Confusion Instead of Clarity: Publicly Reported Cardiac Surgery Ratings for Coronary Artery Bypass Grafting and Aortic Valve Replacement

Anjali C Raghuram, BA, Tejasvi K Dasari, BA, Brendan Chou, BS, Sujana Balla, MBBS, Sergio M Navarro, BS, Rohan M Shah, MPH, Ankur Bakshi, MD, Matthew J Wall, MD, FACS, Todd K Rosengart, MD, FACS, Ravi K Ghanta, MD, FACS

BACKGROUND: Public reporting of cardiac surgery ratings has been advocated to inform patient selection of hospitals. Although Society of Thoracic Surgeons (STS) ratings are based on audited risk-adjusted patient outcomes, other rating systems rely on administrative databases. In this study, we evaluate correlation among 4 widely used hospital rating systems for coronary artery bypass grafting (CABG) and aortic valve replacement (AVR).

STUDY DESIGN: We identified an initial cohort of 602 hospitals from *US News & World Report*'s (USN) listing of the 2016-2017 "Best Hospitals for Cardiology & Heart Surgery." From this cohort, current publicly available CABG and AVR ratings were collected from the STS, USN, Centers for Medicare & Medicaid Services, and Healthgrades. All 4 rating systems rated hospitals as high, average, or below average performers for each procedure. We then determined the match rate between rating systems for individual hospitals and assessed interrater reliability with Cohen's κ.

RESULTS: Rating systems had different distributions of high and low performing ratings assigned. USN rated hospitals as high performing for both CABG and AVR more frequently compared with STS, Healthgrades, and Centers for Medicare & Medicaid Services. For CABG, the match rate between systems varied from 50% to 85%, with the best match between STS and Centers for Medicare & Medicaid Services. Similarly for AVR, the match rate varied from 50% to 73%, with the best match between STS and Healthgrades. Interrater reliability was poor among the 4 rating systems ($\kappa < 0.2$) and consistent with no agreement for CABG and AVR ratings.

CONCLUSIONS: Publicly reported cardiac surgery ratings have significant discrepancy and poor correlation. This might confuse instead of clarify public perception of hospital quality for cardiac surgery. (J Am Coll Surg 2019;228:180−187. © 2018 by the American College of Surgeons. Published by Elsevier Inc. All rights reserved.)

Hospital ratings are used increasingly to influence patients' hospital selections. Consumer demand for these ratings is clear. Medical queries and reviews are the fastest-growing Internet search category, and the ease of accessibility of rating systems is unmistakable.[1] For cardiac surgery, healthcare consumers are able to access procedure-specific ratings from private companies, surgical societies, and governmental organizations. In some states, such as New York and California, public reporting of cardiac surgery outcomes has been mandated. In most

Disclosure Information: Nothing to disclose.

Anjali Raghuram and Teja Dasari contributed equally to this work.

Presented at the American College of Surgeons 104th Annual Clinical Congress, Scientific Forum, Boston, MA, October 2018.

Received July 1, 2018; Revised July 24, 2018; Accepted July 25, 2018.

From the Michael E DeBakey Department of Surgery, Baylor College of Medicine, Houston, TX.

Correspondence address: Ravi K Ghanta, MD, FACS, Michael E DeBakey Department of Surgery, Baylor College of Medicine, One Baylor Plaza, MC-390, Houston, TX 77030. email: ravi.ghanta@bcm.edu

© 2018 by the American College of Surgeons. Published by Elsevier Inc. All rights reserved.

https://doi.org/10.1016/j.jamcollsurg.2018.07.663
ISSN 1072-7515/18

Abbreviations and Acronyms

AVR = aortic valve replacement
CABG = coronary artery bypass grafting
CMS = Centers for Medicare & Medicaid Services
HG = Healthgrades
STS = Society of Thoracic Surgeons
USN = *US News & World Report*

states, however, consumers can only obtain a qualitative rating, typically in the form of a "star" rating. Although highly publicized and used, these rating systems differ in data sources and methodologic approaches.[2]

