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

Sean M. O'Brien, PhD; David J. Cohen, MD; John S. Rumsfeld, MD, PhD; J. Matthew Brennan, MD, MPH; David M. Shahian, MD; David Dai, PhD; David R. Holmes, MD; Rosemarie B. Hakim, PhD; Vinod H. Thourani, MD; Eric D. Peterson, MD, MPH; Fred H. Edwards, MD; for the STS/ACC TVT Registry

Background—The use of transcatheter aortic valve replacement (TAVR) to treat aortic stenosis in the United States is growing, yet little is known about the variation in procedural outcomes in community practice. We developed a TAVR in-hospital mortality risk model and used it to quantify variation in mortality rates across United States (US) TAVR centers.

Methods and Results—We analyzed data from 22 248 TAVR procedures performed at 318 sites participating in the Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy Registry (November 2011 to October 2014). A Bayesian hierarchical model was developed to estimate hospital-specific risk-adjusted mortality rates adjusting for 40 patient baseline factors. A total of 1130 in-hospital deaths (5.1%) were observed. Reliability-adjusted risk-adjusted mortality rate estimates ranged from 3.4% to 7.7% with an interquartile range of 4.8% to 5.4%. A patient's predicted odds of dying was 80% higher if treated by a hospital 1 standard deviation above the mean compared with a hospital 1 standard deviation below the mean (odds ratio =1.8; 95% credible interval, 1.4%–2.2%).

Conclusions—Risk modeling of TAVR in-hospital mortality revealed variation in risk-adjusted mortality rates during the US early commercial experience. Transcatheter Valve Therapy Registry analyses using this model will support research, feedback reporting, and the identification of factors associated with quality. **(*Circ Cardiovasc Qual Outcomes*. 2016;9:560-565. DOI: 10.1161/CIRCOUTCOMES.116.002756.)**

Key Words: aortic stenosis ■ case-mix adjustment ■ outcomes analysis ■ outcomes research ■ transcatheter aortic valve replacement

After regulatory approval by the United States Food and Drug Administration in 2011, transcatheter aortic valve replacement (TAVR) has been used with increasing frequency to treat patients with severe aortic stenosis who are at high risk of mortality or morbidity with conventional surgical aortic valve replacement.[1,2] As a result, the use of TAVR has expanded rapidly throughout the United States to centers with varying degrees of experience using TAVR. Recent studies from the Society of Thoracic Surgeons (STS)/American College of Cardiology (ACC) Transcatheter Valve Therapy (TVT) Registry have documented 30-day and 1-year TAVR outcomes in the United States,[1,3] but the extent to which these outcomes vary across hospitals has not been reported.

Editorial, see p 498

For many procedures performed by interventional cardiologists and cardiac surgeons, risk-adjusted outcome metrics have been developed, applied, and validated using large clinical registries, such as the STS Adult Cardiac Surgery Database and the ACC National Cardiovascular Data Registry. For example, models to assess institutional performance have been developed for coronary artery bypass grafting,[4] valve

Received February 19, 2016; accepted August 2, 2016.

From the Duke University Medical Center, Durham, NC (S.M.O., J.M.B., D.D., E.D.P.); Saint Luke's Mid America Heart Institute, Kansas City, MO (D.J.C.); Denver VA Medical Center, CO (J.S.R.); Massachusetts General Hospital, Boston (D.M.S.); Mayo Clinic, Rochester, MN (D.R.H.); Centers for Medicare & Medicaid Services, Baltimore, MD (R.B.H.); Emory University School of Medicine, Atlanta, GA (V.H.T.); and University of Florida, Jacksonville (F.H.E.).

This article was handled independently by Karen E. Joynt, MD, MPH, as a Guest Editor. The editors had no role in the evaluation or in the decision about its acceptance.

The views expressed in this article represent those of the authors and do not necessarily represent the official views of the American College of Cardiology or the Society of Thoracic Surgeons.

The Data Supplement is available at http://circoutcomes.ahajournals.org/lookup/suppl/doi:10.1161/CIRCOUTCOMES.116.002756/-/DC1.

