5-5-2021

A Composite Metric for Benchmarking Site Performance in TAVR: Results from the STS/ACC TVT Registry

Nimesh D. Desai

Sean M. O'Brien

David J. Cohen

John Carroll

Sreekanth Vemulapalli

See next page for additional authors

Recommended Citation

Desai ND, O'Brien SM, Cohen DJ, Carroll J, Vemulapalli S, Arnold SV, Forrest JK, Thourani VH, Kirtane AJ, O'Neill B, Manandhar P, Shahian DM, Badhwar V, and Bavaria JE. A Composite Metric for Benchmarking Site Performance in TAVR: Results from the STS/ACC TVT Registry. Circulation 2021.

Authors

Nimesh D. Desai, Sean M. O'Brien, David J. Cohen, John Carroll, Sreekanth Vemulapalli, Suzanne V. Arnold, John K. Forrest, Vinod H. Thourani, Ajay J. Kirtane, Brian P. O'Neill, Pratik Manandhar, David M. Shahian, Vinay Badhwar, and Joseph E. Bavaria

10.1161/CIRCULATIONAHA.120.051456

A Composite Metric for Benchmarking Site Performance in TAVR: Results from the STS/ACC TVT Registry

Running Title: *Desai et al.; STS/ACC TVT Registry Composite Metric*

Nimesh D. Desai, MD, PhD[1,2]; Sean M. O'Brien, PhD[3]; David J. Cohen, MD, MSc[4,5]; John Carroll, MD[6]; Sreekanth Vemulapalli, MD[3]; Suzanne V. Arnold, MD, MHA[7]; John K. Forrest, MD[8]; Vinod H. Thourani, MD[9]; Ajay J. Kirtane, MD[10]; Brian O'Neil, MD[11]; Pratik Manandhar, MS[3]; David M. Shahian, MD[12]; Vinay Badhwar, MD[13]; Joseph E. Bavaria MD[1,2]

[1]Division of Cardiovascular Surgery, Hospital of the University of Pennsylvania, Philadelphia, PA; [2]Penn Cardiovascular Outcomes, Quality, and Evaluative Research Center, Leonard Davis Institute of Health Economics, Philadelphia, PA; [3]Duke Clinical Research Institute, Duke University, Durham, NC; [4]St. Francis Hospital, Rosylyn, NY; [5]Cardiovascular Research Foundation, New York, NY; [6]Division of Cardiology, Department of Medicine, University of Colorado School of Medicine, Aurora, CO; [7]Department of Cardiovascular Medicine, Saint Luke's Mid America Heart Institute, Kansas City, MO; [8]Section of Cardiovascular Medicine, Yale School of Medicine, New Haven, CT; [9]Department of Surgery, Piedmont Hospital, Atlanta, GA; [10]Department of Medicine, Columbia University, New York, NY; [11]Division of Cardiology, Henry Ford Hospital, Detroit, MI; [12]Division of Cardiac Surgery and Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA; [13]Department of Cardiovascular and Thoracic Surgery, West Virginia University, Morgantown, WV

Address for Correspondence:
Nimesh D. Desai, MD, PhD
Hospital of the University of Pennsylvania
3400 Spruce Street, Philadelphia, PA 19104
Email: nimesh.desai@pennmedicine.upenn.edu

Abstract

Background: Transcatheter aortic valve replacement (TAVR) is a transformative therapy for aortic stenosis. Despite rapid improvements in technology and techniques, serious complications remain relatively common and are not well described by single outcome measures. The purpose of this study was to determine if there is site-level variation in TAVR outcomes in the United States using a novel 30-day composite measure.

Methods: We performed a retrospective cohort study using data from the STS/ACC TVT Registry to develop a novel ranked composite performance measure that incorporates mortality and serious complications. The selection and rank order of the complications for the composite was determined by their adjusted association with 1-year outcomes. Sites whose risk-adjusted outcomes were significantly more or less frequent than the national average based on a 95% probability interval were classified as performing worse or better than expected.

Results: The development cohort consisted of 52,561 patients who underwent TAVR between January 1, 2015 and December 31, 2017. Based on the associations with 1-year risk-adjusted mortality and health status, we identified four periprocedural complications to include in the composite risk model in addition to mortality. Ranked empirically according to severity, these included stroke, major, life-threatening or disabling bleeding, stage III acute kidney injury, and moderate or severe peri-valvular regurgitation. Based on these ranked outcomes, we found that there was significant site-level variation in quality of care in TAVR in the United States. Overall, better than expected site performance was observed in 25/301 (8%) of sites; performance as expected was observed in 242/301 sites (80%); and worse than expected performance was observed in 34/301 (11%) of sites. Thirty-day mortality, stroke, major, life-threatening or disabling bleeding, and moderate or severe peri-valvular leak were each substantially more common in sites with worse than expected performance as compared with other sites. There was good aggregate reliability of the model.

Conclusions: There are substantial variations in the quality of TAVR care received in the United States, and 11% of sites were identified as providing care below the average level of performance. Further study is necessary to determine structural, process-related, and technical factors associated with high- and low-performing sites.

Key Words: cardiac valve prosthesis; performance measurement; quality outcomes; transcatheter aortic valve implantation

Clinical Perspective

What is new?

- Transcatheter Aortic Valve Replacement(TAVR) is a breakthrough technology that has revolutionized the care of patients with aortic stenosis.
- While the use of TAVR has grown rapidly, it is unknown if there are site-level differences in the quality of care
- Using data from the Transcatheter Valve Therapy Registry, a national registry of all commercial TAVRs in the United States, we found significant variation in the quality of care delivered occurring at the site level.

What are the clinical implications?

- TAVR programs may use the feedback generated from this analysis to study and improve processes of care and clinical outcomes.

