Journal of the American Heart Association

ORIGINAL RESEARCH

Center Variability in Medicare Claims–Based Publicly Reported Transcatheter Aortic Valve Replacement Outcome Measures

Michael P. Thompson, PhD; Hechuan Hou, MS; Alexander A. Brescia, MD, MSc; Francis D. Pagani, MD, PhD; Devraj Sukul, MD, MS; Jeffrey S. McCullough, PhD; Donald S. Likosky, PhD

BACKGROUND: Public reporting of transcatheter aortic valve replacement (TAVR) claims–based outcome measures is used to identify high- and low-performing centers. Whether claims-based TAVR outcomes can reliably be used for center-level comparisons is unknown. In this study, we sought to evaluate center variability in claims-based TAVR outcomes used in public reporting.

METHODS AND RESULTS: The study sample included 119 554 Medicare beneficiaries undergoing TAVR between January 2014 and October 2018 based on procedure codes in 100% Medicare inpatient claims. Multivariable hierarchical logistic regression was used to estimate center-specific adjusted rates and reliability (R) of 30-day mortality, discharge not to home/self-care, 30-day stroke, and 30-day readmission. Reliability was defined as the ratio of between-hospital variation to the sum of the between- and within-hospital variation. The median (interquartile range [IQR]) center-level adjusted outcome rates were 3.1% (2.9%–3.4%) for 30-day mortality, 41.4% (31.3%–53.4%) for discharge not to home, 2.5% (2.3%–2.7%) for 30-day stroke, and 14.9% (14.4%–15.5%) for 30-day readmission. Median reliability was highest for the discharge not to home measure (R=0.95; IQR, 0.94–0.97), followed by the 30-day stroke (R=0.92; IQR, 0.87–0.94), 30-day mortality (R=0.86; IQR, 0.81–0.91), and 30-day readmission measures (R=0.42; IQR, 0.35–0.51). Across outcomes, there was an inverse relationship between center volume and measure reliability.

CONCLUSIONS: Claims-based TAVR outcome measures for mortality, discharge not to home, and stroke were reliable measures for center-level comparisons, but readmission measures were unreliable. Stakeholders should consider these findings when evaluating claims-based measures to compare center-level TAVR performance.

Key Words: hospital profiling ■ outcomes ■ transcatheter aortic valve replacement

Transcatheter aortic valve replacement (TAVR) is a minimally invasive therapy for patients suffering from symptomatic severe aortic stenosis, the most common heart valve condition in adults.[1] Since its approval in 2012, clinical trials have shown TAVR to be similarly safe and effective compared with surgery across all surgical risk categories, which has driven rapid growth in its use.[2–5] To facilitate continued expansion of TAVR across the country, the Centers for Medicare and Medicaid Services adopted less-restrictive reimbursement requirements for hospitals providing or seeking to provide TAVR.[6] At each step, providers, payers, and policymakers have sought to ensure that quality of care is maintained as TAVR expands to new patients and hospitals, often through tracking and benchmarking hospital performance using risk-adjusted outcomes.[7,8] Recently, *US News & World Report* began publicly rating hospital performance according to TAVR outcomes in 2020.[9]

Correspondence to: Michael P. Thompson, PhD, Section for Health Services Research and Quality, Department of Cardiac Surgery, Michigan Medicine, 5331K Frankel Cardiovascular Center, 1500 E. Medical Center Drive, SPC 5864, Ann Arbor, MI 48109.E-mail: mthomps@med.umich.edu

Supplementary Material for this article is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.121.021629

For Sources of Funding and Disclosures, see page 8.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- The rapid growth in transcatheter aortic valve replacement (TAVR) has led to the development of claims-based outcome measures for public reporting, which rely on risk adjustment models to detect differences in outcomes across centers.
- Our analysis of national Medicare claims data suggests that claims-based TAVR outcome measures for mortality, discharge not to home, and stroke can reliably detect differences among hospitals, but readmission measures were unreliable.
- Center-level case volume was directly related to the reliability of TAVR outcome measures, with low-volume centers frequently having unreliable outcome measures.

What Are the Clinical Implications?

- Measures of TAVR mortality, discharge not to home, and stroke can reliably detect center-level differences over and above statistical noise, which could be used to encourage improvements in TAVR quality.
- Readmission rates for TAVR may not be an appropriate measure for public reporting, because it may reflect issues beyond the quality of care provided by a center.
- Ensuring TAVR quality in low-volume centers is critical to improving access, but reliably measuring quality in these centers remains challenging.

Nonstandard Abbreviations and Acronyms

MedPAR	Medicare Provider Analysis and Review
TAVR	transcatheter aortic valve replacement

Whether outcome measures derived from administrative claims data can reliably detect differences in outcomes among hospitals remains unclear. Prior studies have already shown variability in hospital-level mortality and readmission rates for TAVR.[10,11] It is assumed that differences in outcomes among hospitals can be attributed to systematic differences in quality after adjusting for patient risk.[12,13] On the other hand, variation in outcomes among hospitals could be the product of random statistical variation, particularly in hospitals with low case volumes.[14,15] The statistical reliability of TAVR outcome measures used to compare hospital performance is critical to future benchmarking, quality improvement efforts, and potentially reimbursement.

Therefore, this study sought to evaluate variability in claims-based TAVR outcome measures used in public reporting. We estimated center-level rates of 30-day mortality, discharge not to home/self-care, 30-day stroke, and 30-day readmission after TAVR, and the proportion of center-level outcome variation that can be attributed to systematic differences among hospitals, referred to as reliability (R). Additionally, this study sought to establish the case-volume threshold needed to achieve adequate reliability for hospital-level comparisons for each outcome measure, and the proportion of current hospitals that meet the threshold.

METHODS

This study was deemed exempt from human subject protections by the University of Michigan Institutional Review Board (HUM00163969). M.P.T. and H.H. both have full access to all of the data in the study and take responsibility for their integrity and the data analysis. Medicare data will not be made publicly available because of data use agreement restrictions, but methods and materials can be made available to any researcher for purposes of reproducing the results. All analyses were performed using SAS 9.4 (SAS Institute, Cary, NC), and statistical tests were deemed significant at $\alpha<0.05$ (2-sided).