Private media rating organizations, such as *US News & World Report* (USN) and Healthgrades (HG) receive considerable publicity, but the optimal criteria for identification of high-quality hospitals are unclear.[3,4] These ratings are derived from administrative databases, surveys, and proprietary algorithms. The Society of Thoracic Surgeons (STS) adult cardiac surgery database has grown to include more than 1,119 cardiac surgery programs and is estimated to contain information on >95% of the adult cardiac operations performed annually in the US.[5] The database is audited for completeness and accuracy and not derived from administrative data. From the comprehensive clinical data collected, STS has developed well-validated risk algorithms for assessing coronary artery bypass grafting (CABG) and aortic valve replacement (AVR) performance. Since 2010, STS has been publicly reporting hospital ratings derived from its risk-adjusted data, providing star ratings for all participating hospitals. Similarly in 2005, Centers for Medicare & Medicaid Services (CMS), through the Hospital Compare initiative, began offering the American public a nationally standardized and validated measure of hospital quality.[6]

It is clear that these rating systems offer data that influence public perception of hospitals' cardiac surgery performance. However, it is unclear whether these rating systems can appropriately and consistently identify high-quality hospitals for complex operations, such as CABG or AVR. To our knowledge, no study has identified whether publicly reporting cardiac surgery rating systems are consistent. The objective of this study was to determine the correlation among and within 4 different rating systems—STS, USN, HG, and CMS—for hospital CABG and AVR ratings. We hypothesized that ratings systems correlate poorly and, as such, could limit reliability and interpretability by the healthcare consumer.

METHODS

Hospital selection

To evaluate hospital ratings from the perspective of a healthcare consumer, we used the highly publicized list of the USN's "Best Hospitals for Cardiology & Heart Surgery" (n = 602) to establish an initial hospital cohort of likely high-volume cardiac surgery hospitals. From this cohort, we obtained individual hospital cardiac surgery ratings for CABG and AVR in May 2018 from 4 publicly reporting rating systems: USN, STS, HG, and CMS. Centers for Medicare & Medicaid Services does not report AVR or valve operation-specific outcomes, so AVR ratings data were obtained from the other 3 rating systems. The CABG and AVR procedure ratings assigned to the hospitals were compared among the different rating systems to assess for agreement.

Society of Thoracic Surgeons

Society of Thoracic Surgeons reports a number of adult cardiac surgery measures. Among these, STS includes composite scores for CABG and AVR. Based on a combination of 11 National Qualify Forum-endorsed CABG process and outcomes measures, STS generates CABG composite scores. These measures include perioperative medication use, the operative care process, risk-adjusted operative mortality, and risk-adjusted morbidity.[7] Of note, the operative care process measure describes use of an internal mammary artery in CABG. The STS AVR composite score reflects a similar methodological approach, with the exception that it does not include the operative care process and perioperative medication use measures.[8] The STS accounts for these measure exclusions with AVR by describing that there is no comparative measure for use of the internal mammary artery in valve operations, and perioperative medication use is not well-defined for AVR compared with CABG.[8] Hospitals participating in the STS national database receive scores in these individual measures and subsequently overall CABG and AVR composite scores. In addition to these numeric scores, participants receive assignment to 1 of 3 star-rating categories for these procedures, as determined by their higher than average (3-star), average (2-star), or lower than average (1-star) performances.[9] The most current CABG (n = 354) and AVR (n = 333) star ratings were obtained for all publicly reported hospitals in the STS database that were also present in the initial cohort.

USN

USN's 2016-2017 "Best Hospitals for Cardiology & Heart Surgery" rankings were obtained directly from its

website.[10] The USN company offers procedure-specific ratings to evaluate hospitals as performing within 1 of 3 rating categories: high-performing (5-star), average (3-star), or below average (1-star).[11] The most current USN CABG (n = 602) and AVR (n = 593) ratings for all publicly reported hospitals that were also present in our initial cohort were obtained.

Healthgrades

Healthgrades assigns ratings based on performance data extracted from 3 years of recent MedPAR (Medicare Provider Analysis and Review) data.[12] The company uses a proprietary formula to predict mortality rates during hospitalization, as well at 30 days after discharge for CABG and AVR. The company then assigns star ratings for each of the hospitals to reflect better than expected (5-star), similar (3-star), or poorer than expected (1-star) predicted mortality rates compared with the observed rate. The most current HG CABG (n = 543) and valve operation (n = 543) star ratings were obtained for all publicly reported hospitals that were also present in our initial cohort.

Centers for Medicare & Medicaid Services

The CMS's proprietary Hospital Quality initiative reflects quality measures agreed on by the National Quality Form, AHRQ, and Joint Commission. These measures include timely and effective care, complications and deaths, readmissions, use of medical imaging, and survey of patient hospital experiences. Ratings for CABG are derived from final actions claims data for Medicare fee-for-service beneficiaries, the Medicare Enrollment Database, the New York Cardiac Surgery Reporting System, and California Patient Discharge Data. The CMS reports outcomes and procedure-specific mortality rates as a comparison between predicted and observed rates during a defined time period. Hospital performance is reported as "better than the national rate," "no different than the national rate," or "worse than the national rate."[13] Current CMS CABG (n = 562) mortality ratings were obtained for all publicly reported hospitals that were also present in our initial cohort.