Introduction

Transcatheter aortic valve replacement (TAVR) is a rapidly developing technology for the treatment of aortic stenosis that was first approved for commercial use in 2011 by the Food and Drug Administration(FDA) and subsequent coverage by the Centers for Medicare and Medicaid Services (CMS)in 2012[1-3]. Both mortality and major complications after TAVR have been declining due to advances in technology, techniques, and expansion of TAVR to healthier patients.[4] Nonetheless, several publications from the STS/ACC Transcatheter Valve Therapies (TVT) Registry as well as other international registries have documented hospital-specific variation in short-term mortality after TAVR, suggesting differential quality of care is being provided at the hospital level.[5,6]

As part of the initial coverage process, the Centers for Medicare and Medicaid Services (CMS) mandated that all TAVR procedures in the United States be captured in a prospective clinical registry with one year follow-up.[3] In 2019, CMS published an update to their 2012 coverage determination for TAVR recommending continuation of the registry and the development of evidence linking procedure-related complications with longer term patient health outcomes. [7] In their National Coverage Determination (NCD), CMS specifically stated that a periprocedural composite metric incorporating relevant patient health outcomes was a priority and may eventually replace volume thresholds in their Coverage with Evidence Decision(CED) regarding TAVR reimbursement.[7]

In an effort to assist patients in their health care choices and to respond to CMS guidance, we therefore sought to develop and validate a performance measure for TAVR. The broad goals of this performance measure were to serve as a national benchmark for quality-of-care monitoring and to support both local quality improvement efforts as well as national efforts to

use quality rather than volume requirements to maintain TAVR programs. Based, in part, on CMS recommendations, we modeled a ranked composite endpoint including both fatal and non-fatal complications with rankings chosen based on the association of these complications with late mortality as well as patient-reported health status.

Methods

The data, analytic methods, and study are available from the data analytic center on reasonable request via established access pathways to the TVT registry.

STS/ACC TVT Registry

The STS/ACC TVT registry was created in 2011 through partnership between the Society of Thoracic Surgeons and the American College of Cardiology in collaboration with the FDA and the Centers for Medicare and Medicaid services.[8] One of the original mandates of the TVT registry was to develop risk models specifically tailored for the TAVR population to allow for benchmark comparisons of risk-adjusted outcomes among centers. Valve Academic Research Consortium (VARC) standardized definitions of outcomes variables are captured peri-procedurally and at one year.[9] Data elements are completed according to a standardized data dictionary and validated for completeness and accuracy by the analytic center. A random audit of 10% of data elements is performed either on-site or remotely. This study was approved by the Duke University Institutional Review Board (ClinicalTrials.gov Identifier: NCT01737528).

Patient-Reported Health Status

In addition to clinical outcomes, based on the CMS mandate, the TVT registry collects data on patient-reported health status at baseline, 1 month, and 1-year after TAVR using the short version of the Kansas City Cardiomyopathy Questionnaire (KCCQ-12). [10] The KCCQ-OS is

an overall summary score derived from the four measured KCCQ-12 domains (physical limitation, symptom frequency, quality of life, social limitation). It ranges from 0 to 100 with higher scores indicating less symptom burden and better quality of life. Differences in the KCCQ-OS of 5, 10, and 20 points correspond to small, moderate or large clinical changes at the individual patient-level.[11]

Study Cohort

For the purposes of model development, we included all patients undergoing TAVR for symptomatic aortic stenosis between Jan 1, 2015 and Dec 31, 2017. Based on conventions established for the TVT 30-day mortality model, data from hospitals with >10% missing data for the outcome variable and other key study variables were excluded. Baseline patient characteristics are presented in Table 1. We limited the analysis to a 3-year time period to ensure the use of recent generation devices and to allow for one-year follow-up in all patients. Additionally, with the expansion of TAVR centers over the years since commercial approval in 2011, the longer time interval may potentially mitigate learning curve issues a program may experience in its early implementation.[12] Only the first TAVR per patient during the study period was included.

Ranked Composite Outcome Measure

Composite performance measures of procedures in the cardiovascular field are generally composed of mortality and non-fatal outcomes that cause substantial morbidity, which are typically selected by expert panels. Although the most common approach to analyzing such composite endpoints for performance measurement is to combine them into a single binary outcome, different nonfatal outcomes often have very different effects on the long term health status of the patient; as such, it may not be appropriate to treat them equally. Subjective

weighting of non-fatal outcomes by expert panels may introduce bias by failing to account for patient preferences.[13] Based on the assumption that patients would prefer to be alive and well after their TAVR, we used a strategy that incorporated one-year survival to empirically guide the ranking of non-fatal outcomes.

The components of the composite endpoint were determined by establishing which non-fatal periprocedural complications were independently with one-year mortality and the strength of this association. Details of this analysis, which was similar to previous work by Arnold and colleagues [14] are described in detail in Section 1 of the Statistical Supplement . The final hierarchical composite endpoint, as derived from the one year mortality association, consisted of (1) death; (2) stroke; (3) Stage III acute kidney injury (AKI); (4) major, life-threatening, or disabling bleeding; (5) moderate or severe peri-valvular regurgitation; and (6) none of the aforementioned complications. Definitions of the outcomes included in the final global rank composite measure are presented in Table 2. Peri-valvular regurgitation was determined at the latest echocardiogram within the 30-day time frame. If a patient experienced multiple outcomes captured in the overall rank composite measure, the outcome with the highest rank was assigned. We also studied the association between these non-fatal complications and one year patient reported health status by KCCQ-OS and found them to be similar in both direction and magnitude to the mortality associations.

