This is interesting considering that Becker's Review is compiled using nominations along with other hospital rating systems and CMS Hospital Compare often serves as a data source for many other hospital quality public reporting organizations.²¹ Magnet Nursing designation showed a negative relationship with outcomes. While literature exists to the contrary²⁵, this highlights the inconsistency that exists in using these designations to assess quality.

As we recognize that inpatient mortality is not the best indicator of hospital quality²⁶, we performed a more in-depth analysis focusing on the ability of these rankings to adequately predict inpatient postoperative complication rates. We found that high-risk procedures such as open AAA and pancreatectomy had similar complication rates across all hospitals regardless of the number of quality rankings. It is possible that high-risk procedures have already been regionalized and pre-selected into a limited group of hospitals and are therefore performed in settings with the requisite infrastructure and experience. The quality designations would likely not differentiate these hospitals unless more specific metrics are developed. Low-risk procedures had lower complication rates for hospitals with a higher number of quality rankings. However, we also found that those hospitals with higher quality rankings had higher procedure volume, which may confound the association. This study did not specifically address hospital volume, but the volume-outcome relationship is well studied.^{27, 28} As a result, we did look at procedures that are associated with a volume-outcome relationship and noted that hospital ranking was important in predicting postoperative complications. This may suggest that these rankings could be useful in this subset of procedures, but cannot be generalized to all hospitals since a hospital that does a high volume of bariatric surgery may not necessarily perform a high volume of Whipples.

Finally, our procedure-specific decision tree analysis showed that the presence of a quality ranking was only an important feature for total hip replacement and mastectomy to a lesser degree. It is difficult to elucidate the exact reason, but hospital ranking may matter for total hip replacement because they are common, often managed with standard care pathways, and are the focus of a Medicare program to link hospital quality with reimbursement.²⁹ Across most of the procedures considered, we did not find a definite association between quality rankings and surgical outcomes.

Several studies have examined the relationship between individual rankings and quality of care for various medical conditions via outcomes such as risk-adjusted mortality, length of stay, and patient satisfaction. There are, however, few that have examined the impact of these rankings on surgical outcomes.³⁰ Osborne et al. examined whether selecting highly rated hospitals led to improved outcomes in cardiovascular surgery procedures, finding that they did, though similar outcomes could be attained at high volume institutions.²⁸ Conversely, Lascano et al. evaluated the correlation between publicized health ranking systems and outcomes following radical cystectomy and found that rankings were not associated with any clinical outcomes.³¹

It is also possible that these rankings are simply better suited to predict hospital quality for medical outcomes. In a study by Wang et al.³² they discussed the association between the

Centers for Medicare and Medicaid Services (CMS) hospital star rating and patient outcomes across three non-surgical conditions: acute myocardial infarction, pneumonia, and heart failure. Their primary outcome was a composite for mortality and readmissions. Unlike in our study, they found that the number of stars received by a hospital was inversely associated with risk-adjusted mortality rate and adjusted readmission rates. They concluded that CMS star ratings were reliable in guiding healthcare decision making for patients.

Public reporting of clinical data is now commonplace and therefore has spawned much research. There is evidence that many of these quality ratings are not sufficient to assess hospital quality. Prior literature reveals discrepancies between different hospital ranking systems and methodologies.³³ In general, most hospitals do not receive a high ranking from more than one designation group.

According to one study investigating the relationship between mortality and Magnet status of hospitals, 30-day mortality for surgical patients was significantly lower in Magnet Hospitals as compared to non-Magnet Hospitals.³⁴ Conversely, two studies investigating the correlation of ranking systems and surgical outcomes status post radical cystectomy or oncologic surgery failed to find a strong association between publicized rankings and superior surgical outcomes.^{31,35} When the information provided by these rankings is inconsistent, it may give patients and hospitals a false sense of security regarding the care they receive and provide. While it is important to ensure the validity of such quality metrics for the prospects of patient care, we cannot overlook the financial and temporal investments that come with designations such as Joint Commission accreditation and Magnet. Hospital administrations want to know that there is a return on their investment in the form of measurable improvement in structure and process measures as well as patient outcomes.

Our study has several limitations in that the data are retrospective and taken from an administrative database. First, we were limited in the years available for study. We examined years 2011–2013 which are dated, particularly given the change in healthcare climate in 2014 with phasing in of the Affordable Care Act (ACA). Repeating this study in the post-ACA era when there are possibly more people utilizing healthcare would be interesting, though we suspect the results may be similar.