Correspondence to Sean M. O'Brien, PhD, Duke Clinical Research Institute, Duke Box 2721, Durham, NC 27715. E-mail sean.m.obrien@duke.edu

WHAT IS KNOWN

- Since its approval by the United States (US) Food and Drug Administration in 2011, transcatheter aortic valve replacement (TAVR) has been used with increasing frequency to treat patients with severe aortic stenosis who are at high risk of mortality and morbidity with conventional surgical aortic valve replacement.
- A risk model customized for TAVR institutional performance assessment was not previously available.

WHAT THIS STUDY ADDS

- Risk modeling of TAVR in-hospital mortality revealed variation in risk-adjusted mortality rates during the early US commercial experience.
- Transcatheter Valve Therapy Registry analyses using this model will support research, feedback reporting, and the identification of factors associated with quality.

surgery,[5] and percutaneous coronary intervention.[6,7] Recent studies have developed parsimonious prediction models to facilitate patient selection and counseling for TAVR.[8–10] However, a TAVR risk model customized for institutional performance assessment was not previously available.

This study was conducted using data from the STS/ACC TVT Registry. Our specific goals were to (1) develop a TAVR in-hospital mortality risk model suitable for research and feedback reporting in the TVT Registry and (2) characterize variation in mortality across centers during the US early commercial experience.

Methods

The STS/ACC TVT Registry

The STS/ACC TVT Registry was created in 2011 to monitor patient safety and real-world outcomes of emerging treatments for valve disease. The registry's major goals are to monitor outcomes of transcatheter valve procedures in real-world patients, assess the learning curve among centers beginning to adopt the technology, and provide a ready-made infrastructure for research and safety surveillance. As a condition for Medicare coverage, all US TAVR hospitals are required to collect data on consecutively enrolled patients receiving a Food and Drug Administration–approved transcatheter heart valve. Centers participating in the TVT Registry receive quarterly reports comparing each center's case mix, practice patterns, and outcomes to the national experience. Performance measures developed within the TVT Registry are intended to promote quality improvement, identify characteristics of high-performing centers, and inform the design of interventions to improve patient outcomes.

Cohort Selection

We included all TVT records for patients undergoing TAVR with a commercially approved device between November 1, 2011, and October 31, 2014. From a starting population of 22 271 TAVR cases, we excluded 23 cases (0.1%) with missing data for discharge mortality status. TAVR procedures aborted intraoperatively were included in this cohort in keeping with the analysis principle of intention to treat.[11,12]

End Point

We used in-hospital rather than 30-day mortality status as the primary outcome measure because in-hospital mortality status was reported in 99.9% of records, whereas 30-day mortality status was missing in 20% at the time of model development initiation.

Descriptive Analyses

Patient baseline characteristics were summarized for the overall cohort, as well as for each individual center, to assess variation in case mix across hospitals. Baseline summaries included percentages and median values as appropriate. For describing center-specific unadjusted mortality rates, each center's unadjusted (observed) discharge mortality rate was calculated and plotted in relation to the center's number of eligible cases (ie, the center's sample size). Lines depicting exact 95% binomial prediction limits were overlaid to show the expected range of sampling variation that would normally occur under the hypothesis of no true center-level variation in mortality.[13]

Covariates for Case-Mix Adjustment

Details of the covariate selection process are described in the Methods in the Data Supplement. Briefly, our multivariable analysis was designed to estimate center-specific mortality rates after adjusting for patient baseline factors and to determine the amount of variation in these rates that is because of true hospital-level differences (ie, signal) as opposed to random error (ie, noise). Accordingly, covariates for multivariable analysis were selected with the goal of minimizing potential confounding by case mix and without regard for the model's parsimony (Table I in the Data Supplement). A total of 40 baseline factors were selected. The main considerations for covariate selection included clinical relevance, variation across hospitals, and data quality. Frailty, as assessed by the 5-m walk assessment, and cardiovascular-specific functional status, as assessed by the Kansas City Cardiomyopathy Questionnaire, are known to be prognostically significant,[14,15] but were missing too frequently (60% and 61%, respectively) to be included in the model. Remaining model covariates were highly complete, with most covariates having <1% missing data and only one covariate having >3% missing data (number of prior cardiac operations; missing 3.6%). In the rare case of missing data, unknown values were imputed to the most common category of categorical variables and to the median or subgroup-specific median of continuous variables. More computationally intensive missing data strategies, such as multiple imputation, were not used for this analysis because of the low rate of missing data and because it would be impractical to implement them in combination with the computationally intensive Markov Chain Monte Carlo estimation procedure described below.