Methodology

Each of the rating systems assigns hospitals 1 of 3 possible ratings, on either a 3-2-1 scale or a 5-3-1 scale. Centers for Medicare & Medicaid Services is an exception in that it reports CABG mortality rates with qualitative descriptors in comparison with the national mortality rate. To create a consistent framework for comparing CABG ratings across the 4 rating systems, 5-3-1 ratings were transformed into 3-2-1 ratings. Centers for Medicare & Medicaid Services' 3 qualitative descriptors were then assigned numeric values of 3 (better than the national rate), 2 (no different than the national rate), or 1 (worse than the national rate) rating. The distribution of high-performing, average-performing, and low-performing ratings in each rating system was then compared. The number of high-performing hospitals for CABG and AVR within each state was determined as well.

To compare individual hospital ratings for CABG and AVR between rating systems, we used 2 methods, both featuring pairwise comparisons between rating systems. First, 2 rating systems were compared and if a given hospital had the same rating in both rating systems, this was categorized as a ratings match. Otherwise, this hospital's ratings were categorized as unmatched. We determined the match rate among hospital ratings in our cohort. Second, hospital ratings were compared between rating systems using Cohen's linearly weighted κ coefficients to assess interrater reliability. To assess the correlation between hospital CABG and AVR ratings within individual rating systems, Spearman's ρ correlation coefficients were calculated.[14] Centers for Medicare & Medicaid Services CABG ratings were excluded from this analysis because there are no comparative CMS AVR or valve operation ratings.

Cohen's κ is a robust statistic that is similar to a correlation coefficient in that it ranges in value from -1 to $+1$, with a value of 0 representing agreement that can be expected from random chance and a value of 1 representing perfect agreement between 2 raters.[15] We used linear weights to account for the greater significance attributable to a ratings discrepancy of 3 vs 1 between 2 systems when compared with a less significant ratings discrepancy of 2 vs 1 between 2 systems. Cohen's κ also accounts for chance agreement. For example, if there is a match of ratings for a given hospital in 2 different rating systems, Cohen's κ designates a linear weight of 0 to consider that these rating systems are using some common knowledge base about hospital cardiac surgery performance that does not reflect the proprietary methodologies proposed by each rating system.[16] Cohen's κ additionally considers the likelihood that a high prevalence of hospitals with an average rating (rating = 2) reflects a component of chance and therefore inaccurate identification of hospitals that might be high, medium, or low CABG and AVR performers.

To allow for direct interpretation of congruence accuracy among the rating systems, we squared the values of the Cohen's κ coefficients to obtain coefficients of determination. We also obtained SEs and 95% CIs for the κ coefficients. Hospitals that were not included in both

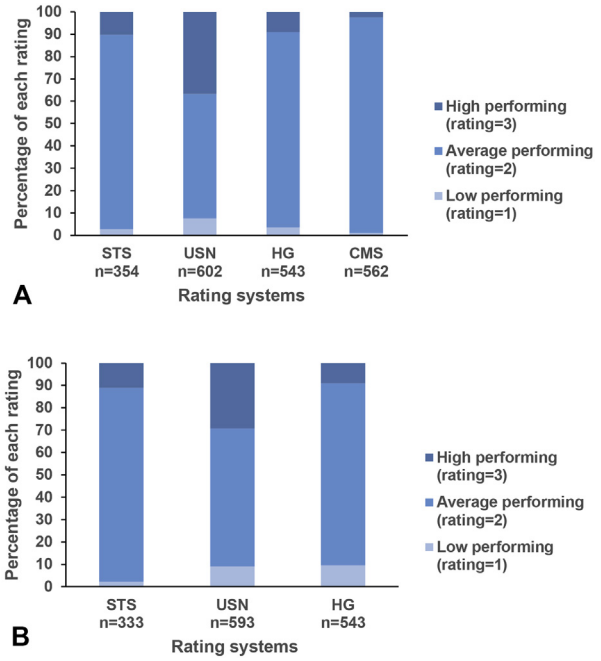

Figure 1. Percent distribution of ratings assigned by each rating system: (A) coronary artery bypass grafting ratings and (B) aortic valve replacement ratings. n = the number of hospitals within our cohort rated by each system. *US News & World Report* (USN) had the highest proportion of hospitals rated as high performing in comparison with the other rating systems. CMS, Centers for Medicare & Medicaid Services; HG, Healthgrades; STS, Society of Thoracic Surgeons.