Additionally, our study only examines short-term (inpatient) outcomes. The shortcomings of using inpatient mortality as a measure of hospital quality have been previously discussed. The addition of other outcomes such as readmissions as well as extending the time for outcomes (i.e., 30 days) would add value.

Our study examines a small subset of surgical procedures and may not be generalizable to every surgical specialty or procedure. Low-risk procedures like laparoscopic cholecystectomy that are often performed on an outpatient basis, but were included to evaluate a diverse case mix. Future work may include a study that focuses solely on a larger group of high-risk procedures.

Finally, the results of this study cannot be deemed causal. They are meant to be hypothesis-generating to fuel further research. The data is also subject to substantial selection bias.

However, the use of training and test sets in our machine learning analysis serves to address this through cross-validation.

There is no doubt that the varying methodologies of hospital quality public reporting contribute to the inconsistency in results. A centralized quality reporting mechanism that exists separately from the current quality industry, using specific consensus metrics that are periodically evaluated for relevance, would be ideal.^{4,20,23} With the advent of EHRs and Health Information Exchanges, these data can be theoretically mined at a very granular level to eventually match the patient with surgeons, hospitals, and/or regions to maximize outcomes and minimize cost. This would represent personalized medicine of the phenotypic variety.

CONCLUSIONS

Hospital quality rankings are inconsistent in their ability to predict perioperative mortality and morbidity. This may suggest that these designations are inadequate indicators of hospital quality when assessing surgical outcomes. Hospital and provider quality must be evaluated with an emphasis on creating consistent, reliable, and accurate measures that translate to improved patient outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Support:

This work is supported by NIH T32 GM08750-16.

Thank you to the Loyola University Chicago Burn Shock Trauma Institute for their continued support of this work, the Loyola Department of Surgery and One: MAP Analytics family for their contributions to the conceptualization and editing of this manuscript. We thank the School of Computing of DePaul University for their continued collaboration and advice concerning statistical and machine learning analyses.

References

1. Rothberg MB, Morsi E, Benjamin EM, Pekow PS, Lindenauer PK. Choosing the best hospital: the limitations of public quality reporting. *Health Affairs*. 2008 Nov 1; 27(6):1680–7. [PubMed: 18997226]
2. Emmert Martin, et al. Public reporting of hospital quality shows inconsistent ranking results. *Health Policy*. 2017; 121(1):17–26. [PubMed: 27890391]
3. Shahian David M, , et al. Rating the raters: the inconsistent quality of health care performance measurement. *Annals of surgery*. 2016; 264(1):36–38. [PubMed: 26756770]
4. Shahian DM, Mort EA, Pronovost PJ. The quality measurement crisis: An urgent need for methodological standards and transparency. *Joint Commission Journal on Quality and Patient Safety*. 2016; 42(10):435–438. [PubMed: 27712601]
5. Austin J Matthew, et al. National hospital ratings systems share few common scores and may generate confusion instead of clarity. *Health Affairs*. 2015; 34(3):423–430. [PubMed: 25732492]
6. Bilimoria Karl Y, Barnard Cynthia. The New CMS Hospital Quality Star Ratings: The Stars Are Not Aligned. *Jama*. 2016; 316(17):1761–1762. [PubMed: 27802552]