Multivariable Analysis

For modeling procedural acuity, a categorical variable was created by combining the procedure status field with 4 relatively rare high-risk preprocedure conditions: (1) cardiac arrest within 24 hours; (2) cardiac shock within 24 hours; (3) inotropes; and (4) mechanical assist device. Patients with none of these high-risk conditions were assigned to category 1 if status was elective or category 2 if status was urgent. Remaining patients were assigned to category 3 if status was elective or urgent and there was no cardiac arrest within 24 hours or category 4, otherwise.

To estimate center-specific mortality rates, we fit a hierarchical logistic regression model with adjustment for 40 prespecified patient baseline factors and center-specific random intercepts (Table II in the Data Supplement). Before making inferences about hospital performance, the model's fit to the data was assessed internally. Calibration was assessed by comparing observed versus expected mortality rates overall and within subgroups, based on deciles of predicted risk. Discrimination was assessed by the C statistic. After confirming satisfactory internal calibration and discrimination, the model was

Table. Frequency of Patient Baseline Factors, Overall and By Center

Risk Factor	All Centers	Range Across Centers*
Age, y (median)	84.0	79.0–86.0
Age, 85+ y	45.1%	23.4%–68.2%
BSA, m^2 (median)	1.8	1.7–1.9
BSA, ≥2.20 m^2	7.2%	1.0%–14.5%
Sex, female	50.3%	33.9%–63.2%
Nonwhite or Hispanic	9.1%	0.0%–69.6%
GFR <45 mL/min/1.73 m^2 or dialysis	27.9%	8.3%–45.0%
Renal function, dialysis	4.1%	0.0%–11.6%
Ejection fraction (median)	57.0	53.5–64.0
Ejection fraction, <35%	14.5%	1.9%–21.8%
Hemoglobin, g/dL (median)	11.7	10.6–12.7
Hemoglobin, <10 g/dL	17.5%	3.9%–33.1%
Platelet count, μL (median)	190 000	168 000–212 000
Platelet count, <100k μL	5.7%	1.5%–14.5%
Procedure date, after March 31, 2013†	69.3%	37.4%–99.3%
Left main disease ≥50%	10.9%	3.6%–20.3%
Proximal LAD ≥70%	20.2%	4.7%–38.8%
Prior MI	25.4%	9.0%–51.8%
Endocarditis	0.9%	0.0%–6.0%
Prior TIA or stroke	19.0%	7.9%–31.0%
Carotid stenosis	16.4%	3.1%–43.3%
Prior PAD	32.0%	11.1%–71.0%
Smoker	5.3%	0.0%–16.6%
Diabetes mellitus	37.1%	22.8%–52.1%
NYHA Class IV	21.9%	3.3%–96.9%
Atrial fibrillation or flutter	40.8%	26.7%–55.5%
Conduction defect	30.8%	0.0%–67.3%
Severe chronic lung disease	13.9%	2.2%–46.0%
Home oxygen	13.6%	3.7%–33.1%
Hostile chest	8.7%	0.0%–34.4%
Porcelain aorta	7.1%	0.0%–23.7%
Access site, femoral	58.8%	24.3%–93.3%
Previous ICD	4.3%	0.0%–10.2%
Prior PCI	35.7%	18.6%–64.9%
Prior CABG	31.6%	14.1%–48.9%
Prior aortic procedure	16.4%	0.6%–70.2%
Prior nonaortic procedure	2.8%	0.0%–10.0%
Aortic etiology, degenerative	94.5%	30.5%–100.0%
Aortic valve morphology, tricuspid	91.9%	0.0%–100.0%
Aortic valve morphology, other‡	2.0%	0.0%–20.8%
Aortic insufficiency, at least moderate	20.4%	6.8%–46.5%
Mitral insufficiency, at least moderate	31.1%	12.3%–67.6%
Tricuspid insufficiency, at least moderate	24.2%	6.5%–46.2%
Preoperative inotropes	2.9%	0.0%–59.1%
Acuity, category 2	8.8%	0.0%–74.4%
Acuity, category 3	4.0%	0.0%–58.6%
Acuity, category 4	0.5%	0.0%–3.4%