Statistical Methods

Statistical analyses were performed using SAS software version 9.4 (SAS Institute, Inc., Cary, North Carolina) and R version 3.4 (R Core Team, 2019). Hospital-specific outcomes were adjusted for 46 case mix factors chosen to match the most recently published TVT TAVR mortality model (See Supplemental Table 1) [15]. We used a generalized linear mixed model

(GLMM) proportional odds model with logit link and multinomial distributional assumption, incorporating a random effect for site to account for potential clustering and to produce unbiased standard errors. Parameters were estimated by maximum likelihood using SAS PROC GLIMMIX software. The statistical framework was a hierarchical proportional odds regression model with random hospital intercepts [16]. The proportional odds methodology is an extension of binary logistic regression to accommodate multiple ordered outcome categories. We included hospital-level random effects in order to estimate hospital-level variation while accounting for the multi-level data structure and to obtain reliability-adjusted estimates of each individual hospital's performance.

To facilitate interpretation, performance estimates from the model are expressed in terms of a "site difference" metric that resembles the "win ratio" and "net benefit" approaches developed for assessing ranked outcomes in clinical trials based on the numbers of winners and losers in paired analyses [17, 18]. Conceptually, the site difference involves pairing each patient treated by the TAVR hospital of interest with a hypothetical patient having identical risk factors who is treated by an average-performing reference hospital. The site difference is calculated as the model-predicted proportion of winning pairs minus losing pairs where "winning pair" means that the hospital of interest's patient had a better outcome compared to the reference hospital and "losing pair" means that the reference hospital's patient had a worse outcome compared to the reference hospital. A site difference greater than zero implies that a hospital's outcomes are better than expected in light of its case mix, whereas a site difference value less than zero implies the opposite.

To account for uncertainty due to chance fluctuations in a hospital's outcomes, we calculated a 95% empirical Bayes probability interval around each hospital's site difference estimate. Hospitals were then categorized as performing better than expected, worse than expected, or as expected based on whether the 95% probability interval fell entirely below zero, entirely above zero, or was overlapping with zero. There was no attempt to pre-specify the proportion of hospitals that would be categorized as performing within or outside the expected range. Full details of the model are presented in the Supplemental Methods Appendix.

Analysis of Individual Endpoints Stratified by Hospital Categories:

To ensure that hospital classifications based on the ordinal composite endpoint were consistent with clinical expectations, we fit hierarchical binary logistic regression models for each individual endpoint in the composite and used these models to compare risk-adjusted outcomes across hospitals stratified by their performance on the composite metric (better than expected, as expected, worse than expected). Covariates for this analysis were identical to those for the composite model. Expected rates of each complication were calculated for each group of hospitals (better, same, worse than expected) by averaging predicted risk estimates across all patients within the group. These expected rates were then compared with the observed outcomes by calculating observed to expected ratios (observed rate/expected rate).

Reliability

Aggregate reliability (i.e., the ability of the performance measure to distinguish between actual differences in performance and measurement error) was estimated using Monte Carlo simulation as described in the Supplemental Methods Appendix. Based on previous work, ideal thresholds for reliability are typically between 0.7 and 1.0 with acceptable thresholds as low as 0.5 [19].

Missing data

Covariates for case mix adjustment were generally complete with the exception of carotid stenosis (14.9%), baseline KCCQ-OS (missing in 2.6%) and gait speed (missing in 14.5%). To reduce bias and maintain consistency with the CMS NCD mandate to include patient health status, we limited the analysis to sites with ≥90% completeness for baseline KCCQ-OS and gait speed. Excluded sites were generally very similar to sites included in the analysis. Excluded sites had a numerically higher mean annualized volume but similar geographic distribution by region, urban versus rural setting and teaching versus non-teaching. Comparison of the characteristics of included and excluded sites is presented in Supplemental Table 2. Among included sites, we imputed missing covariate data to the median of continuous variables and the most common category of categorical variables. We performed sensitivity analyses regarding the inclusion of KCCQ-OS and gait speed in the risk model. This included running the model without the exclusion for >90% completeness for the KCCQ-OS and gait speed which left a total of 444 eligible sites and 112 sites remained excluded due to less than 90% completeness for the primary endpoint. This increased the number of patients from 52,561 to 85,650. We analysed the classification of sites both in the larger overall cohort of 444 sites as well as specifically looking at any reclassifications of the original 301 sites based on the model derived from the 444 sites.

Results

Patient Cohort

Between January 1, 2015 and December 31, 2017, a total of 114,121 patients underwent TAVR and were included in the TVT Registry. After excluding patients with missing data for

components of the global rank endpoint (n=1656) and patients treated at sites with <90% complete baseline data for the KCCQ or 5MWT (n=59,904 patients), the analytic cohort included 52,561 records from 301 hospitals. Baseline characteristics of those patients who were included in the primary analytic cohort and those who were excluded are summarized in Supplemental Table 3. Overall the included versus the excluded patients were very similar in baseline demographic features.

Selection of Composite Measure Endpoints

Non-fatal 30-day complications that were associated with increased risk of one year mortality included stroke (adjusted HR 2.10; 95% CI 1.65 to 2.87; p<0.001), major or life-threatening bleeding (adjusted HR 1.92; 95% CI 1.42 to 2.60, p<0.001), modified AKIN Stage III acute kidney injury (adjusted HR 1.81; 95% CI 1.38 to 2.37, p<0.001), and moderate or severe peri-valvular aortic regurgitation (adjusted HR 1.50; 95% CI 1.24 to 1.81; p<0.001). Major vascular complications (in the absence of bleeding), mild peri-valvular aortic regurgitation, and new permanent pacemaker implantation were not associated with one year mortality.

Similarly, non-fatal periprocedural complications were also associated with 1-year patient reported health status as assessed by the KCCQ-OS score. With only minor exceptions, the overall directionality of the associations were similar to the association of these complications on mortality. Any stroke (adjusted impact on 1-year KCCQ-OS -5.8 points; 95% CI -9.2 to -2.4, p<0.001) and moderate or severe peri-valvular regurgitation (adjusted KCCQ-OS impact -2.0 points; 95% CI-3.8 to -0.30, p=0.021) were independently associated with poorer adjusted KCCQ-OS at one year. Modified AKIN Stage III acute kidney injury (adjusted KCCQ-OS impact -3.3 points; 95% CI -6.8 to 0.28, p=0.07) and major or life-threatening bleed (adjusted KCCQ-OS impact 0.4 points; 95%CI -2.0 to 1.2, p=0.619) were not associated with one year

KCCQ-OS but were retained in the global rank composite measure, given their strong associations with 1-year mortality.