7. Friese Christopher R, , et al. Hospitals in ‘Magnet’ program show better patient outcomes on mortality measures compared to non-‘Magnet’ hospitals. *Health Affairs*. 2015; 34(6):986–992. [PubMed: 26056204]
8. Barnes HilaryRearden JessicaMcHugh Matthew D. Magnet® Hospital Recognition Linked to Lower Central Line-Associated Bloodstream Infection Rates. *Research in nursing & health*. 2016; 39(2): 96–104. [PubMed: 26809115]
9. Lee TH, Marcantonio ER, Mangione CM, et al. Derivation and Prospective Validation of a Simple Index for Prediction of Cardiac Risk of Major Noncardiac Surgery. *Circulation*. 1999; 100(10) <http://circ.ahajournals.org.archer.luhs.org/content/100/10/1043.long>.
10. Donati A, Ruzzi M, Adrario E, et al. A new and feasible model for predicting operative risk. *Br J Anaesth*. 2004; 93(3):393–399. DOI: 10.1093/bja/ae210 [PubMed: 15220171]
11. Fleishcer LA. American College of Cardiology/American Heart Association. Cardiac risk stratification for noncardiac surgery: Update from the American College of Cardiology/American Heart Association 2007 guidelines. *Cleve Clin J Med*. 2009; 76(Suppl_4):S9–S15. DOI: 10.3949/ccjm.76.s4.02 [PubMed: 19880841]
12. HCUP Databases. Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality; Rockville, MD: Available at: www.hcup-us.ahrq.gov/sidoverview.jsp [Accessed May 2017]
13. Truven Health Analytics. [Accessed June 2016] 100 Top hospitals: study overview. <http://100tophospitals.com/portals/2/assets/100-Top-Study-March-2014.pdf>
14. HealthGrades. [Accessed June 2016] <http://cdn.mm-health.com/7e/652a00f8d011e08a3c12313d033e31/file/NEW%20CONSUMERISM%20PDF.pdf>
15. The Joint Commission. [Accessed June 2016] https://www.jointcommission.org/accreditation/top_performers.aspx; https://www.jointcommission.org/facts_about_the_tracer_methodology/
16. Gamble H, Herman B. [Accessed June 2016] 100 Great hospitals in America. 2012. <http://www.beckershospitalreview.com/lists/100-great-hospitals.html>
17. US News & World Report. [Accessed June 2016] U.S. News Hospitals Rankings and Ratings. 2011–2012. https://www.usnews.com/pubfiles/7-17AdultMethReport2012_Final.pdf
18. Medicare.gov. [Accessed June 2016] What is Hospital Compare?. <https://www.medicare.gov/hospitalcompare/about/what-is-HOS.html>
19. American Nurses Credentialing Center Web site. [Accessed February 15, 2017] Magnet. <http://www.nursecredentialing.org/Magnet/ProgramOverview>, <http://www.nursecredentialing.org/MagnetScheduleFees>
20. Himss Analytics. [Accessed June 2016] EMRAM Overview and Criteria Update. <http://www.himssanalytics.org/presentations/EMRAM-overview>
21. Hwang W, Derk J, LaClair M, Paz H. Finding Order in Chaos: A Review of Hospital Ratings. *Am J Med Qual*. 2016 Mar-Apr;31(2):147–155. [PubMed: 25381001]
22. AHRQ Quality Indicators—Guide to Inpatient Quality Indicators: Quality of Care in Hospitals—Volume Mortality, and Utilization. Rockville, Md: Agency for Healthcare Research and Quality; 2002. AHRQ Publication 02-R0204
23. Best WR, Khuri SF, Phelan M, Hur K, Henderson WG, Demakis JG, Daley J. Identifying patient preoperative risk factors and postoperative adverse events in administrative databases: results from the Department of Veterans Affairs National Surgical Quality Improvement Program. *Journal of the American College of Surgeons*. 2002 Mar 31; 194(3):257–66. [PubMed: 11893128]
24. Podgorelec V, Kokol P, Stiglic B, Rozman I. Decision trees: an overview and their use in medicine. *Journal of medical systems*. 2002 Oct 1; 26(5):445–63. [PubMed: 12182209]
25. Jayawardhana J, Welton JM, Lindrooth RC. Is there a business case for Magnet hospitals? Estimates of the cost and revenue implications of becoming a Magnet. *Medical care*. 2014 May 1; 52(5):400–6. [PubMed: 24535022]
26. Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality: the problem with small sample size. *Jama*. 2004 Aug 18; 292(7):847–51. [PubMed: 15315999]
27. Birkmeyer JD, Siewers AE, Finlayson EV, Stukel TA, Lucas FL, Batista I, Welch HG, Wennberg DE. Hospital volume and surgical mortality in the United States. *New England Journal of Medicine*. 2002 Apr 11; 346(15):1128–37. [PubMed: 11948273]