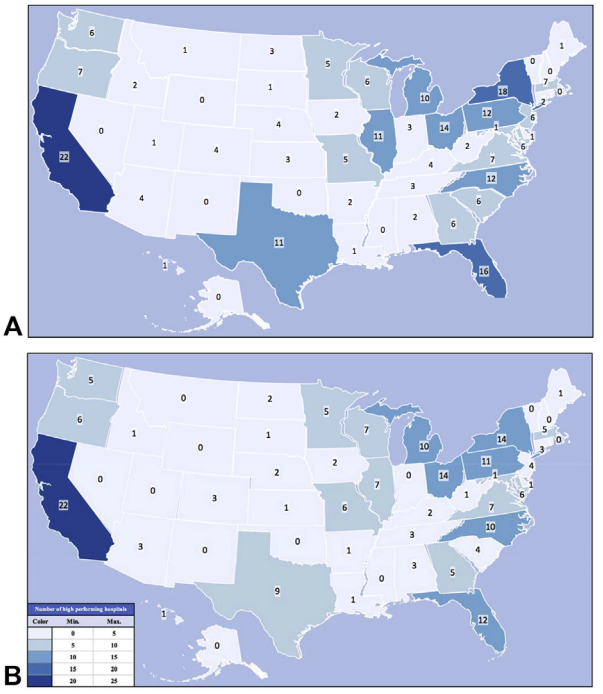

Figure 2. A US map of high-performing (A) coronary artery bypass grafting and (B) aortic valve replacement hospitals, across all applicable rating systems (Society of Thoracic Surgeons, *US News & World Report*, Healthgrades, Centers for Medicare & Medicaid Services), within each state. The cohort of hospitals evaluated in this study were distributed across all regions of the US.

rating systems in a given pairwise comparison were excluded from this analysis.

RESULTS

Ratings distribution

Figure 1 illustrates the distribution of CABG and AVR ratings obtained across the different rating systems. For CABG, USN ratings skewed toward high performing when compared with the other rating systems, with 36% (219 of 602) rated as high performing and 8% (51 of 602) rated as low performing. Centers for Medicare & Medicaid Services rated the least number of hospitals as high performers, with 3% (14 of 562) rated as high performing and 96% (542 of 562) rated as average. Similarly, among the 3 AVR rating systems, USN ratings skewed more toward high performing. In USN, more hospitals were rated as high performing (29%; 173 of 593) compared with HG (9%; 48 of 543) and STS (11%; 37 of 333). Figure 2 presents the number of hospitals in each state rated as high performing, with respect to CABG and AVR ratings, in at least 1 of the 4 rating systems. Hospitals in this cohort were present in all regions of the US. Hospitals with high CABG or AVR ratings were found primarily in California (22 CABG, 22 AVR hospitals), Florida (16 CABG, 12 AVR hospitals), and New York (18 CABG, 14 AVR hospitals).

Match rates between USN, Healthgrades, Society of Thoracic Surgeons, and Centers for Medicare & Medicaid Services

Only 2 hospitals in the initial cohort (n = 602) were rated as high performers by all 4 rating systems for CABG. Likewise, only 5 hospitals were rated as high performers by all 3 AVR rating systems. Figure 3 illustrates the match rate of hospital ratings between each rating system for CABG and AVR. For CABG, STS and CMS had the highest rating agreement at 85%, followed by CMS and HG with an 81% match rate. The poorest CABG ratings matches were seen with USN and STS (50%), USN and HG (52%), and USN and CMS (58%). For AVR, STS ratings were the most consistent with HG ratings (73%) and, as seen with CABG ratings, STS AVR ratings were poorly consistent with USN ratings (60%).