Data Sources and Study Sample

Medicare administrative claims data from 2014 to 2018 were used for this study, including the 100% Medicare Provider Analysis and Review (MedPAR) and Medicare Beneficiary Summary File data sets (Medicare Data Use Agreement No. 21521). The MedPAR data include information for all inpatient discharges, including beneficiary demographics, admission and discharge information, and diagnosis and procedure codes. The Medicare Beneficiary Summary File includes information on beneficiary Medicare enrollment and dates of death. Additionally, hospital characteristics were obtained from the American Hospital Association Annual Survey and Medicare Impact Files.

The study sample included any Medicare beneficiary with a procedure code for TAVR during an inpatient admission and fully entitled to Medicare Part A and Part B for the 6 months before admission and at 90 days after discharge. *International Classification of Diseases Ninth Revision* (*ICD-9*; 35.05, 35.06) and *Tenth Revision* (*ICD-10*; 02RF3*) procedure codes were used to identify TAVR during inpatient admissions.

Outcomes

The primary outcomes for this study were 30-day mortality, discharge not to home, 30-day stroke, and

30-day readmission, which are currently used to profile hospital TAVR performance in *US News & World Report*.[9] Date of procedure from the MedPAR data and date of death from the Medicare Beneficiary Summary File data were used to define death within 30 days of procedure. Discharge location was used to define discharge not to home, which included discharge to a skilled nursing facility, intermediate care facility, swing bed, inpatient rehabilitation, long-term care facility, or other health care facility. Stroke within 30 days was defined as an inpatient admission with diagnosis of stroke within 30-days of the procedure (*ICD-9*: 433*, 434*, 436*, 437.9, 997.02; *ICD-10*: I63*, I97.81, I97.82). Beneficiaries who died in-hospital were excluded from the readmission definition. Admission and discharge dates from MedPAR were used to identify readmissions within 30-days of discharge. Readmissions that were potentially planned readmissions as defined by Medicare were also excluded from this definition.[16]

Covariates

Our study included covariates at the beneficiary and hospital level. Beneficiary-level factors were based on published criteria and drawn from the MedPAR and the Medicare Beneficiary Summary File, and included age, sex (men versus women), transfer status, year of admission, Medicare eligibility status (age, disability, or end-stage renal disease), Medicaid dual eligibility, and 29 Elixhauser comorbidities.[9,17,18] Race or ethnicity and elective procedure status were not included in risk-adjusted models because they are not used in adjustment for publicly reported outcomes.

Hospital Risk-Adjusted Outcomes

Hospital risk-adjusted outcomes were estimated for TAVR using a 2-stage approach applied in publicly reported outcomes.[9] The first stage used a standard logistic regression model to estimate a beneficiary-level probability of the outcome adjusting for beneficiary factors to generate the expected probability. The second stage used a hierarchical logistic regression model to estimate the probability of the outcome adjusting for the expected probability generated from the first model and a hospital-level random effect to get the predicted probability. Predicted and expected outcome probabilities were summed at the hospital level and used to estimate the ratio of predicted to expected for each hospital . Finally, we multiplied the hospital-specific predicted to expected ratio with the overall sample outcome rate to calculate the hospital risk-adjusted outcome rate. Model fit information for publicly reported TAVR outcomes can be found in Table S1.

Statistical Analysis

Reliability of hospital risk-adjusted outcome measures were estimated using established methods.[14,15,19] Briefly, measure reliability is defined as the ratio of between-hospital variation in outcome rates (the signal) to the sum of the between-hospital (the signal) and within-hospital variation in outcome rates (the noise), or signal/(signal+noise). In other words, an outcome measure's R value (range, 0–1) is the proportion of hospital-level variation that can be attributed to systematic differences among hospitals, which are assumed to be a proxy for hospital quality. The between-hospital variation estimate was drawn from the hierarchical logistic regression models as the variance in hospital random intercepts and is uniform across all hospitals. The within-hospital variation was estimated as the standard error of a proportion, or $\sqrt{[P(1-P)/n]}$, where P represents the probability of the outcome within a given hospital (ie, the hospital observed mortality or readmission rate), and n is the hospital case volume. Because both the observed outcome rates and case volume vary by hospital, each hospital has a unique within-hospital variation estimate and therefore a unique estimate of reliability.

Dot plots were created to show the relationship between a hospital's case volume and reliability estimates for individual outcome measures. A fitted Loess curve was added to illustrate the relationship between reliability and case volume at the hospital level. Median reliability (interquartile range [IQR]) was estimated for unadjusted and adjusted TAVR outcomes. Using prior studies, we used common benchmarks of reliability for group-level comparisons (R=0.7 and 0.9) and estimated the procedural volume threshold required to exceed those benchmarks.[14,15,19] Finally, we estimated the number of reliable risk-adjusted outcome measures per hospital (range, 0–4) for both benchmark levels. As a sensitivity analysis, we repeated our analysis using a 3-year period (2016–2018), which reflects time horizons used in other TAVR quality measures in the Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy Registry.

RESULTS

Between 2014 and 2018, a total of 119 554 Medicare beneficiaries underwent TAVR at 576 hospitals, with a mean of 211 beneficiaries per hospital (SD, 226). The characteristics of TAVR beneficiaries are shown in Table 1. Within this sample, 3762 (3.2%) died within 30 days, 50 672 (43.2%) were discharged not to home/self-care, 2980 (2.5%) had a stroke within 30 days, and 17 498 (14.9%) were readmitted within 30 days of discharge.

Table 1. Characteristics of Patients With Transcatheter Aortic Valve Replacement Outcome Status