BSA indicates body surface area; CABG, coronary artery bypass grafting; GFR, glomerular filtration rate; ICD, internal cardioverter defibrillator; LAD, left anterior descending; MI, myocardial infarction; NYHA, New York Heart Association; PAD, peripheral artery disease; PCI, percutaneous coronary intervention; and TIA, transient ischemic attack.

*Among centers with at least 100 eligible cases.

†Approximate midpoint of study period.

‡Includes unicuspid, bicuspid, quadracuspid, and uncertain morphology.

re-estimated using a 70% random sample of records and assessed for calibration and discrimination in the remaining 30% of records.

A center's standardized mortality ratio was defined as the ratio of the center's true mortality rate according to the model divided by the mortality rate that would be expected for a hypothetical center with the same case mix and an intercept parameter equal to the population average (see Methods in the Data Supplement for details). A hospital's risk-adjusted mortality rate (RAMR) was defined by the formula RAMR=(standardized mortality ratio)×(overall mortality rate), where "overall mortality rate" denotes the observed discharge mortality rate among all patients in the overall study cohort. Definitions of the variables used in the multivariable analysis can be found in Table III in the Data Supplement.

Estimation

Model parameters were estimated in a Bayesian statistical framework by specifying a prior probability distribution for unknown parameters and using Markov Chain Monte Carlo simulations for inference. Briefly, Bayesian inference uses the language of probability to express beliefs about clinically interesting hypotheses and quantities. The output of a Bayesian analysis is a probability distribution describing the most likely numeric estimates of unknown model parameters. Markov Chain Monte Carlo simulations are used to generate representative samples of parameter values, which are then analyzed to create appropriate estimates and summary measures. Advantages of fully Bayesian estimation include the ability to perform inference about complex functions of unknown quantities (eg, the variance of RAMRs across hospitals) and the ability to calculate the probability of any clinically interesting hypothesis (eg, the probability that a given center's RAMR is greater than the national average). Unlike frequentist confidence intervals, Bayesian interval estimates (known as credible intervals [CrIs]) have an intuitively direct interpretation as an interval containing the true value with a specified probability (eg, 95%).

Bayesian estimators of hospital-specific mortality incorporate a reliability adjustment (also known as shrinkage) that diminishes the apparent variation in mortality rates across hospitals.[16–18] Although such estimators are optimal for the purpose of estimating each hospital's unknown mortality rate,[19] the distribution of reliability-adjusted mortality rates may severely underestimate the true amount of signal variation.[20–22] Therefore, to appropriately characterize the magnitude of true signal variation, the following approach was adopted. First, using the model's estimated variance parameter, we

calculated the odds ratio (OR) comparing a patient's predicted odds of dying if treated by a hospital 1 standard deviation above the mean relative to a hospital 1 standard deviation below the mean. Second, we calculated a Bayesian estimate of the histogram of unknown RAMRs across TAVR hospitals.[20–22] Numeric summaries of the histogram of RAMRs were estimated by their posterior mean and reported with 95% Bayesian CrIs (see Methods in the Data Supplement for details).[23,24]

Institutional Review Board

This study was approved by the Duke Institutional Review Board (Durham, NC).