Among all hospitals with CABG ratings across STS, USN, HG, and CMS, 16 (3%) hospitals were simultaneously rated as high performing by at least 1 rating

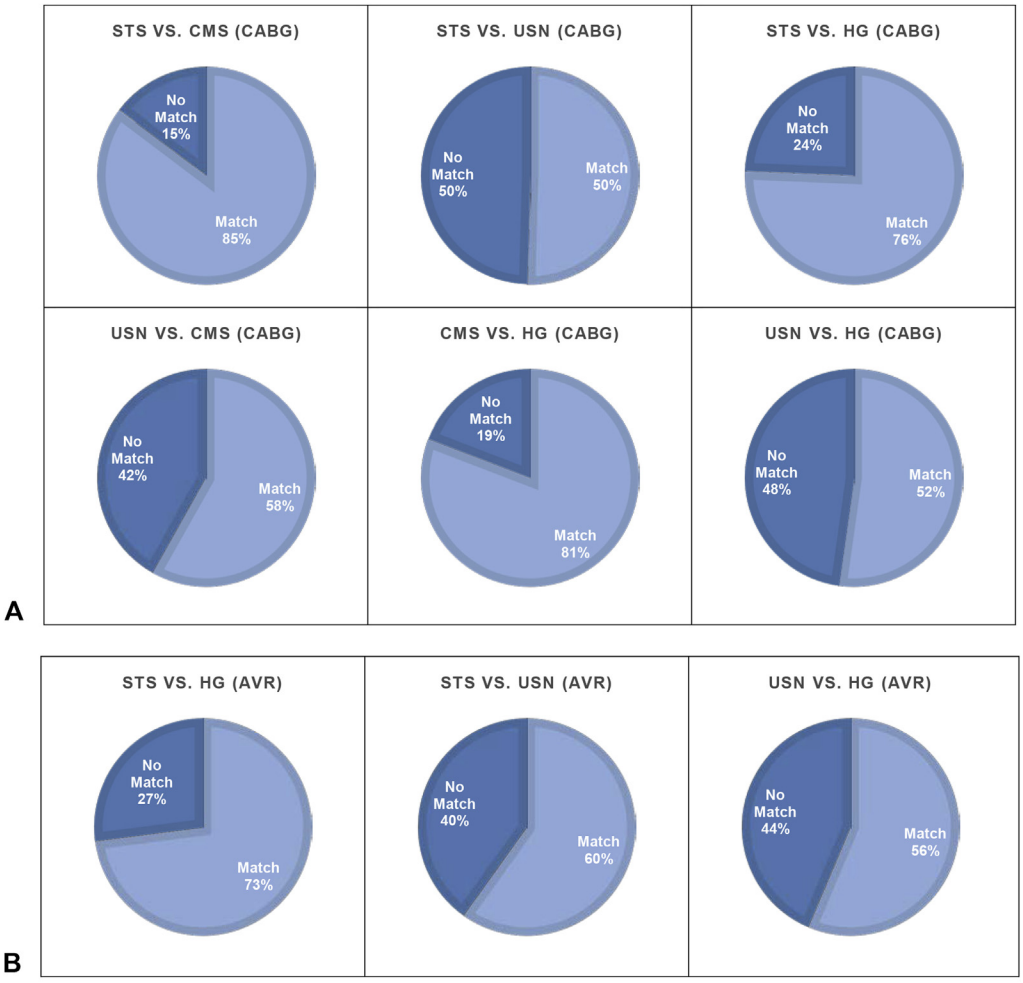

Figure 3. Consistency in assigned ratings as captured by percent agreement of match vs no match in ratings: (A) coronary artery bypass grafting (CABG) ratings and (B) aortic valve replacement (AVR) ratings. Highest match rates were found between Society of Thoracic Surgeons (STS) and Centers for Medicare & Medicaid Services (CMS) for CABG and STS and Healthgrades (HG) for AVR. USN, *US News & World Report*.

system and low performing by at least 1 other rating system. No hospitals were rated as low performing for CABG by all rating systems. Similarly, for all hospitals with AVR ratings across the systems, 11 (2%) hospitals were simultaneously rated as high performing by at least 1 rating system and low performing by at least 1 other rating system. As with CABG, no hospitals were rated as low performing for AVR by all rating systems.

Table 1. Weighted Cohen's κ Coefficients Between Rating Systems for Coronary Artery Bypass Grafting Ratings

Rating system comparison	Hospitals rated by both rating systems, n	Value of observed κ	Coefficient of determination (%)	SE	95% CI	Interpretation of agreement between rating systems (κ)
STS vs CMS	340	0.126	0.016 (1.6)	0.064	0.001–0.251	No agreement (<0.2)
STS vs USN	349	0.113	0.013 (1.3)	0.027	0.060–0.166	No agreement (<0.2)
STS vs HG	337	0.121	0.015 (1.5)	0.054	0.015–0.227	No agreement (<0.2)
USN vs CMS	557	0.089	0.008 (0.8)	0.022	0.046–0.132	No agreement (<0.2)
CMS vs HG	506	0.099	0.010 (1.0)	0.042	0.017–0.181	No agreement (<0.2)
USN vs HG	540	0.108	0.012 (1.2)	0.028	0.054–0.161	No agreement (<0.2)

CMS, Centers for Medicare & Medicaid Services; HG, Healthgrades; STS, Society of Thoracic Surgeons; USN, *US News & World Report*.