Based on these associations, the final global ranking for the ranked composite measure in order of decreasing severity were death, stroke, major or life-threatening/disabling bleeding, modified AKIN Stage III acute kidney injury, moderate or severe peri-valvular aortic regurgitation (Table 3). Overall, at least one of the complications occurred in 14.1% of patients, whereas 85.9% of patients had none. In the hierarchical analysis, the most common complication was major bleeding, which occurred in 5.8% of patients; this was followed by death (3.2%), moderate or severe PVL (2.5%), stroke (2.0%), and stage III AKI (0.6%). The associations between endpoints is presented in Supplemental Table 4.

Site-Specific Performance Estimates

Site-specific performance estimates according to the global rank composite metric are displayed using the site-difference methodology in Figure 1. The estimated site difference compared with the average site ranged from +6% in the best performing site to -16% in the worst performing site. Overall, better than expected site performance was observed in 25/301 (8%) of sites, as expected performance was observed in 241/301 sites (80%), and worse than expected performance was observed in 34/301 (11%) of sites. The overall risk-adjusted model is provided in Supplemental Table 5.

Outcomes of individual endpoints within composite performance categories

Adjusted observed to expected (O:E) ratios of the individual endpoint components according to the 3 levels of site performance are summarized in Table 4. Sites with better than expected performance on the global rank composite metric showed lower O:E ratios for all components of the global rank composite measure compared with the sites that performed as expected or worse

than expected. Similarly, sites with worse than expected performance on the global rank composite demonstrated consistently higher O:E ratios than the other sites. The largest differences favoring the better than expected sites were observed in the incidence of major, life threatening or disabling bleeding and moderate or severe peri-valvular leak.

Model Performance in Contemporary Data

The initial model was derived using data from January 1, 2015 to December 31, 2017. In order to assess outcomes in the most recent available data, we also ran the composite model on procedures performed between July 1, 2016 and June 30, 2019. We chose a 3-year time period for this analysis despite the fact that it overlaps partly with the derivation cohort, because the performance measure is intended to be applied to 3-year rolling data. This updated dataset included 364 eligible sites and 86,006 patients. The overall rate of mortality or major complications was 11.9%, similar to that in the development cohort (Supplemental Table 6). In the contemporary cohort, 45/365 (12%) of sites performed worse than expected, 279/364 (77%) performed as expected, and 40/364 (11%) performed better than expected.

Model Reliability

Based on Monte Carlo simulation, reliability for the overall study cohort was 0.64 indicating moderate reliability. When the analytic cohort was restricted to sites with at least 25 cases over 3 years, reliability increased to 0.65 (Supplemental Table 7), and reliability exceeded 0.7 (indicating high reliability) once site volume exceeded 100 cases over the 3 year period.

Impact of Missing KCCQ-OS and Gait Speed

The model results run without exclusions for data completeness of KCCQ-OS and gait speed were nearly identical to our main findings with 36/444 (8%) sites performed better than expected, 353/444 (80%) sites performed as expected, and 55/444 (12%) of sites performed

worse than expected. When we reexamined the classification of the 301 sites within the original cohort when using the larger 444 site model, 3 of the 301 sites were reclassified: one site was reclassified from “Better Than Expected” to “As Expected”, one site was reclassified from “As Expected” to “Better Than Expected”; and one site was reclassified from “Worse Than Expected” to “As Expected”.

Discussion

TAVR is a breakthrough technology that has been systematically evaluated in large scale randomized clinical trials that have established its safety and efficacy compared with both medical therapy and surgical aortic valve replacement in a variety of patient populations. Reports from the TVT registry have demonstrated that, on average, real-world outcomes are similar to those reported in the pivotal trials [20]. As TAVR is disseminated to an increasing number of sites and a wide range of aortic stenosis patients, however, concerns remain regarding variations in the quality of care delivered. Although greater procedural volume has been shown to correlate with better TAVR outcomes this correlation only accounts for fraction of variability in site performance [5]. As such, there is interest on the part of multiple stakeholders in developing more direct approaches to evaluating site-level performance. In the current study, we have used a novel approach to benchmarking, based on an empirically-derived ranked composite endpoint that accounts for the association between early morbid outcomes and 1-year survival and patient health status. Using this approach, we found that in current U.S. practice, there is a significant variation in performance between hospitals. This was observed in both the derivation

cohort encompassing calendar years 2015-2017 and in the more recent cohort from July 1, 2016 to June 30, 2019.

In order to understand the quality of care that patients receive, performance measurement has become a standard in cardiovascular intervention and surgery. Quality initiatives originally focused on risk-adjusted mortality after coronary artery bypass grafting (CABG) but over the past 30 years have expanded to include valvular surgery, congenital heart surgery and a wide range of interventional cardiology and electrophysiology procedures [21-24]. As mortality rates for most cardiovascular procedures have declined over the past 2-3 decades, there has been greater emphasis on understanding quality of care among the patients who survive these interventions [15]. This shifting focus has led to the development of composite measures which include both morbidity and mortality, which now form the basis of the publicly reported ratings for major cardiac procedures.[25]

To our knowledge, our approach to development and reporting of the TVT composite outcome model for TAVR is unique. Rather than relying on a traditional binary composite endpoint with endpoint selection based on expert consensus, we used a combination of novel statistical approaches and empirical data on the association between short-term procedural complications and mortality and health status outcomes to derive an outcome benchmark that is rigorous, statistically reliable, and patient-centered. Since TAVR is performed nationally in a large volumes (currently >70,000 procedures/year), the need to include surrogate structural or process measures was minimized. Based on the NCD by CMS regarding qualifications for a TAVR center, TAVR is typically performed in centers with a substantial volume of PCI, structural heart, and surgical AVR procedures, and performed by experienced multidisciplinary teams including a surgeon and an interventional cardiologist. Participation in the STS ACC TVT

registry is mandatory for reimbursement. Hence, traditional structural measures of quality such as program volume, access to advanced imaging capabilities or procedural suites, use of multidisciplinary teams and participation in outcomes registries have already been systematically incorporated into the site selection process for TAVR.