Characteristic	No. (%) or mean (SD)
Age, y, mean (SD)	82 (8)
Men, n (%)	63 341 (53.0)
Year of admission, n (%)	
2014	12 410 (10.4)
2015	17 586 (14.7)
2016	25 435 (21.3)
2017	31 934 (26.7)
2018	32 189 (26.9)
Medicare eligibility, n (%)	
Aged without ESRD	113 260 (94.7)
Aged with ESRD	3832 (3.2)
Disabled without ESRD	1475 (1.2)
Disabled with ESRD	751 (0.6)
ESRD only	236 (0.2)
Dual eligibility, n (%)	13 990 (11.7)
Transferred patient, n (%)	6461 (5.4)
Clinical factors, n (%)	
Acquired immune deficiency syndrome	66 (0.1)
Alcohol abuse	1609 (1.4)
Chronic blood loss anemia	2895 (2.4)
Chronic pulmonary disease	39 535 (33.1)
Coagulopathy	20 795 (17.4)
Congestive heart failure	15 021 (12.6)
Deficiency anemias	38 168 (31.9)
Depression	13 323 (11.1)
Diabetes with chronic complications	24 515 (20.5)
Diabetes without chronic complications	29 353 (24.6)
Drug abuse	464 (0.4)
Fluid and electrolyte disorders	35 977 (30.1)
Hypertension	77 462 (64.8)
Hypothyroidism	28 569 (23.9)
Liver disease	3775 (3.2)
Lymphoma	1692 (1.4)
Metastatic cancer	999 (0.8)
Obesity	25 258 (21.1)
Other neurological disorders	11 137 (9.3)
Paralysis	3610 (3.0)
Peptic ulcer disease/bleeding	1241 (1.0)
Peripheral vascular disease	35 153 (29.4)
Psychoses	1561 (1.3)
Pulmonary circulation disease	1369 (1.2)
Renal failure	45 735 (38.3)
Rheumatoid arthritis	6969 (5.8)
Solid tumor without metastasis	3722 (3.1)
Valvular disease	21 588 (18.1)
Weight loss	6707 (5.6)

ESRD indicates end-stage renal disease.

The distribution of crude hospital-level outcomes can be found in Figure 1. After risk-adjustment, the median (IQR) outcome rates were as follows: 3.1% (2.9%–3.4%) for 30-day mortality, 41.4% (31.3%–53.4%) for discharge not to home/self-care, 2.5% (2.3%–2.7%) for 30-day stroke, and 14.9% (14.4%–15.5%) for 30-day readmission.

The relationship between hospital-level volume and reliability for 30-day risk-adjusted outcomes are shown in Figure 2. In general, as volume increased, the reliability of TAVR outcome measures increased. Median reliability was greatest for the 30-day stroke measure for the discharge not to home/self-care measure (R=0.95; IQR, 0.94–0.97), which exceeds both reliability benchmarks for group comparisons (Table 2). In other words, 95% of the variation in 30-day stroke among hospitals can be attributed to hospital-level differences, with the remaining 5% representing statistical noise. Median reliability for 30-day stroke (R=0.92; IQR, 0.87–0.94) and 30-day mortality measures (R=0.86; IQR, 0.81–0.91) were also above the benchmarks for group-level comparisons. The reliability of the 30-day readmission measure was R=0.42 (IQR, 0.35–0.51), which was below the benchmark for group-level comparisons.

The volume thresholds for highly reliable measures also differed by outcome. Hospitals would need to have at least 28 cases to have a highly reliable measure (R=0.90 benchmark) for discharge not to home/self-care, which was met by 93.3% of hospitals. The volume requirements for highly reliable 30-day stroke and mortality measures were 161 and 368 cases over the 5-year measurement period. These volume thresholds were met by 324 (57.1%) and 165 (29.1%) hospitals in this sample for 30-day stroke and mortality, respectively. For hospitals to have acceptable reliability in 30-day readmission (R=0.7), they would need to have at least 1438 cases over the 5-year measurement period. This threshold was only met by 4 hospitals (0.7%) in the sample.

Using a 3-year period, estimates of between-hospital variation and measure reliability were qualitatively similar for all measures (Table S2). However, the shorter time horizon led to fewer hospitals having reliable outcome measures for all measures because of smaller volume estimates.

When using the R=0.7 benchmark for reliable outcome measures, only 3 hospitals (0.5%) met the benchmark for all 4 outcome measures (Table 3). A total of 509 (89.8%) hospitals achieved reliability for 3 of the 4 measures, and 42 (7.4) hospitals achieved reliability for 2 of the 3 measures. Using the benchmark for high reliability (R=0.9), no hospitals achieved the benchmark for all 4 measures, and 109 (19.2%) achieved this benchmark for 3 of 4 measures. A plurality of hospitals met the high-reliability benchmark for only 2 of the 4 measures.