Table 2. Weighted Cohen's κ Coefficients Between Rating Systems for Aortic Valve Replacement Ratings

Rating system comparison	Hospitals rated by both rating systems, n	Value of observed κ	Coefficient of determination (%)	SE	95% CI	Interpretation of agreement between rating systems (κ)
STS vs HG	318	0.056	0.003 (0.3)	0.048	0−0.151	No agreement (<0.2)
STS vs USN	329	0.182	0.033 (3.3)	0.037	0.110−0.254	No agreement (<0.2)
USN vs HG	534	0.096	0.009 (0.9)	0.030	0.037−0.156	No agreement (<0.2)

HG, Healthgrades; STS, Society of Thoracic Surgeons; USN, *US News & World Report*.

Correlation between *US News & World Report*, Healthgrades, Society of Thoracic Surgeons, and Centers for Medicare & Medicaid Services

Tables 1 and 2 present the Cohen's κ coefficients for the rating systems. For CABG, all of the rating systems demonstrated poor interrater reliability. The STS demonstrated the highest interrater reliability with CMS, albeit yielding a low κ of 0.126 (95% CI 0.001 to 0.251). Because all of the κ coefficients seen in pairwise rating comparisons were low, there is poor agreement evidenced between the rating systems. Consistent with this pattern of poor interrater reliability, the highest value of Cohen's κ for AVR ratings was 0.182 (95% CI 0.110 to 0.251), seen with a comparison between STS and USN ratings. Additionally, coefficients of determination capturing the percentage of data that can be interpreted as reliable between 2 rating systems were very low for all of the CABG and AVR rating system pairwise comparisons.

Correlation between coronary artery bypass grafting and aortic valve replacement ratings

To evaluate whether there was a correlation between high ratings in CABG and high ratings in AVR, ratings for individual hospitals for both procedures within a single rating system were compared. Table 3 identifies the Spearman's ρ correlation coefficients between hospital CABG and AVR ratings within each rating system. *US News & World Report* and HG CABG and AVR ratings were strongly correlated (Spearman's $\rho = 0.80$, $p < 0.01$; Spearman's $\rho = 0.97$, $p < 0.01$, respectively). However, STS CABG and AVR ratings (Spearman's $\rho = 0.32$, $p < 0.01$) were weakly correlated.

DISCUSSION

This study demonstrates that publicly reported hospital ratings for CABG and AVR are inconsistent, and there is poor interrater reliability between hospital rating systems. Only 2 hospitals for CABG and 5 hospitals for AVR were rated as high performing in all of the rating systems evaluated. The distribution of ratings, specifically the number of high-performing hospitals, differed between the ratings agencies. For CABG, STS and CMS had the highest ratings agreement, with an 85% match rate. For AVR, STS ratings were the most consistent with HG ratings (73%). Interrater reliability between the 4 rating systems was evaluated to be poor with the Cohen's κ statistic. The relative inconsistencies in the distributions of ratings assigned indicate that consumers inevitably encounter some bias when visiting different hospital rating websites. Notably, there is poor correlation of private media organizations with the STS adult cardiac surgery database, which, due to its extensive capture of cardiac surgery procedures, has gained recognition as one of the most comprehensive and well-recognized clinical data registries in healthcare.[5] Other ratings organizations, however, also capture data that are not incorporated by STS, which can also be important in gauging quality.

In healthcare research, interrater reliability enables the creation of recommendations for clinical practice.[15] The low Cohen's κ coefficients between rating systems in this study indicate that the hospital ratings data presented do not correlate and, as such, cannot be reliably interpreted by a healthcare consumer. CMS, USN, and HG rely on administrative data from the Medicare patient population, which might not be representative of outcomes in all cardiac surgery populations. Therefore, one significant source

Table 3. Spearman's ρ Coefficients Between Coronary Artery Bypass Grafting and Aortic Valve Replacement Ratings Within Rating Systems

Rating system procedure comparison	Spearman's ρ	p Value	Interpretation of correlation between procedure ratings
STS CABG vs STS AVR	0.32	<0.01	Weak correlation
USN CABG vs USN AVR	0.80	<0.01	High positive correlation
HG CABG vs HG AVR	0.97	<0.01	High positive correlation

HG, Healthgrades; STS, Society of Thoracic Surgeons; USN, *US News & World Report*.

of variability in the ratings of these systems is the underlying patient data set. In addition, within an administrative database, coding errors might be present and procedure-specific risk adjustment is not well established.[17] Second, the time periods of evaluation differ for each rating system, which can impact the mortality outcomes. Finally, private media organizations use proprietary algorithms that differ substantially from one another.