With a focus on outcomes for performance measurement, various strategies to select non-fatal outcomes were considered. These strategies included patient focus groups, expert consensus or Delphi panels, and data-driven empirical approaches. In contrast to most other procedural registries that focus exclusively on in-hospital or 30-day outcomes, the STS/ACC TVT registry, as part of its CMS mandate, captures select 1-year clinical outcomes including mortality and also patient-reported health status as measured by the KCCQ. Informing hospital-level performance measures with patient reported outcomes has potentially distinct advantages over traditional techniques by allowing the quality of an episode of care to be assessed according to the impact of the therapy and periprocedural complications on a patient's health status over the life of the patient. We selected the endpoints and their rank order using empirical data from the TVT registry on the association between the complications of interest and both 1-year mortality and health status. This represents one of the first hospital-level performance models to incorporate formal assessment of health status—both as a risk-adjustment factor as well as in the selection and ranking of non-mortality endpoints for the composite measure. Future endeavors may involve directly using one-year KCCQ as one of the primary endpoints, although one year health status may be influenced by a large number factors that are either not related to the index TAVR procedure, or out of the control of the TAVR implanting center.

The results of our analyses performed as part of model development and validation have important clinical and policy implications. Importantly, after adjusting for case mix, sites with worse than expected performance had substantially higher rates of mortality and all categories major morbidities compared with sites that performed as expected or better than expected. These findings suggest that there is a meaningful performance gap in TAVR at a modest number of US centers.

Our approach to measuring and benchmarking quality has a number of important strengths. In particular, the use of ranked endpoints in this TAVR performance measure provides a high degree of overall reliability despite the variability in site volumes. An additional strength of our approach is that the model is highly adaptable to fundamental changes in the patient population. As TAVR expands into the lower risk cohort of patients, different outcomes may become more relevant to quality of life and mortality may become less common. The technique utilized in this analysis allows for these changes to be incorporated into an updated ordinal ranking without re-visiting a Delphi-type consensus process for each iteration of the model. Finally, it is important to note that this measure has undergone a period of public commentary and sensitivity analyses to ensure internal and external validity and alignment with other quality efforts. This is in accordance with the 2019 update of the CMS TAVR NCD which specifically proposed both public reporting and linking early procedural outcomes to later patient health status as an overall measure of quality [7]. Indeed, in the most recent TAVR NCD, CMS explicitly stated that such a quality metric may eventually replace TAVR volume as one of the primary requirements for maintaining a TAVR program.

Limitations

This study should be considered in light of several limitations. Missing baseline KCCQ-12 and gait speed data significantly limited the number of sites included in this analysis. Their inclusion in the final model was based on evidence that these covariates are associated with both early and late TAVR outcomes [15,26] as well as concern that exclusion of these covariates could potentially bias the risk model against sites that treat more complex patients. In light of the challenges posed by incomplete baseline data, significant educational efforts are being made to improve data completeness, and the inclusion of these variables within the TVT registry and their collection remain mandated by CMS. Given recent experience with similar data gaps in by the STS registry (personal communication, David M. Shahian, MD 9/4/2020), we anticipate improved compliance with these data elements in future iterations of this model. Sensitivity analyses showed that removal of the 90% completeness exclusion for KCCQ-OS and gait speed, thereby allowing for a larger number of sites (444 sites), resulted in nearly identical proportions of classification into the outcome groups and nearly no re-classification within the original 301 site cohort.

Our use of 30-day (rather than in-hospital) outcomes as the basis for the model is based on the assumption that 30-day outcomes are largely dependent on the performance of the hospital and team performing the TAVR. Nonetheless, we recognize that after discharge other providers and institutions may have impacted the occurrence and management of the outcomes included in this composite measure. On the other hand, a performance measure based on only events occurring during the index hospital admission has its own limitations leading to under-reporting of complications related to the TAVR procedure and the patient selection process. Thus, the 30-day time frame was considered as a reasonable compromise.

Conclusions

Using a novel measure based on ranked 30-day outcomes, we have identified significant hospital-level variation in mortality and major complications after TAVR procedures in the United States. This metric is scheduled to be publicly reported beginning in mid-2021 as part of ongoing STS/ACC TVT Registry quality improvement initiatives. Further study is necessary to determine the impact of these measures on TAVR outcomes and structural, process-related, and technical factors associated with high and low-performing sites.