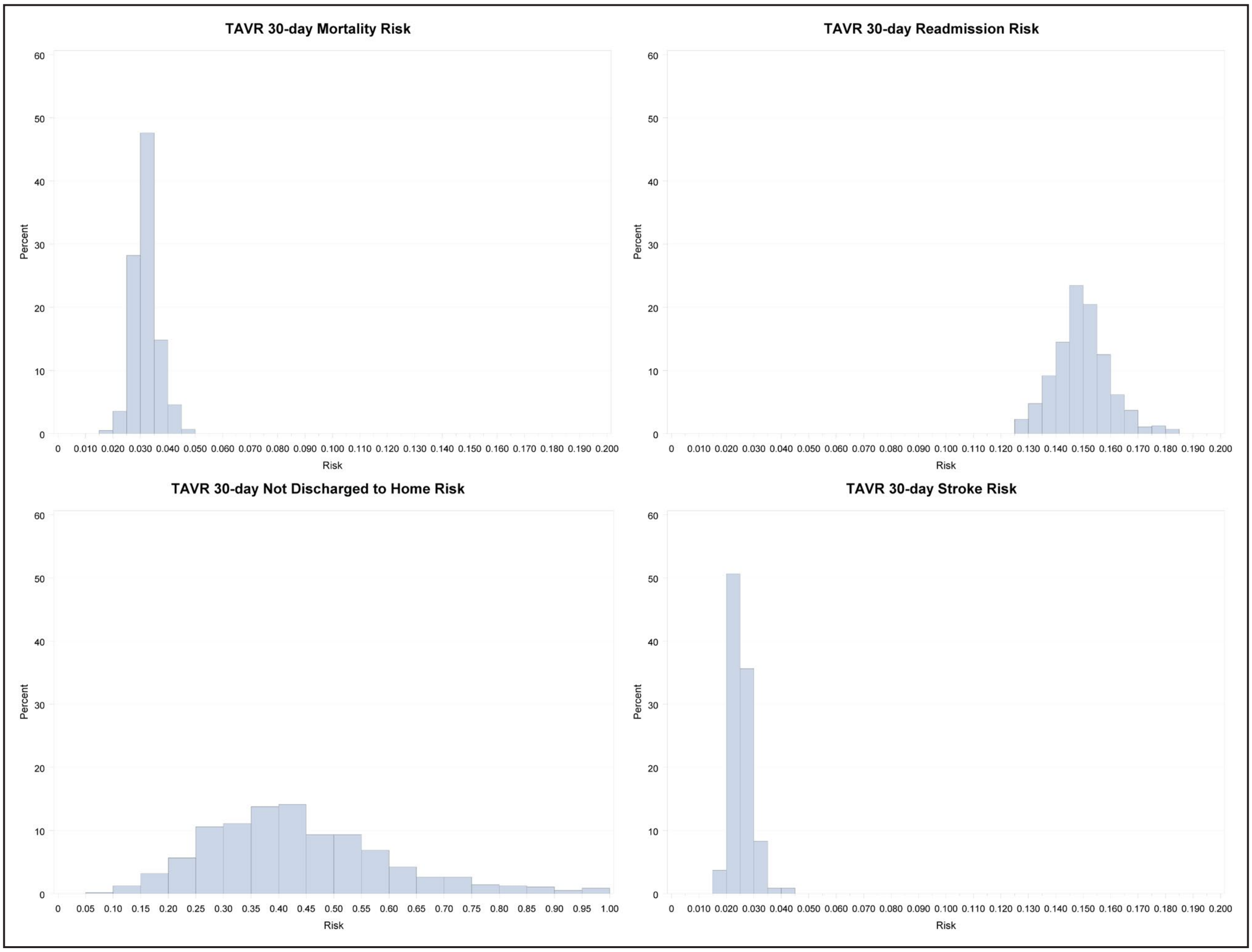

Figure 1. Distribution of center-level TAVR outcomes.
TAVR indicates transcatheter aortic valve replacement.

DISCUSSION

The usefulness of hospital-level quality measures is dependent upon its ability to identify systematic differences among hospitals over and above random statistical noise. Through this analysis, 5-year TAVR outcome measures for 30-day mortality, discharge not to home/self-care, and 30-day stroke demonstrated high reliability to detect systematic difference over random statistical noise. However, 30-day readmission rates for TAVR did not demonstrate sufficient reliability for group-level comparisons, with only 42% of the variability in readmission rates attributed to the hospital. We also found a strong relationship between center volume and measure reliability, with lower volume centers exhibiting less-reliable outcome measures.

There are limitations to our study that should be considered. First, Medicare claims data lack important clinical data on severity and extent of disease, which may explain some of the observed hospital-level variation in TAVR outcomes. However, the purpose of this study was to evaluate publicly reported outcomes that are derived from administrative claims data.[20] Second, our study focused on a single example of publicly reported outcome measures for TAVR.[9] Different risk-adjustment models may result in different estimates of between-hospital variation, and should be evaluated when they become available. Third, TAVR volumes have continually risen during and after the time period used in our analysis (2014–2018), which were reflected in our sample. Future work should seek to understand how the inclusion of more recent data may impact the findings of our study as TAVR volume continues to grow.[21,22] Fourth, our study did not analyze the same data used in public reporting. Although attempts were made to mimic existing approaches, differences in data management may produce results different from current publicly reported measures. Finally, our study sample includes only Medicare fee-for-service beneficiaries, which was done to mimic current publicly reported TAVR outcomes that use Medicare claims data. These findings may not be generalizable to non-Medicare fee-for-service populations.

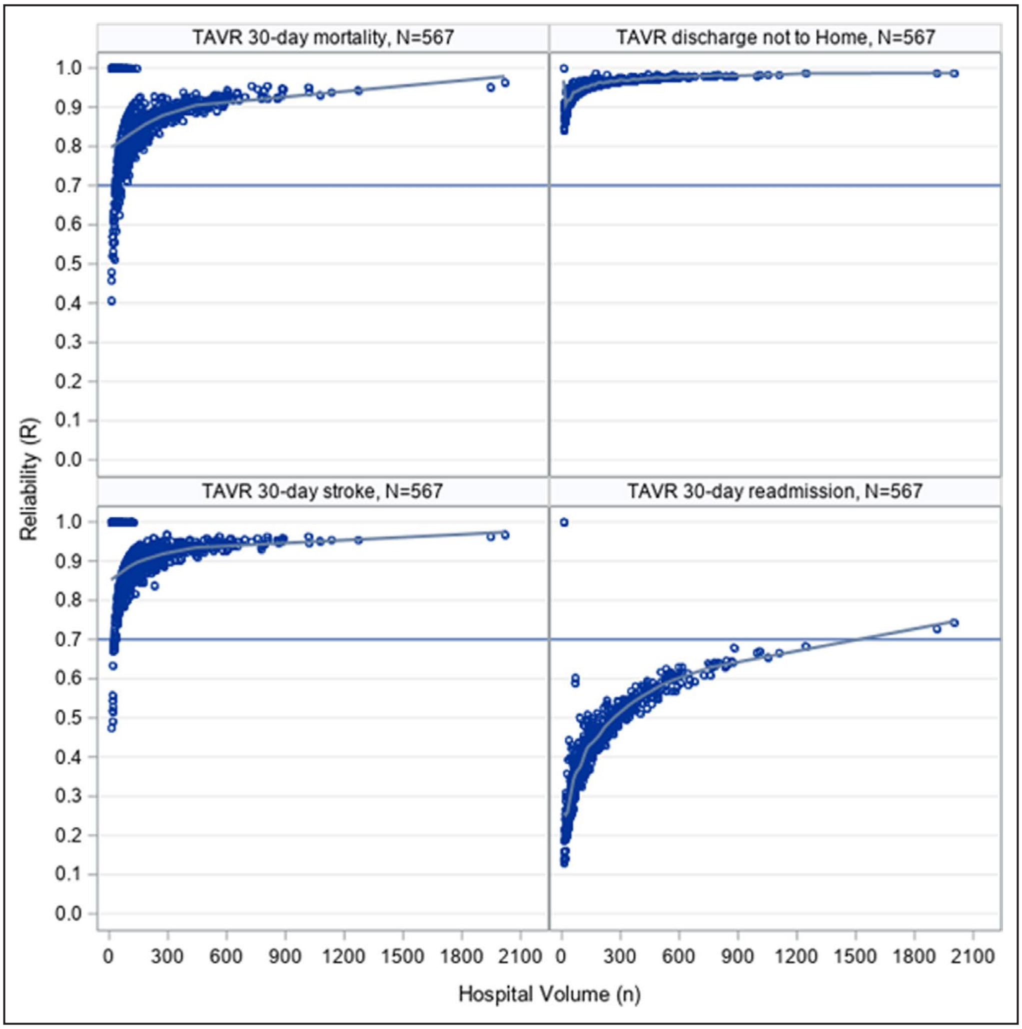

Figure 2. Relationship between center volume and reliability of TAVR outcome (n=567).
TAVR indicates transcatheter aortic valve replacement.