Results

Study Cohort

The final study population included 22 248 index TAVR procedures from 318 US hospitals with a median of 58 (interquartile range [IQR] 22–96) eligible cases per hospital. The median age was 84 years, and 50% were female. Patient characteristics varied substantially across hospitals. Even after excluding hospitals with <100 cases, there was more than a 5-fold difference across hospitals for the majority of baseline factors examined (Table). For example, the proportion of patients with a prior aortic procedure varied >100-fold, ranging from 0.6% to 70.2% across hospitals. New York Heart Association Class IV heart failure varied 29-fold, ranging from 3.3% to 96.9% across hospitals. Among centers with at least 100 cases and at least 30 cases with nonmissing 5-m walk assessment, the proportion of patients requiring >10 seconds to walk 5 m varied from 6% to 81% across centers.

Unadjusted Mortality

There were a total of 1130 in-hospital deaths during the study period (5.1%). Hospital-specific observed mortality rates ranged from 0% to 33.3% (IQR 2.4%–7.1%). For the vast majority of centers (96%), the observed mortality rate fell within the expected range of normal sampling variation (Figure 1).

Multivariable Analysis

The multivariable association between each model covariate and in-hospital mortality based on the final model is summarized in Table II in the Data Supplement. Binary factors with the largest estimated ORs included dialysis (OR=2.4), nonfemoral access (OR=1.9), and severe chronic lung disease (OR=1.5). As expected, mortality risk was highest in the highest category of procedural acuity (OR=3.7).

Calibration plots demonstrated high agreement between observed and expected mortality rates, both overall and within prespecified subgroups (Figures I and II in the Data Supplement). The C statistic was 0.71 in the overall sample and 0.67 in the 30% cross-validation sample (Table IV in the Data Supplement).

A patient's predicted odds of dying was 80% higher if treated by a hospital 1 standard deviation above the mean compared with 1 standard deviation below the mean (OR=1.8; 95% CrI, 1.4%–2.2%). Figure 2 summarizes the estimated

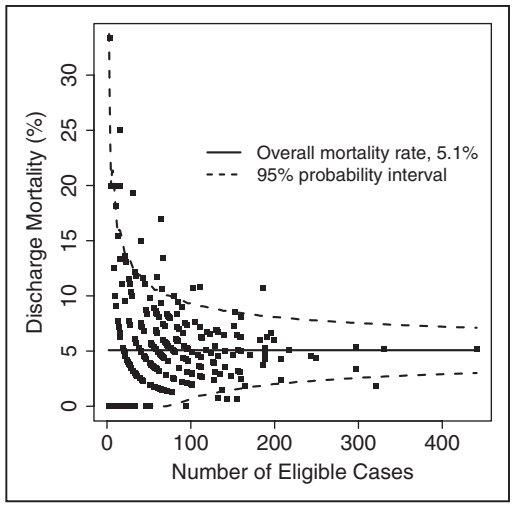


Figure 1. Hospital-specific observed mortality rates. Funnel plot of hospital-specific observed mortality rates with 95% prediction limits.

distribution of hospital-specific true RAMRs in the form of a histogram. The estimated IQR was 4.3% to 6.1%. The estimated average RAMR ranged from 3.2% to 8.1% among hospitals in the lowest and highest mortality deciles, respectively (risk difference=5.0% [95% CrI, 3.0%–6.9%]; risk ratio=2.6 [95% CrI, 1.8%–3.5%]).

Reliability-adjusted estimates of individual hospital-specific RAMRs ranged from 3.4% to 7.7% with an IQR of 4.8% to 5.4%. Only one hospital had an RAMR statistically distinguishable from the TVT average (Figure 3). As noted in the Methods section, the variation in these reliability-adjusted estimates is a downwardly biased estimate of the true variation in RAMRs across hospitals. As a result, the IQR reported here is slightly narrower than the correctly estimated IQR of true RAMRs, as reported in Figure 2.

Discussion

In this report of the US early commercial experience with TAVR, we found significant hospital-level variation in risk-adjusted TAVR mortality rates among US hospitals. These findings support the hypothesis that institutional factors may play a role in affecting patient outcomes. As TAVR continues to evolve, risk models developed in the TAVR population will support research, feedback reporting, and the identification of factors associated with quality.