Disclosures

Dr. Desai has received Institutional research grants from Edwards Lifesciences, Medtronic, Abbott, Terumo and W.L. Gore and consulting income from Medtronic, Terumo and W.L. Gore. Dr. Cohen reports institutional research grant support from Edwards LifeSciences, Boston Scientific, Abbott, Medtronic and consulting income from Edwards LifeSciences, Boston Scientific, Abbott, Medtronic. Dr. Carroll reports institutional research grant support from Edwards LifeSciences and Medtronic and consulting income from Abbott. Dr. Vemulapalli reports grants/contracts:] from Abbott, Boston Scientific, American College of Cardiology, Society of Thoracic Surgeons, National Institutes of Health (R01 and SBIR), Food and Drug Administration (NEST) and consulting/advisory board relationships with Janssen, Boston Scientific, HeartFlow, American College of Physicians, Ikon. Dr. Forrest reports grant support for structural fellowship and proctoring/consulting fees from Edwards Lifesciences and Medtronic. Dr. Thourani reports consulting relationships with Boston Scientific, Edwards Lifesciences, Abbott Vascular, Shockwave, and Jenavalve. Dr Kirtane reports Institutional funding to Columbia University and/or Cardiovascular Research Foundation from Medtronic,

Boston Scientific, Abbott Vascular, Abiomed, CSI, CathWorks, Siemens, Philips, ReCor Medical. In addition to research grants, institutional funding includes fees paid to Columbia University and/or Cardiovascular Research Foundation for speaking engagements and/or consulting. He also reports personal consulting fees from Neurotronic. Dr. O'Neil reports research support and consulting fees from Edwards Lifesciences. Dr. Bavaria reports research grants and consulting fees from Edwards Lifesciences, Medtronic and Abbott.

Sources of Funding

This work was funded through the STS/ACC Transcatheter Valve Therapy Registry.

Supplemental Materials

Statistical Appendix

Supplemental Tables 1-7

References 27-35

References

1. Andersen HR, Knudsen LL, Hasenkam JM. Transluminal implantation of artificial heart valves. Description of a new expandable aortic valve and initial results with implantation by catheter technique in closed chest pigs. Eur Heart J. 1992;13(5):704–708
2. Leon MB. Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. N Engl J Med. 2010;363(17):1597–1607.
3. National Coverage Determination (NCD) for Transcatheter Aortic Valve Replacement (TAVR) (20.32). Centers for Medicare and Medicaid Services Website. https://www.cms.gov/medicare-coverage-database/details/ncd-details.aspx?NCDId=355 . Accessed on February 9, 2020.
4. Mack MJ, Leon MB, Thourani VH, Makkar R, Kodali SK, Russo M, Kapadia SR, Malaisrie SC, Cohen DJ, Pibarot P, Leipsic J, Hahn RT, Blanke P, Williams MR, McCabe JM, Brown DL, Babaliaros V, Goldman S, Szeto WY, Genereux P, Pershad A, Pocock SJ, Alu MC, Webb JG, Smith CR; PARTNER 3 Investigators; the PARTNER 3 Investigators. Transcatheter Aortic-Valve Replacement with a Balloon-Expandable Valve in Low-Risk Patients. N Engl J Med. 2019 May 2;380(18):1695-1705.

5. Vemulapalli S, Carroll JD, Mack MJ, Li Z, Dai D, Kosinski AS, Kumbhani DJ, Ruiz CE, Thourani VH, Hanzel G, Gleason TG, Herrmann HC, Brindis RG, Bavaria JE. Procedural Volume and Outcomes for Transcatheter Aortic-Valve Replacement. N Engl J Med. 2019 Jun 27;380(26):2541-2550.
6. O'Brien SM, Cohen DJ, Rumsfeld JS, Brennan JM, Shahian DM, Dai D, Holmes DR, Hakim RB, Thourani VH, Peterson ED, Edwards FH. 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. Circulation: Cardiovascular Quality and Outcomes. 2016 Sep;9(5):560-5.
7. National Coverage Analysis (NCA) Tracking Sheet for Transcatheter Aortic Valve Replacement (TAVR) (CAG-00430R). Centers for Medicare and Medicaid Services Website. https://www.cms.gov/medicare-coverage-database/details/nca-tracking-sheet.aspx?NCAId=293 . Accessed on February 9, 2020.
8. Mack MJ, Brennan JM, Brindis R, Carroll J, Edwards F, Grover F, Shahian D, Tuzcu EM, Peterson ED, Rumsfeld JS, Hewitt K, Shewan C, Michaels J, Christensen B, Christian A, O'Brien S, Holmes D; STS/ACC TVT Registry. Outcomes following transcatheter aortic valve replacement in the United States. JAMA. 2013 Nov 20;310(19):2069-77.
9. Kappetein AP, Head SJ, Généreux P, Piazza N, van Mieghem NM, Blackstone EH, Brott TG, Cohen DJ, Cutlip DE, van Es GA, Hahn RT, Kirtane AJ, Krucoff MW, Kodali S, Mack MJ, Mehran R, Rodés-Cabau J, Vranckx P, Webb JG, Windecker S, Serruys PW, Leon MB; Valve Academic Research Consortium (VARC)-2. Updated standardized endpoint definitions for transcatheter aortic valve implantation: the Valve Academic Research Consortium-2 consensus document (VARC-2). Eur J Cardiothorac Surg. 2012 Nov;42(5):S45-60.
10. Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. J Am Coll Cardiol. 2000 Apr;35(5):1245-55
11. Spertus J, Peterson E, Conard MW, Heidenreich PA, Krumholz HM, Jones P, McCullough PA, Pina I, Tooley J, Weintraub WS, Rumsfeld JS, Cardiovascular Outcomes Research Consortium. Am Heart J. 2005 Oct; 150(4):707-15.
12. McCarthy FH, Groeneveld PW, Kobrin D, McDermott KM, Wirtalla C, Desai ND. Effect of Clinical Trial Experience on Transcatheter Aortic Valve Replacement Outcomes. Circ Cardiovasc Interv. 2015;8:e002234.
13. Stolker JM, Spertus JA, Cohen DJ, Jones PG, Jain KK, Bamberger E, Lonergan BB, Chan PS. Re-thinking composite endpoints in clinical trials: insights from patients and trialists. Circulation 2014;130:1254-1261
14. Arnold SV, Manandhar P, Vemulapalli S, Kosinski A, Desai ND, Bavaria JE, Carroll JD, Mack MJ, Thourani VH, Cohen DJ. Impact of Short-Term Complications of TAVR on Longer-Term Outcomes: Results from the STS/ACC Transcatheter Valve Therapy Registry. Eur Heart J Qual Care Clin Outcomes. 2020 Jan 11. pii: qcaa001. doi: 10.1093/ehjqcco/qcaa001
15. Arnold SV, Spertus JA, Vemulapalli S, Dai D, O'Brien SM, Baron SJ, Kirtane AJ, Mack MJ, Green P, Reynolds MR, Rumsfeld JS, Cohen DJ. Association of Patient-Reported Health Status With Long-Term Mortality After Transcatheter