Quality improvement databases initiated by surgical societies, such as the STS and the American College of Surgeons NSQIP databases, do not rely on administrative data and are audited for data inaccuracy. Although USN might be the most well-known rating system to the health consumer, STS is frequently considered the gold standard of cardiac surgery reporting. *US News & World Report* and HG are well-known and highly publicized rating systems. Given the disagreement among USN, HG, and STS, surgeon-driven audited quality databases, such as the STS and NSQIP databases, should receive greater emphasis.

Importantly, this study adds to the findings in recent reports that question the validity of rating systems in capturing hospital performance accurately. Osborne and colleagues[18] investigated whether popular media and Internet-based hospital quality rankings can identify hospitals at which patients will attain improved outcomes after cardiovascular surgery procedures. They found that popular hospital rating systems can somewhat predict lower risk-adjusted mortality, but fail to consider factors such as hospital volume, which contribute to high-quality patient outcomes at hospitals without high publicly reported ratings. Expanding the scope of discussion beyond cardiac surgery, Lascano and associates[19] evaluated whether there was a correlation between publicized health ranking systems and surgical outcomes after radical cystectomy in the state of New York. Although this study was limited to one state, it drew conclusions based on performance by hospitals with the highest volume of radical cystectomy within New York and found that rating systems with higher emphasis on prestige do not enable patients to make the best medical care decisions when compared with rating systems that use objective surgical outcomes. Chau and coworkers[20] reiterate the point that, although imperfect, volume might be a good quality indicator of surgical outcomes, specifically with respect to pancreatectomy. However, they affirm that the current offering of publicly available hospital rating systems fail to show consistent agreement with one another and also fail to capture true outcomes after operation.

It is reasonable to expect that hospitals that were rated as high performing in one cardiac surgery procedure would also be rated as high performing in another cardiac surgery procedure. Elements contributing to surgical high performance include personnel, infrastructure, and evidence-based processes of care designed to limit morbidity and mortality.[21] As such, hospitals that have successfully taken steps to improve the practice of CABG stand to mirror this improvement in the practice of AVR and vice versa. Johnston and colleagues[22] evaluated the relationship between coronary and valve procedure outcomes within 18 institutions in Virginia but found that mortality rates in CABG did not correlate with mortality rates in valve operations. Similarly, we found that STS CABG and AVR ratings were only weakly correlated. We did identify a correlation between USN and HG CABG and AVR ratings. This might be related to these organizations proprietary methodology, which incorporates different data sets.

There are some important limitations of this study. Our initial cohort was chosen from the USN's 602 ranked "Best Hospitals in Cardiology & Heart Surgery" and not the entire national cohort of hospitals. The distribution of ratings for the overall national cohort will be different. We chose to focus our analysis on this listing of hospitals that are likely high-volume cardiac surgery centers and those that consumers might interpret as the best. Second, CABG and AVR ratings are only obtained as a result of hospitals choosing to publicly report these data, and hospitals that do not disclose their cardiac surgery performance cannot be assessed by these rating systems. Some hospitals participated in some rating systems but not in all of them, which reduced our data set for analysis in pairwise comparisons. Our data set, however, remained sizeable in all of our analyses and compared with earlier studies of hospital ratings.[3,4] The data included in this study reflect findings specific to the US and therefore might not be generalizable to non-US rating systems. In addition, each rating system assesses hospitals in specific regions or states that are receiving patient volumes large enough to merit rating consideration by that particular system. It is equally likely that many hospitals that might be rated in one rating system and "unrated" in another could have comparable or significantly different cardiac surgery performance. With this study's statistical approach, the percent agreement statistic is both easily calculated and interpretable but can overestimate the true agreement among raters. Conversely, Cohen's κ can excessively underestimate agreement between 2 rating systems. However, with the goal of capturing interrater reliability with the most precision, both percent agreement and Cohen's κ complement one another in qualifying the data.

CONCLUSIONS

This study demonstrates a lack of agreement and inter-rater reliability among rating systems, contributing to confusion when gauging hospital cardiac surgery performance. Interestingly, few hospitals are rated as high performing in all rating systems. Patients might benefit from publicly reported quality metrics from risk-adjusted and audited databases to provide consistent ratings of hospital outcomes for complex surgical procedures.