The findings from this study have important implications for ensuring TAVR quality through public reporting. Performance profiling is reliant on the ability to distinguish systematic differences among providers. Our results suggest that claims-based TAVR outcomes for mortality, discharge location, and stroke, achieves this purpose. However, a critical question remains: Are claims-based TAVR outcomes valid measures of performance? There is evidence to suggest that hospital performance derived from claims data may differ from performance derived from clinical registries.[23] Nevertheless, it is likely that public reporting will have significant implications for TAVR quality. Evaluations of public reporting for other cardiovascular procedures have demonstrated significant improvements in quality,[24,25] but also unintended effects on access to care and disparities.[26–28] More research is needed to better understand the concordance between publicly reported TAVR outcomes derived from claims data and clinical registries, and the subsequent impact of public reporting on TAVR care.

One unanticipated finding of our study pertains to the high reliability of the 30-day stroke measure for TAVR. An analysis of the Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy Registry demonstrated that a risk model for in-hospital stroke using clinical data had difficulty predicting the outcome using rich clinical registry data (C statistic, 0.622).[29] The present study did not seek to develop a prediction model for 30-day stroke following TAVR, but rather estimated the extent to which patient-level variation in 30-day stroke outcomes can be attributed to the hospital in which the patient was treated. Similar to our study, an analysis of postoperative stroke rates following cardiac surgery identified a strong hospital-level effect after adjusting for patient risk.[30] The analytic approach used in this study and in publicly reported outcome measures more broadly assumes that variation in risk-adjusted outcomes attributed to the hospital level is the direct result of differences in quality. However, other hospital-specific practices unrelated to quality may also explain systematic

Table 2. Median Reliability of Transcatheter Aortic Valve Replacement 30-Day Outcome Measures

Measure	Frequency (%)	Risk-adjusted rate, median (IQR)	Between-hospital variance	Reliability, median (IQR)	Volume threshold for R=0.70, median (IQR)	Hospitals with R≥0.70, n (%)	Volume threshold for R=0.90, median (IQR)	Hospitals with R≥0.90, n (%)
30-day mortality	3762 (3.2)	3.1% (2.9%–3.4%)	0.08	0.86 (0.81–0.91)	24 (17–34)	527 (93.0%)	368 (255–508)	165 (29.1%)
Discharge not to home*	50 672 (43.2)	41.4% (31.3%–53.4%)	0.81	0.95 (0.94–0.97)	1.8 (1.5–2.0)	567 (100%)	27.5 (22.2–30.2)	529 (93.3%)
30-day stroke	2980 (2.5)	2.5% (2.3%–2.7%)	0.10	0.92 (0.87–0.94)	11 (6–16)	552 (97.4%)	161 (89–238)	324 (57.1%)
30-day readmissions*	17 498 (14.9)	14.9% (14.4%–15.5%)	0.02	0.42 (0.35–0.51)	1438 (1220–1650)	4 (0.7%)	21 400 (18 148–24 544)	2 (0.4%)

IQR indicates interquartile range; and R, reliability.
*Excludes patients who died in-hospital: n=2263.

differences in outcomes, such as diagnostic and coding patterns. A comparison of post-TAVR stroke outcomes in clinical trial data versus administrative claims data showed stroke rates in trial data to be higher than those captured in claims data, with the sensitivity of claims-based algorithms to detect stroke between 19% and 67%.[31] This could be explained by active stroke surveillance in clinical trials or limitations in administrative claims coding of stroke. More work is needed to better understand the mechanisms underlying between-hospital differences in 30-day stroke rates for TAVR.

Using readmission rates to profile hospital TAVR quality may be more problematic. Readmission following surgery is widely accepted as a measure of care quality. The Society of Thoracic Surgeons benchmarks hospitals on readmissions after surgical aortic valve replacement, although they use risk models and comprehensive clinical data from the Society of Thoracic Surgeons Adult Cardiac Surgery Database, rather than Medicare administrative claims.[32] Perhaps more importantly, readmission measures are foundational in Medicare pay-for-performance and public reporting programs, although they have not yet been implemented in aortic valve replacement. Prior work has shown wide variability in TAVR readmissions across hospitals, suggesting it may be a useful quality measure.[33,34] However, our findings show that less than half of the between-hospital variation in readmissions was attributed to the hospital itself, because the variance estimate attributed to the hospital is markedly lower compared with other measures. Our findings add to the literature addressing the reliability of readmission rates for cardiac procedures, which have had mixed findings for coronary artery bypass grafting.[15,35] These results may not be surprising, because studies have consistently demonstrated that hospital readmissions are often a reflection of socioeconomic factors, rather than the clinical care provided in the initial admission.[36,37] Readmissions may still be an important quality indicator for TAVR, and efforts to reduce readmissions may be beneficial to patients, but their use in performance profiling may be limited.

Profiling TAVR outcomes in low-volume hospitals using claims-based outcomes may continue to be a challenge. Current hospital performance profiling methods are often sensitive to volume, with lower-volume hospitals exhibiting greater statistical noise.[38,39] Our findings similarly

Table 3. Number of Reliable Outcome Measures Per Hospital (n=567)

Reliability benchmark	No. (%) of reliable outcome measures				
	0	1	2	3	4
Reliability ≥0.70	0 (0.0)	13 (2.3)	42 (7.4)	509 (89.8)	3 (0.5)
Reliability ≥0.90	14 (2.5)	211 (37.2)	233 (41.1)	109 (19.2)	0 (0.0)

highlight a strong relationship between TAVR volume and outcome measure reliability, with lower-volume hospitals exhibiting less-reliable outcomes. The claims-based outcomes evaluated in this study attempt to address the low-volume problem by using a rolling 5-year window for measurement, which will be less sensitive to year-to-year changes in performance. If the measurement window were shortened to include fewer years, the resulting decline in volume would likely have detrimental effects to the reliability of TAVR outcome measures. Alternative methods to minimizing statistical noise in low-volume hospitals may be preferred, referred to as reliability adjustment or shrinkage methods.[19,40] However, these methods may mask legitimate outcome deviations in low-volume hospitals, where consistent evidence has demonstrated volume-outcome relationships with outcomes improving as volume increases.[41–45] Concerns about ensuring quality in low-volume TAVR centers have intensified with Medicare's recent decision to lower volume requirements for hospitals seeking to maintain or begin TAVR programs.[6] Professional societies have also echoed these concerns, and proposed additional paths to ensure quality in low-volume hospitals, such as requiring root cause analyses of adverse events.[46] Policymakers should consider our findings when developing and implementing performance metrics to track and benchmark quality in aortic valve replacement.