28. Osborne NH, Nicholas LH, Ghaferi AA, Upchurch GR, Dimick JB. Do popular media and internet-based hospital quality ratings identify hospitals with better cardiovascular surgery outcomes? *Journal of the American College of Surgeons*. 2010 Jan 31; 210(1):87–92. [PubMed: 20123337]
29. Keswani A, Uhler LM, Bozic KJ. What quality metrics is my hospital being evaluated on and what are the consequences? *The Journal of arthroplasty*. 2016 Jun 30; 31(6):1139–43. [PubMed: 27101770]
30. Pronovost PJ, Lilford R. Analysis & commentary: A road map for improving the performance of performance measures. *Health Affairs (Project Hope)*. 2011; 30(4):569–573. [PubMed: 21471474]
31. Lascano D, Finkelstein J, Barlow L, et al. The correlation of media ranking's "best" hospitals and surgical outcomes following radical cystectomy for urothelial cancer. *Urology*. 2015; 86(6):1104. [PubMed: 26408500]
32. Wang David E, , et al. Association between the Centers for Medicare and Medicaid Services hospital star rating and patient outcomes. *JAMA internal medicine*. 2016; 176(6):848–850. <https://jamanetwork-com.archer.luhs.org/journals/jamainternalmedicine/fullarticle/2513630>. [PubMed: 27062275]
33. Ross C. [Accessed February 2017] 5 reasons to take hospital ratings with a big grain of salt. *Stat*. 2016. Aug 2, <https://www.statnews.com/2016/08/02/hospital-ratings-skepticism/>
34. McHugh MD, Kelly LA, Smith HL, Wu ES, Vanak JM, Aiken LH. Lower mortality in magnet hospitals. *Medical Care*. 2013; 51(5):382–388. [PubMed: 23047129]
35. Wright JD, Tergas AI, Ananth CV, Burke WM, Chen L, Neugut AI, Richards CA, Hershman DL. Relationship between surgical oncologic outcomes and publically reported hospital quality and satisfaction measures. *Journal of the National Cancer Institute*. 2015 Jan 24.107(3):dju409. [PubMed: 25618899]

SUMMARY SENTENCES

We examined nine common quality awards and their impact on surgical outcomes; finding that they are not a surrogate for quality surgical care. The significance of this finding demonstrates the need for reliable, accurate quality measures to guide decision making.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

Complication Rate in High vs Low Risk Procedures by Hospital Ranking. Complication rate plotted against number of rankings obtained. Low risk surgeries are in blue. High risk surgeries are in red. Each data point represents the complication rate of a particular surgery in a hospital in a certain year. The size of the dot indicates the number of patients receiving that procedure at a particular hospital. High risk surgeries include pulmonary lobectomy, abdominal aortic aneurysm repair, and pancreatectomy.

Figure 2. Odds of postoperative complication by number of quality rankings. An OR of 1 represents any hospital with 5 rankings as it served as the referent group.

Table 1

Quality Hospital Rating System Overview. Includes background and methodology information regarding each quality award included for study.

Organization (2011–2013)	Manner of Reporting	Background/Methodology	Fee for Service/ Request Designation
Truven	Top 100 Hospitals	“Truven Health 100 Top Hospitals®... incorporates independent public data, risk-adjusted and peer-reviewed methodology.” Key performance measures: mortality (in-hospital, 30 day), complications, length of stay, 30 day readmission, inpatient expense per discharge, HCAHPS score ¹³	No
Health Grades	100 Best Hospitals	“HealthGrades releases a comprehensive evaluation of hospital care using 40 million hospital records obtained from the Centers for Medicare and Medicaid Services.” Key performance measures: clinical outcomes only; risk-adjusted mortality and in hospital complications ¹⁴	No, but charge for consulting; licensing fees
Joint Commission Distinguished	Top Performer-Surgical	“The Joint Commission’s Top Performer on Key Quality Measures® program recognizes accredited hospitals that attain excellence on accountability measure performance. The tracer methodology uses information from the organization to follow the experience of care, treatment or services...through the entire care delivery process.” ¹⁵	Yes, fees associated with accreditation
Becker’s Review	Best Hospitals (2011) Great Hospitals (2012, 2013)	“Becker’s Hospital Review accepted nominations, conducted research and considered other reputable hospital ranking sources such as U.S. News & World Report, HealthGrades, Magnet Recognition by the American Nurses Credentialing Center and Malcolm Baldrige National Quality Award recipients.” ¹⁶ No discrete methodology provided.	No
US News & World Report (Orthopedics, Gastroenterology, Cancer, Cardiology)	Best Hospitals	“The analysis covers clinical specialties, procedures and medical conditions. Scores are based on data that include survival, patient safety, nurse staffing and other factors” Key performance measures: clinical outcomes (risk-adjusted mortality), patient safety, expert opinion (reputation), structural (hospital volume) measures. ¹⁷	No
Hospital Compare (Medicare)	Difference from mean in deaths from treatable complications post-op	“Hospital Compare was created through the efforts of the Centers for Medicare & Medicaid Services (CMS)... in collaboration with organizations representing consumers, hospitals, doctors, employers, accrediting organizations, and other federal agencies.” Key performance measures: seven groups of structural, process and outcome measures. Uses latent variable model to calculate scores. ¹⁸	No
Nursing Magnet	Certification	“The Magnet Recognition Program® recognizes health care organizations for quality patient care, nursing excellence and innovations in professional nursing practice.” Key performance measures: Transformational Leadership, Structural Empowerment, Exemplary Professional Practice, New Knowledge, Innovation & Improvements, Empirical Quality Results ¹⁹	Yes Appraisal fee: \$18,000–65,000 depending on hospital size
Electronic Medical Record Adoption Model (EMRAM) (2016)	Stages of EMR integration	“The EMR Adoption Model (EMRAM) is a progressively sophisticated roadmap that enables quality, safety, and operational efficiency for healthcare providers. ... incorporates methodology and algorithms to automatically score hospitals around the world relative to their Electronic Medical Records (EMR) capabilities.” ²⁰	Free score, but must seek out. Consulting fees.