Aortic Valve Replacement: Report From the STS/ACC TVT Registry. Circ Cardiovasc Interv. 2015 Dec;8:e002875
16. Agresti A, Booth JG, Hobert JP, Caffo B. 2. Random-Effects Modeling of Categorical Response Data. Sociological Methodology. 2000 Aug;30(1):27-80
17. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. Eur Heart J. 2012 Jan;33(2):176-82.
18. Péron J, Roy P, Ozenne B, Roche L, Buyse M. The Net Chance of a Longer Survival as a Patient-Oriented Measure of Treatment Benefit in Randomized Clinical Trials. *JAMA Oncol*. 2016;2(7):901-905. doi:10.1001/jamaoncol.2015.6359
19. Adams JL, Mehrotra A, Thomas JW, et al. Physician cost profiling--reliability and risk of misclassification. New Engl J Med. 2010; 362:1014–1021.
20. Grover FL, Vemulapalli S, Carroll JD, Edwards FH, Mack MJ, Thourani VH,Brindis RG, Shahian DM, Ruiz CE, Jacobs JP, Hanzel G, Bavaria JE, Tuzcu EM,Peterson ED, Fitzgerald S, Kourtis M, Michaels J, Christensen B, Seward WF, Hewitt K, Holmes DR Jr; STS/ACC TVT Registry. 2016 Annual Report of The Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy Registry. J Am Coll Cardiol. 2017 Mar 14;69(10):1215-1230.
21. Jacobs JP, Lacour-Gayet FG, Jacobs ML, Clarke DR, Tchervenkov CI, Gaynor JW, Spray TL, Maruszewski B, Stellin G, Gould J, Dokholyan RS, Peterson ED, Elliott MJ, Mavroudis C. Initial application in the STS congenital database of complexity adjustment to evaluate surgical case mix and results. Ann Thorac Surg. 2005 May;79(5):1635-49;
22. O'Brien SM, Shahian DM, Filardo G, Ferraris VA, Haan CK, Rich JB, Normand SL, DeLong ER, Shewan CM, Dokholyan RS, Peterson ED, Edwards FH, Anderson RP; Society of Thoracic Surgeons Quality Measurement Task Force. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2--isolated valve surgery. Ann Thorac Surg. 2009 Jul;88(1 Suppl):S23-42.
23. Shahian DM, He X, Jacobs JP, Rankin JS, Welke KF, Edwards FH, Filardo G, Fazzalari FL, Furnary A, Kurlansky PA, Brennan JM, Badhwar V, O'Brien SM. The STS AVR+CABG composite score: a report of the STS Quality Measurement Task Force. Ann Thorac Surg. 2014 May;97(5):1604-9.
24. Acharya T, Salisbury AC, Spertus JA, Kennedy KF, Bhullar A, Reddy HKK, Joshi BK, Ambrose JA. In-Hospital Outcomes of Percutaneous Coronary Intervention in America's Safety Net: Insights From the NCDR Cath-PCI Registry. JACC Cardiovasc Interv. 2017 Aug 14;10(15):1475-1485.
25. Shahian DM, He X, Jacobs JP, et al. The Society of Thoracic Surgeons Isolated Aortic Valve Replacement (AVR) Composite Score: a report of the STS Quality Measurement Task Force. Ann Thorac Surg. 2012;94(6):2166-2171. doi:10.1016/j.athoracsur.2012.08.120
26. Arnold SV, O'Brien SM, Vemulapalli S, Cohen DJ, Stebbins A, Brennan JM, Shahian DM, Hakim R, Grover FL, Holmes DR, Thourain V, Peterson ED, Edwards FH. Inclusion of functional status measures in the risk adjustment of 30-day mortality after transcatheter aortic valve replacement: A report from the STS/ACC TVT Registry. *J Am Coll Cardiol Intv* 2018; 11:581-89

27. Agresti A, Booth JG, Hobert JP, Caffo B. 2. Random-Effects Modeling of Categorical Response Data. Sociological Methodology. 2000 Aug;30(1):27-80.
28. Harrell Jr FE. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer; 2015 Aug 14.
29. Daniels MJ, Gatsonis C. Hierarchical polytomous regression models with applications to health services research. Statistics in Medicine. 1997 Oct 30;16(20):2311-25.
30. Daniels MJ, Gatsonis C. Hierarchical generalized linear models in the analysis of variations in health care utilization. Journal of the American Statistical Association. 1999 Mar 1;94(445):29-42.
31. Krumholz HM, Brindis RG, Brush JE, Cohen DJ, Epstein AJ, Furie K, Howard G, Peterson ED, Rathore SS, Smith Jr SC, Spertus JA. Standards for statistical models used for public reporting of health outcomes: an American Heart Association Scientific Statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group: cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. Circulation. 2006 Jan 24;113(3):456-62.
32. Dimick JB, Ghaferi AA, Osborne NH, Ko CY, Hall BL. Reliability adjustment for reporting hospital outcomes with surgery. Annals of surgery. 2012 Apr 1;255(4):703-7.
33. Ten Have TR, Localio AR. Empirical Bayes estimation of random effects parameters in mixed effects logistic regression models. Biometrics. 1999 Dec;55(4):1022-9.
34. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. Eur Heart J. 2012 Jan;33(2):176-82.
35. Péron J, Roy P, Ozenne B, Roche L, Buyse M. The net chance of a longer survival as a patient-oriented measure of treatment benefit in randomized clinical trials. JAMA oncology. 2016 Jul 1;2(7):901-5