CONCLUSIONS

Claims-based TAVR outcome measures for mortality, discharge not to home, and stroke were reliable measures for hospital comparisons, but readmission measures were not reliable. Stakeholders should consider our findings when deciding to use claims-based measures to evaluate and compare hospital-level TAVR performance.

ARTICLE INFORMATION

Received May 5, 2021; accepted September 23, 2021.

Affiliations

Department of Cardiac Surgery (M.P.T., H.H., A.A.B., F.D.P., D.S.L.) and Division of Cardiovascular Medicine, Department of General Internal Medicine (D.S.), Michigan Medicine, Ann Arbor, MI; and Institute for Healthcare Policy and Innovation (M.P.T., A.A.B., F.D.P., D.S.L.) and Department of Health Management and Policy, School of Public Health (J.S.M.), University of Michigan, Ann Arbor, MI.

Sources of Funding

This project was supported by the Institute for Healthcare Policy and Innovation at the University of Michigan. Access to data was supported by Dr Dimick through a federal grant from the National Institute on Aging (R01AG039434).

Disclosures

Dr Thompson receives partial salary support from Blue Cross Blue Shield of Michigan and receives funding from the Agency for Healthcare Research and Quality (K01HS027830-01A1) outside of the submitted work. Dr Brescia is supported by the National Research Service Award postdoctoral fellowship (5T32HL076123) outside of the submitted work. Dr Sukul receives partial salary support from Blue Cross Blue Shield of Michigan outside of the submitted work and is a member of the Government Relations Committee for the Society for Cardiovascular Angiography & Interventions, which had no role in the work herein. Dr Likosky is a consultant to the American Society of Extracorporeal Technology and receives funding from the Agency for Healthcare Research and Quality and National Institutes of Health, all outside of the submitted work. The remaining authors have no disclosures to report.

Supplementary Material

Tables S1–S2

REFERENCES

1. Leon MB, Smith CR, Mack M, Miller DC, Moses JW, Svensson LG, Tuzcu EM, Webb JG, Fontana GP, Makkar RR, et al. Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. *N Engl J Med*. 2010;363:1597–1607. doi: 10.1056/NEJMoa1008232
2. Adams DH, Popma JJ, Reardon MJ, Yakubov SJ, Coselli JS, Deeb GM, Gleason TG, Buchbinder M, Hermiller J, Kleiman NS, et al. Transcatheter aortic-valve replacement with a self-expanding prosthesis. *N Engl J Med*. 2014;370:1790–1798. doi: 10.1056/NEJMoa1400590
3. Mack MJ, Leon MB, Thourani VH, Makkar R, Kodali SK, Russo M, Kapadia SR, Malaisrie SC, Cohen DJ, Pibarot P, et al. Transcatheter aortic-valve replacement with a balloon-expandable valve in low-risk patients. *N Engl J Med*. 2019;380:1695–1705. doi: 10.1056/NEJMoa1814052
4. Popma JJ, Deeb GM, Yakubov SJ, Mumtaz M, Gada H, O'Hair D, Bajwa T, Heiser JC, Merhi W, Kleiman NS, et al. Transcatheter aortic-valve replacement with a self-expanding valve in low-risk patients. *N Engl J Med*. 2019;380:1706–1715. doi: 10.1056/NEJMoa1816885
5. Leon MB, Smith CR, Mack MJ, Makkar RR, Svensson LG, Kodali SK, Thourani VH, Tuzcu EM, Miller DC, Herrmann HC, et al. Transcatheter or surgical aortic-valve replacement in intermediate-risk patients. *N Engl J Med*. 2016;374:1609–1620. doi: 10.1056/NEJMoa1514616
6. Centers for Medicare and Medicaid Services. Decision Memo for Transcatheter Aortic Valve Replacement (TAVR) (CAG-00430R). Medicare Coverage Database 2019. https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=293&type=Open&bc=ACAAAAAAQCAA&. Accessed September 3, 2019.
7. Mack MJ, Brennan JM, Brindis R, Carroll J, Edwards F, Grover F, Shahian D, Tuzcu EM, Peterson ED, Rumsfeld JS, et al. Outcomes following transcatheter aortic valve replacement in the United States. *JAMA*. 2013;310:2069–2077. doi: 10.1001/jama.2013.282043
8. Grover FL, Vemulapalli S, Carroll JD, Edwards FH, Mack MJ, Thourani VH, Brindis RG, Shahian DM, Ruiz CE, Jacobs JP, et al. 2016 annual report of the Society of Thoracic Surgeons/American College of Cardiology transcatheter valve therapy registry. *J Am Coll Cardiol*. 2017;69:1215–1230.
9. U.S. News Releases 2020-21 Best Hospitals Rankings and Special 'Hospital Heroes' Series During Historic Year for Health care. US News & World Report 2020. Available at: https://www.usnews.com/info/blogs/press-room/articles/2020-07-28/us-news-releases-2020-21-best-hospitals-rankings-and-special-hospital-heroes-series-during-historic-year-for-health-care. Accessed December 3, 2020.
10. O'Brien SM, Cohen DJ, Rumsfeld JS, Brennan JM, Shahian DM, Dai D, Holmes DR, Hakim RB, Thourani VH, Peterson ED, et al. Variation in hospital risk–adjusted mortality rates following transcatheter aortic valve replacement in the United States. A report from the Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy Registry 2016. *Circ Cardiovasc Qual Outcomes*. 2016;9:560–565.
11. Murugiah K, Wang Y, Desai NR, Nuti SV, Krumholz HM. Hospital variation in outcomes for transcatheter aortic valve replacement among medicare beneficiaries 2011–2013. *J Am Coll Cardiol*. 2015;66:2678–2679.
12. Shahian DM, Normand SL. Comparison of "risk-adjusted" hospital outcomes. *Circulation*. 2008;117:1955–1963. doi: 10.1161/CIRCULATIONAHA.107.747873
13. Krumholz HM, Wang Y, Mattera JA, Wang Y, Han LF, Ingber MJ, Roman S, Normand SLT. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients

with an acute myocardial infarction. *Circulation*. 2006;113:1683–1692. doi: 10.1161/CIRCULATIONAHA.105.611186
14. Adams JL, Mehrotra A, Thomas JW, McGlynn EA. Physician cost profiling — reliability and risk of misclassification. *N Engl J Med*. 2010;362:1014–1021. doi: 10.1056/NEJMsa0906323
15. Thompson MP, Kaplan CM, Cao Y, Bazzoli GJ, Waters TM. Reliability of 30-day readmission measures used in the hospital readmission reduction program. *Health Serv Res*. 2016;51:2095–2114. doi: 10.1111/1475-6773.12587
16. Centers for Medicare and Medicaid Services. Readmissions Reduction Program 2014. Available at: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html. Accessed December 12, 2014.
17. van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. *Med Care*. 2009;47:626–633. doi: 10.1097/MLR.0b013e31819432e5
18. Moore BJ, White S, Washington R, Coenen N, Elixhauser A. Identifying increased risk of readmission and in-hospital mortality using hospital administrative data: the AHRQ Elixhauser Comorbidity Index. *Med Care*. 2017;55:698–705. doi: 10.1097/MLR.0000000000000735
19. Krell RW, Hozain A, Kao LS, Dimick JB. Reliability of risk-adjusted outcomes for profiling hospital surgical quality. *JAMA Surg*. 2014;149:467–474. doi: 10.1001/jamasurg.2013.4249
20. Brescia AA, Rankin JS, Cyr DD, Jacobs JP, Prager RL, Zhang M, Matsouaka RA, Harrington SD, Dokholyan RS, Bolling SF, et al. Determinants of variation in pneumonia rates after coronary artery bypass grafting. *Ann Thorac Surg*. 2018;105:513–520. doi: 10.1016/j.athoracsur.2017.08.012
21. Kundi H, Strom JB, Valsdottir LR, Elmariah S, Popma JJ, Shen C, Yeh RW. Trends in isolated surgical aortic valve replacement according to hospital-based transcatheter aortic valve replacement volumes. *JACC Cardiovasc Interv*. 2018;11:2148–2156.
22. Culler SD, Cohen DJ, Brown PP, Kugelmass AD, Reynolds MR, Ambrose K, Schlosser ML, Simon AW, Katz MR. Trends in aortic valve replacement procedures between 2009 and 2015: has transcatheter aortic valve replacement made a difference? *Ann Thorac Surg*. 2018;105:1137–1143.
23. Lawson EH, Louie R, Zingmond DS, Brook RH, Hall BL, Han L, Rapp M, Ko CY. A comparison of clinical registry versus administrative claims data for reporting of 30-day surgical complications. *Ann Surg*. 2012;256:973–981. doi: 10.1097/SLA.0b013e31826b4c4f
24. Hannan EL, Kilburn H Jr, Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA*. 1994;271:761–766. doi: 10.1001/jama.1994.03510340051033
25. Romano PS, Marcin JP, Dai JJ, Yang XD, Kravitz RL, Rocke DM, Dharmar M, Li Z. Impact of public reporting of coronary artery bypass graft surgery performance data on market share, mortality, and patient selection. *Med Care*. 2011;49:1118–1125. doi: 10.1097/MLR.0b013e3182358c78
26. Joynt KE, Blumenthal DM, Orav EJ, Resnic FS, Jha AK. Association of public reporting for percutaneous coronary intervention with utilization and outcomes among Medicare beneficiaries with acute myocardial infarction. *JAMA*. 2012;308:1460–1468. doi: 10.1001/jama.2012.12922
27. Resnic FS, Welt FGP. The public health hazards of risk avoidance associated with public reporting of risk-adjusted outcomes in coronary intervention. *J Am Coll Cardiol*. 2009;53:825–830. doi: 10.1016/j.jacc.2008.11.034
28. McCabe JM, Joynt KE, Welt FGP, Resnic FS. Impact of public reporting and outlier status identification on percutaneous coronary intervention case selection in Massachusetts. *JACC Cardiovasc Interv*. 2013;6:625–630. doi: 10.1016/j.jcin.2013.01.140
29. Thourani VH, O'Brien SM, Kelly JJ, Cohen DJ, Peterson ED, Mack MJ, Shahian DM, Grover FL, Carroll JD, Brennan JM, et al. Development and application of a risk prediction model for in-hospital stroke after transcatheter aortic valve replacement: a report from the Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy Registry. *Ann Thorac Surg*. 2019;107:1097–1103.
30. LaPar DJ, Quader M, Rich JB, Kron IL, Crosby IK, Kern JA, Tribble CG, Speir AM, Ailawadi G. Institutional variation in mortality after stroke after cardiac surgery: an opportunity for improvement. *Ann Thorac Surg*. 2015;100:1276–1282; discussion 1282–1283.
31. Strom JB, Zhao Y, Faridi KF, Tamez H, Butala NM, Valsdottir LR, Curtis J, Brennan JM, Shen C, Boulware M, et al. Comparison of clinical trials and administrative claims to identify stroke among patients undergoing aortic valve replacement. *Circ Cardiovasc Interv*. 2019;12:e008231. doi: 10.1161/CIRCINTERVENTIONS.119.008231
32. Shahian DM, He X, O'Brien SM, Grover FL, Jacobs JP, Edwards FH, Welke KF, Suter LG, Drye E, Shewan CM, et al. Development of a clinical registry-based 30-day readmission measure for coronary artery bypass grafting surgery. *Circulation*. 2014;130:399–409. doi: 10.1161/CIRCULATIONAHA.113.007541
33. Kolte D, Khera S, Sardar MR, Gheewala N, Gupta T, Chatterjee S, Goldsweig A, Aronow WS, Fonarow GC, Bhatt DL, et al. Thirty-day readmissions after transcatheter aortic valve replacement in the United States: insights from the nationwide readmissions database. *Circ Cardiovasc Interv*. 2017;10:e004472. doi: 10.1161/CIRCINTERVENTIONS.116.004472
34. Hannan EL, Samadashvili Z, Jordan D, Sundt TM, Stamato NJ, Lahey SJ, Gold JP, Wechsler A, Ashraf MH, Ruiz C, et al. Thirty-day readmissions after transcatheter aortic valve implantation versus surgical aortic valve replacement in patients with severe aortic stenosis in New York state. *Circ Cardiovasc Interv*. 2015;8:e002744. doi: 10.1161/CIRCINTERVENTIONS.115.002744
35. Shih T, Dimick JB. Reliability of readmission rates as a hospital quality measure in cardiac surgery. *Ann Thorac Surg*. 2014;97:1214–1218. doi: 10.1016/j.athoracsur.2013.11.048
36. Meddings J, Reichert H, Smith SN, Iwashyna TJ, Langa KM, Hofer TP, McMahon LF Jr. The impact of disability and social determinants of health on condition-specific readmissions beyond Medicare risk adjustments: a cohort study. *J Gen Intern Med*. 2017;32:71–80. doi: 10.1007/s11606-016-3869-x
37. Herrin J, St. Andre J, Kenward K, Joshi MS, Audet A-M, Hines SC. Community factors and hospital readmission rates. *Health Serv Res*. 2014;50:20–39. doi: 10.1111/1475-6773.12177
38. Ash AS, Fienberg SE, Louis TA, Normand S-LT, Stukel TA. *Statistical Issues in Assessing Hospital Performance*. COPSS-CMS White Paper Committee; 2012.
39. Silber JH, Rosenbaum PR, Brachet TJ, Ross RN, Bressler LJ, Even-Shoshan O, Lorch SA, Volpp KG. The hospital compare mortality model and the volume-outcomes relationship. *Health Serv Res*. 2010;45:1148–1167.
40. Dimick JB, Ghaferi AA, Osborne NH, Ko CY, Hall BL. Reliability adjustment for reporting hospital outcomes with surgery. *Ann Surg*. 2012;255:703–707. doi: 10.1097/SLA.0b013e31824b46ff
41. Vemulapalli S, Carroll JD, Mack MJ, Li Z, Dai D, Kosinski AS, Kumbhani DJ, Ruiz CE, Thourani VH, Hanzel G, et al. Procedural volume and outcomes for transcatheter aortic-valve replacement. *N Engl J Med*. 2019;380:2541–2550. doi: 10.1056/NEJMsa1901109
42. Mao J, Redberg RF, Carroll JD, Marinac-Dabic D, Laschinger J, Thourani V, Mack M, Sedrakyan A. Association between hospital surgical aortic valve replacement volume and transcatheter aortic valve replacement outcomes. *JAMA Cardiol*. 2018;3:1070–1078. doi: 10.1001/jamacardio.2018.3562
43. Khera S, Kolte D, Gupta T, Goldsweig A, Velagapudi P, Kalra A, Tang GHL, Aronow WS, Fonarow GC, Bhatt DL, et al. Association between hospital volume and 30-day readmissions following transcatheter aortic valve replacement. *JAMA Cardiol*. 2017;2:732–741. doi: 10.1001/jamacardio.2017.1630
44. Gonzalez AA, Dimick JB, Birkmeyer JD, Ghaferi AA. Understanding the volume-outcome effect in cardiovascular surgery: the role of failure to rescue. *JAMA Surg*. 2014;149:119–123. doi: 10.1001/jamasurg.2013.3649
45. Dewey TM, Herbert MA, Ryan WH, Brinkman WT, Smith R, Prince SL, Mack MJ. Influence of surgeon volume on outcomes with aortic valve replacement. *Ann Thorac Surg*. 2012;93:1107–1113. doi: 10.1016/j.athoracsur.2011.09.064
46. Bavaria JE, Tommaso CL, Brindis RG, Carroll JD, Deeb GM, Feldman TE, Gleason TG, Horlick EM, Kavinsky CJ, Kumbhani DJ, et al. 2018 AATS/ACC/SCAI/STS expert consensus systems of care document: operator and institutional recommendations and requirements for transcatheter aortic valve replacement: a Joint Report of the American Association for Thoracic Surgery, American College of Cardiology, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. *J Am Coll Cardiol*. 2019;73:340–374. doi: 10.1016/j.jacc.2018.07.002