Nevertheless, our results also demonstrate the challenges in outcomes performance evaluation at this early stage in the procedure's adoption curve. Given the low number of cases performed at many sites, measured mortality rates are expected to vary substantially by chance alone. Consequently, despite evidence of true signal variation, only one hospital had an RAMR statistically distinguishable from the TVT average. Estimating the amount of residual variation not caused by chance required a relatively complex statistical strategy.

Overall, the early message of the US TAVR experience is positive. Our current findings indicate that many community sites are achieving procedural outcomes similar to or better

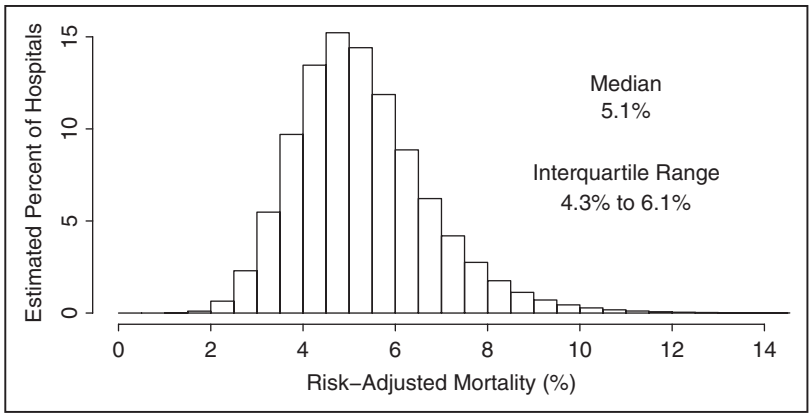


Figure 2. Distribution of true RAMR. Estimated distribution of true RAMR across hospitals. RAMR indicates risk-adjusted mortality rate.

than those seen in the pivotal clinical trial experience. However, there is evidence of site-to-site variation both in the types of patients treated and in adjusted acute outcomes. Understanding what site characteristics and processes may be associated with excellent or poor outcomes will be an important direction for future quality improvement and research efforts.

The STS and ACC have long recognized and emphasized the importance of continual assessment of TAVR outcomes, but outcomes analysis has been limited by the lack of a customized TAVR risk-adjustment methodology.[25] To adjust for case mix in this study, we used hierarchical logistic regression and included all baseline factors with a known or suspected association with mortality as covariates. Before estimating center-level performance, the model was empirically tested to evaluate internal calibration and to assess the impact of various methodological decisions. For estimation, we used a fully Bayesian approach that permits making intuitively appealing probability statements about each hospital's performance. The Bayesian method also appropriately accounts for overfitting by allowing uncertainty in the estimation of regression coefficients to be appropriately reflected in wider confidence intervals. To further enhance validity, the models used in this project were developed by a panel of statistical and clinical experts and were circulated for public comment on the ACC's mailing list and website. Public comments were carefully considered in the approach to model development.

The risk adjustment techniques used in this study were developed not only for quantifying variation across TAVR hospitals, but also for use in a variety of future planned quality improvement and research efforts. Based on this article's risk-adjustment method, centers in the TVT Registry will receive quarterly feedback reports comparing their RAMR to the national experience. Additional performance measures for nonfatal end points (eg, stroke and bleeding) are planned for development and will provide complementary information for a more comprehensive assessment of TAVR outcomes. One-year outcomes are also being tracked in the TVT Registry[3] and will be the focus of future risk modeling efforts.

Limitations

This study has several limitations. First, although the model adjusted for a large number of patient factors, several known or suspected risk factors were not included. Variables such as 5-m walk assessment and baseline functional status are known to be prognostically significant for mortality after TAVR,[14,15] but were missing in >50% of records in this early stage of national data collection. In centers performing the 5-m walk assessment, the distribution of walk times was highly variable across centers, suggesting a potential for bias from differential case mix. To the extent that frailty is correlated with other variables in the risk-adjustment model, the potential for bias is lessened. Second, although our analysis documented significant variation in mortality, it did not identify specific factors explaining this variation. The effects of provider experience and volume on mortality were not analyzed in the current study, but will be an important topic for future study. Third, we analyzed a single-end point discharge mortality and did not analyze long-term mortality or nonfatal end points. The development of TAVR-specific risk models for these other end points is a high STS and ACC priority. Fourth, generalizability is impacted by the exclusion of noncommercial cases and the availability of only one Food and Drug Administration–approved device during the majority of the study period. Finally, for reasons related to statistical power and sample size, we were unable to determine whether between-center variation in mortality has increased or decreased over the study period.