Table 2

Baseline characteristics of the study population

Age (mean, SD)		Total Knee Replacement n=87,032	Total Hip Replacement n=53,866	Bariatric n=8,502	Mastectomy n=12,826	Prostatectomy n=25,263	Pulmonary Lobectomy n=19,578	Vascular AAA Repair n=795	Pancreatotomy n=1,917	Laparoscopic Cholecystectomy n=19,878
Race										
White		76.1%	82.1%	51.8%	64.3%	68.5%	74%	80.9%	72.1%	61.9%
Black		8.2%	6.4%	14.7%	12.7%	10.4%	7.5%	4.8%	9.1%	11.3%
Hispanic		8.4%	4.6%	23.6%	10.5%	10.8%	7.6%	6.7%	7.3%	19.3%
Other		7.3%	6.9%	9.9%	12.5%	10.4%	10.8%	7.7%	11.6%	7.6%
Sex										
Female		54,956 (63.1%)	30,618 (56.8%)	6,647 (78.2%)	12,637 (98.5%)	0 (0%)	10,056 (51.4%)	226 (28.4%)	960 (50.1%)	13,415 (67.5%)
Male		32,076 (36.9%)	23,248 (43.2%)	1,855 (21.8%)	189 (1.5%)	25,263 (100%)	9,522 (48.6%)	569 (71.6%)	957 (49.9%)	6,463 (32.5%)
Median Income*										
1 st quartile		18.7%	15.5%	33.7%	19.1%	19.1%	18.6%	20.4%	20.1%	24.4%
2 nd quartile		23.5%	20.9%	23.7%	19.1%	21.7%	20.9%	26.4%	21.6%	23.5%
3 rd quartile		26.6%	25.4%	24.9%	25.9%	22.9%	25.5%	26.0%	24.0%	24.8%
4 th quartile		31.2%	38.2%	17.7%	35.9%	36.2%	34.9%	27.2%	34.3%	27.3%
Comorbidities										
Hypertension		59,012 (67.8%)	31,609 (58.7%)	4,866 (57.2%)	5,064 (39.5%)	13,613 (53.9%)	10,505 (53.7%)	620 (77.9%)	1,199 (62.6%)	9,178 (46.2%)
Diabetes		17,091 (19.9%)	7,149 (13.3%)	2,875 (33.8%)	1,657 (12.9%)	3,982 (15.8%)	3,181 (16.3%)	127 (15.9%)	628 (32.8%)	3,565 (17.9%)
CHF		1,723 (1.8%)	1,154 (2.1%)	146 (1.7%)	200 (1.6%)	680 (2.7%)	789 (4.0%)	4 (0.5%)	63 (3.3%)	993 (5.0%)
Alcohol		653 (0.75%)	797 (1.5%)	32 (0.38%)	63 (0.49%)	258 (1.1%)	535 (2.7%)	38 (4.8%)	53 (2.8%)	534 (2.7%)

Table 3

(a). Overall Inpatient Complication and Death Rates for Each Surgery for Hospitals in Florida and New York. The complication rate for each surgery and hospital was computed by dividing the number of patients that have complications after receiving the surgery in a certain hospital by the total number of patients for that surgery. **(b)** Average Percentage of Complications for Each Surgery for Hospitals in Florida and New York. AAA: abdominal aortic aneurysm.