SUPPLEMENTAL MATERIAL

Table S1. **Model fit information for publicly reported TAVR outcome measures.**

Measure	Hosmer and Lemeshow Goodness-of-Fit Test, Chi-square (DF), p-value	Area Under the Curve (AUC)	Akaike Information Criterion (AIC)
30-Day Mortality	7.7458 (8), p-value=0.4587	AUC=0.7519	Intercept Only: AIC=22009; Intercept and Covariates: AIC=19933
Discharge not to Home*	93.5899 (8), p-value=<.0001	AUC=0.7179	Intercept Only: AIC=116851; Intercept and Covariates: AIC=104249
30-Day Stroke	9.1241 (8), p-value=0.3319	AUC=0.7685	Intercept Only: AIC=21278; Intercept and Covariates: AIC=17522
30-Day Readmissions*	28.748 (8), p-value=0.0004	AUC=0.6383	Intercept Only: AIC= 71493; Intercept and Covariates: AIC= 68930

Table S2. Median reliability of TAVR 30-day outcome measures using 3-year time window.

Measure	Risk-Adjusted Rate, Median (IQR)	Between-Hospital Variance	Reliability (R), Median (IQR)	Volume Threshold for R=0.70, Median (IQR)	Hospitals with R ≥ 0.70, n (%)	Volume Threshold for R=0.90, Median (IQR)	Hospitals with R ≥ 0.90, n (%)
30-Day Mortality	2.6% (2.5%-2.9%)	0.08	0.86 (0.80-0.90)	24 (14-36)	516 (91.3%)	353 (214-532)	152 (26.9%)
Discharge not to Home	35.6% (27.1%-46.8%)	0.81	0.95 (0.93-0.96)	1.7 (1.4-2.0)	565 (100%)	26 (21-29)	526 (93.1%)
30-Day Stroke	2.5% (2.4%-2.8%)	0.11	0.90 (0.86-0.94)	10 (5-15)	551 (97.5%)	151 (77-224)	292 (51.7%)
30-Day Readmissions	14.0% (13.5%-14.5%)	0.02	0.40 (0.33-0.48)	1,430 (1,211-1,680)	3 (0.5%)	21,273 (18,024-24,993)	2 (0.4%)

TAVR = transcatheter aortic valve replacement, IQR = interquartile range, R = Reliability