Author Contributions

Study conception and design: Raghuram, Dasari, Bakshi, Wall, Rosengart, Ghanta
Acquisition of data: Raghuram, Dasari, Chou
Analysis and interpretation of data: Raghuram, Dasari, Chou, Balla, Navarro, Shah, Bakshi, Ghanta
Drafting of manuscript: Raghuram, Dasari, Balla, Ghanta
Critical revision: Raghuram, Balla, Wall, Ghanta

REFERENCES

1. Pew Research Center's Internet & American Life Project. Health Topics. Available at: http://pewinternet.org/reports/2011/healthtopics.aspx. Accessed June 17, 2018.
2. Shahian D, Edwards F, Jacobs J, et al. Cardiac surgery performance reporting part 1—history, rationale, consequences. Ann Thorac Surg 2011;92:2–11.
3. Dimick J, Staiger D, Osborne N, et al. Composite measures for rating hospital quality with major surgery. Health Serv Res 2012;47:1861–1879.
4. Osborne N, Ghaferi A, Nicholas L, et al. Evaluating popular media and Internet-based hospital quality ratings for cancer surgery. Arch Surg 2011;146:600–604.
5. D'Agostino R, Jacobs J, Badhwar V, et al. The Society of Thoracic Surgeons Adult Cardiac Surgery database: 2018 update on outcomes and quality. Ann Thorac Surg 2018;105:15–23.
6. Hospital Compare-Hospital Quality Initiative. Centers for Medicare & Medicaid Services. Available at: https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalcompare.html. Accessed June 17, 2018.
7. Performance Measure Descriptions-Coronary Artery Bypass Graft (CABG) Composite Score. The Society of Thoracic Surgeons. Available at: https://www.sts.org/quality-safety/performance-measures/descriptions#CABGCompositeScore. Accessed June 17, 2018.
8. Society of Thoracic Surgeons. STS AVR Composite Score. Available at: https://publicreporting.sts.org/avr-composite-score. Accessed June 25, 2018.
9. Society of Thoracic Surgeons. STS CABG Composite Score. Available at: https://publicreporting.sts.org/cabg-composite-score. Accessed June 29, 2018.
10. *U.S. News & World Report*. Best Hospitals for Cardiology & Heart Surgery. Available at: https://health.usnews.com/best-hospitals/rankings/cardiology-and-heart-surgery. Accessed June 17, 2018.
11. *U.S. News & World Report*. 2017-2018 Best Hospitals Procedures & Conditions Ratings. Available at: https://health.usnews.com/health-care/best-hospitals/articles/faq-how-and-why-we-rank-and-rate-hospitals. Accessed June 17, 2018.
12. Healthgrades. Mortality and Complications Outcomes 2018 Methodology. Available at: https://www.healthgrades.com/quality/2018-methodology-mortality-and-complications-outcomes. Accessed June 17, 2018.
13. Centers for Medicare & Medicaid Services. Hospital Compare. Available at: https://www.medicare.gov/hospitalcompare/search.html?. Accessed June 17, 2018.
14. Mukaka M. A guide to appropriate use of correlation coefficient in medical research. Malawi Med J 2012;24:69–71.
15. McHugh M. Interrater reliability: the kappa statistic. Biochem Med 2012;22:276–282.
16. Tang W, Hu J, Zhang H, et al. Kappa coefficient: a popular measure of rater agreement. Shanghai Arch Psychiatry 2015;27:62–67.
17. Johnson E, Nelson C. Utility and pitfalls in the use of administrative databases for outcomes assessment. J Urol 2014;190:17–18.
18. Osborne N, Nicholas L, Ghaferi A, et al. Do popular media and Internet-based hospital quality ratings identify hospitals with better cardiovascular surgery outcomes? J Am Coll Surg 2010;210:87–92.
19. 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;86:1104–1112.
20. Chau Z, West J, Zhou Z, et al. Rankings versus reality in pancreatic cancer surgery: a real-world comparison. HPB (Oxford) 2014;16:528–533.
21. Goodney P, O'Connor G, Wennberg D, et al. Do hospitals with low mortality rates in coronary artery bypass also perform well in valve replacement? Ann Thorac Surg 2003;76:1131–1136.
22. Johnston L, Downs E, Hawkins R, et al. Good at one or good at all? Variability of coronary and valve operation outcomes within centers. Ann Thorac Surg 2018;105:1678–1683.