Table 1. Baseline Patient Characteristics

Characteristic	Overall Population (n=52,561)	Missing (%)
Age, yrs	82 (76, 87)	0.0
Male Sex, %	53.5	0.0
Diabetes mellitus, %	38.4	0.1
Current smoker, %	5.8	<0.1
Currently on dialysis, %	3.7	0.1
GFR, ml/min/1.73 m^2	63 (48, 78)	0.3
LVEF, %	58 (50, 60)	0.6
Prior MI, %	23.2	0.2
Prior pacemaker, %	14.3	0.1
Prior PCI, %	34.4	0.1
Prior CABG, %	23.4	0.1
Prior aortic valve procedure, %	11.9	0.1
Prior non-aortic valve procedure, %	2.2	0.2
NYHA Class IV, %	14.9	0.5
Atrial fibrillation/flutter, %	39.3	0.1
Conduction defect, %	38.2	0.5
Prior stroke or TIA, %	18.5	<0.1
Carotid stenosis, %	22.0	14.9
Peripheral arterial disease, %	28.3	0.1
Severe chronic lung disease, %	11.9	0.5
Home oxygen, %	10.2	<0.1
Hostile chest, %	7.0	<0.1
Porcelain aorta, %	4.0	0.1
Non-femoral access, %	7.6	0.3
Acuity		0.0
Elective, %	90.8	
Urgent, %	6.1	
Shock/Inotrope/Support device, %	2.7	
Emergency/Salvage/Cardiac arrest, %	0.4	
KCCQ-OS	43 (26, 63)	2.6
5MWT, seconds	7 (6, 10)	14.5

Continuous variables are described as medians with 25th and 75th percentiles

GFR = glomerular filtration rate; LVEF = left ventricular ejection fraction; MI = myocardial infarction; PCI = percutaneous coronary intervention; CABG = coronary artery bypass graft surgery; NYHA = New York Heart Association; TIA = transient ischemic attack; KCCQ-OS = Kansas City Cardiomyopathy Questionnaire, Overall Summary Score; 5MWT = 5-meter walk time

Table 2. Definitions of Complications Included in the Global Rank Composite Measure

Endpoint Component	Definition
Mortality	All-cause mortality
Stroke	Stroke defined by Valve Academic Research Consortium 2 (VARC-2) criteria as an acute episode of a focal or global neurological deficit with duration of ≥24 hours caused by ischemic, hemorrhagic, or undetermined etiology and confirmed by neurological or neurosurgical specialist or neuroimaging. Strokes were directly adjudicated by DCRI physicians. Transient ischemic attack (TIA) was excluded.
Bleeding Complication	Major Bleeding or Life-threatening or disabling bleeding was defined as Bleeding Academic Research Consortium(BARC) 3a,b,or c or BARC 5
Acute Kidney Injury	Acute Kidney Injury(AKI) was defined according to a modification of Class III AKIN classification: Increase in serum creatinine to ≥300% (>3 × increase compared with baseline) OR serum creatinine of ≥4.0 mg/dL (≥354 mmol/L) with an acute increase of at least 0.5 mg/dL (44 mmol/L)
Moderate or Severe Peri-valvular Leak	Moderate aortic insufficiency or regurgitation is defined as the following: Qualitative Measurements: Angiographic grade of 2+; Color Doppler jet width greater than mild but no signs of severe aortic regurgitation (insufficiency); Dopplar vena contracta width 0.3-0.6 cm; Quantitative Measures (cath or echo) Regurgitant volume 30-59 ml/beat; Regurgitant fraction 30-49%; Regurgitant orifice area 0.10-0.29 cm(2) Severe aortic insufficiency or regurgitation is defined as the following: Qualitative Measurements: Angiographic grade of 3-4+; Color Doppler jet width (Central jet) >65% of LVOT; Doppler vena contracta width >0.6 cm;. Quantitative Measures (cath or echo) Regurgitant volume >=60 ml/beat; Regurgitant fraction >=50%; Regurgitant orifice area >=0.30 cm(2) Additional essential criteria: Left ventricular size is increased

Table 3. Frequency of Global Ranking Categories in Study Cohort

Endpoint Ranking	Number	Percent
1 = Death	1671	3.2%
2 = Stroke	1077	2.0%
3 = Bleeding Complication	3024	5.8%
4 = Acute Kidney Injury	336	0.6%
5 = Moderate or Severe Peri-valvular leak	1304	2.5%
6 = None of the above	45149	85.9%

(Total N = 52,561)

Table 4. Adjusted Rates of 30-day complications according to overall site performance

	O/E Ratio (Observed* / Expected*)		
Variables	**Better than Expected (Sites = 25) (N = 7993)**	**As Expected (Sites = 242) (N = 37473)**	**Worse than Expected (Sites = 34) (N = 7095)**
Death	0.71 (2.25% / 3.16%)	1.01 (3.21% / 3.17%)	1.25 (4.06% / 3.26%)
Stroke	0.73 (1.74% / 2.38%)	1.03 (2.49% / 2.41%)	1.29 (3.16% / 2.44%)
Major or Life Threatening /Disabling Bleed	0.45 (2.84% / 6.30%)	1.02 (6.48% / 6.33%)	2.13 (13.6% / 6.38%)
Acute Kidney Injury or New Dialysis	0.67 (0.83% / 1.23%)	1.12 (1.34% / 1.20%)	1.17 (1.44% / 1.23%)
Moderate or Severe Peri-valvular Regurgitation	0.77 (1.88% / 2.45%)	1.19 (2.71% / 2.28%)	2.00 (4.76% / 2.38%)

O/E ratio = observed to expected ratio

*Proportions represent observed rates and expected rates (based on risk-adjusted models for each individual complication

Figure Legends

Figure 1. Caterpillar Plot of Site-Specific Outcomes