(a)				
	Surgery	Complication rates	Mortality rates	
Low Risk Surgery	Mastectomy	1.0%	0.1%	
	Prostatectomy	1.6%	0.2%	
	Bariatric Surgery	1.6%	0.1%	
	Laparoscopic Cholecystectomy	4.6%	0.6%	
	Total Hip Replacement	1.6%	0.1%	
	Total Knee Replacement	1.6%	0.1%	
High Risk Surgery	Pulmonary Lobectomy	9.7%	2.2%	
	AAA Repair	20.0%	4.8%	
	Pancreatectomy	17.8%	2.3%	

(b)				
	Surgery	Average % complications	Standard Deviation	
Low Risk Surgery	Mastectomy	1.3%	1.8%	
	Prostatectomy	2.9%	3.0%	
	Bariatric Surgery	1.9%	2.4%	
	Laparoscopic Cholecystectomy	4.8%	3.8%	
	Total Hip Replacement	2.4%	2.9%	
	Total Knee Replacement	2.2%	2.1%	
High Risk Surgery	Pulmonary Lobectomy	12.1%	7.2%	
	AAA Repair	22.2%	11.8%	

Table 4

Number of Hospitals with Quality Designations

Number of quality designations	Number of hospitals (%): 2012	Number of hospitals (%): 2013
0	77 (43.5%)	39 (22%)
1	77 (43.5%)	105 (59.3%)
2	12 (6.8%)	24 (13.6%)
3	5 (2.8%)	3 (1.7%)
4	2 (1.1%)	4 (2.3%)
5	4 (2.3%)	2 (1.1%)
Total	177	177

Table 5

Variable Importance for Predictors of Overall Mortality and Surgical Complications from Decision Tree. **(a)** Models used and Performance Metrics for Predictors of Overall Mortality and Surgical Complications **(b)** Variable importance (Gini) measures a variable's ability to perform in a specific tree of a specific size either as a primary splitter or as a surrogate splitter. The result does not contain units, but is scaled 1 to 100 with the greater number indicating greater importance. The sign indicates positive or negative correlation.

(a)				
Outcome	Accuracy	Sensitivity	Specificity	P value
MORTALITY				
Support Vector Machine	78%	33%	100%	<.001
Decision Tree	83%	78%	92%	<.001
COMPLICATIONS				
Support Vector Machine	70%	24%	97%	0.224
Decision Tree	93%	100%	90%	<.001

(b)	
Hospital Characteristic	Variable importance
MORTALITY	
Hospital Compare	42.435239
Total Procedure Volume	16.107084
Joint Commission	11.986891
Magnet	-11.754712
Becker's	9.738801
COMPLICATIONS	
Hospital Compare	54.832182
Becker's	11.036830
Total Procedure Volume	10.597257
Joint Commission	8.456566
Magnet	-5.312190

Table 6

Procedure Specific Decision Tree Analysis for Importance of Hospital Ranking in Predicting Complication Rate. The top 3 most important variables for predicting complications for each procedure are listed with their respective variable importance (%). High risk procedures are in bold. “Ranking” indicates if hospital ranking was a variable of importance (outside of top 3) and its respective importance. Accuracy of model prediction. LYTES: electrolyte imbalance; WGHTL: weight loss; COAG: coagulopathy; PULM: pulmonary disorders; ANEM: anemia; CHF: congestive heart failure; CHRNL: chronic lung disease; RENLF: renal failure

Surgery Name	1	Top 3 Important Features			Ranking	Accuracy
		%	2	%	3	%
Pulmonary Lobectomy	LYTES	100	WGHTL	94.9	COAG	32.8
					No	68.3%
AAA Repair	LYTES	100	WGHTL	76.2	AGE	49.7
					No	64.2%
Pancreatectomy	LYTES	100	COAG	71.6	WGHTL	57.4
					No	64.3%
Laparoscopic cholecystectomy	LYTES	100	AGE	80.6	CHF	71.5
Total Knee Replacement	LYTES	100	PULM	76.0	CHRNL	35.5
Total Hip Replacement	CHF	100	LYTES	98.6	RANKING	41.9
Mastectomy	LYTES	100	AGE	70.2	RANKING	17.1
Bariatric Surgery (gastric bypass)	LYTES	100	ANEM	24.2	OBESE	11.0
Prostatectomy	LYTES	100	RENLF	48.0	WGHTL	40.6
					No	74.7%

Table 7

Models used and Performance Metrics for Predictors of Surgical Complications in Inpatient and Volume-Outcome Procedures

	Model	Accuracy	Sensitivity	Specificity
Inpatient Procedures (prostatectomy, open AAA repair, and pulmonary lobectomy)	Logistic Regression	79%	68%	80%
	Decision Tree	71%	74%	71%
Volume-Outcome Relationship (bariatric surgery, pancreatectomy, total hip and knee replacements)	Logistic Regression	80%	59%	80%
	Decision Tree	86%	51%	88%