Conclusions

Risk adjustment of TAVR in-hospital mortality revealed center-level variation during the US early commercial experience. TVT Registry analyses using this model will support research, feedback reporting, and the identification of factors associated with quality.

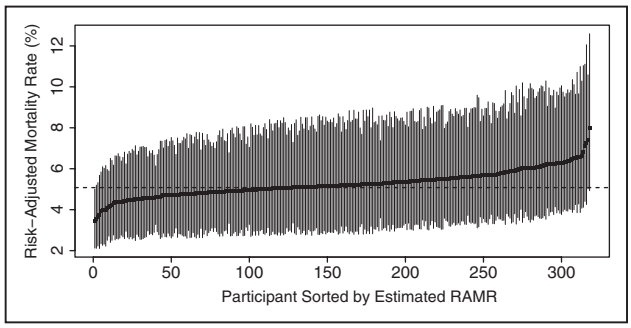


Figure 3. Hospital-specific risk-adjusted mortality rates (RAMR) estimates. Hospital-specific RAMR estimates with 95% Bayesian credible intervals. RAMR indicates risk-adjusted mortality rate.

Acknowledgments

We thank Erin Hanley, MS, for her editorial contributions to this article. Ms Hanley did not receive compensation for her contributions, apart from her employment at the institution where this study was conducted.

Sources of Funding

The Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy Registry is an initiative of the Society of Thoracic Surgeons and the American College of Cardiology Foundation. This research was supported by the American College of Cardiology Foundation's National Cardiovascular Data Registry and the Society of Thoracic Surgeons.

Disclosures

None.

References

1. 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;310:2069–2077.
2. Osnabrugge RL, Mylotte D, Head SJ, Van Mieghem NM, Nkomo VT, LeReun CM, Bogers AJ, Piazza N, Kappetein AP. Aortic stenosis in the elderly: disease prevalence and number of candidates for transcatheter aortic valve replacement: a meta-analysis and modeling study. *J Am Coll Cardiol*. 2013;62:1002–1012. doi: 10.1016/j.jacc.2013.05.015.
3. Holmes DR Jr, Brennan JM, Rumsfeld JS, Dai D, O'Brien SM, Vemulapalli S, Edwards FH, Carroll J, Shahian D, Grover F, Tuzcu EM, Peterson ED, Brindis RG, Mack MJ; STS/ACC TVT Registry. Clinical outcomes at 1 year following transcatheter aortic valve replacement. *JAMA*. 2015;313:1019–1028.
4. Shahian DM, O'Brien SM, 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 1—coronary artery bypass grafting surgery. *Ann Thorac Surg*. 2009;88:S2–S22.
5. 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;88:S23–S42.
6. Peterson ED, Dai D, DeLong ER, Brennan JM, Singh M, Rao SV, Shaw RE, Roe MT, Ho KK, Klein LW, Krone RJ, Weintraub WS, Brindis RG, Rumsfeld JS, Spertus JA; NCDR Registry Participants. Contemporary mortality risk prediction for percutaneous coronary intervention: results from 588,398 procedures in the National Cardiovascular Data Registry. *J Am Coll Cardiol*. 2010;55:1923–1932.
7. National Cardiovascular Data Registry. PCI readmission measure. National Cardiovascular Data Registry web site. http://www.ncdr.com/WebNCDR/analytics/pcireadmissionmeasure. Accessed July 1, 2016.
8. Iung B, Laouénan C, Himbert D, Eltchaninoff H, Chevreul K, Donzeau-Gouge P, Fajadet J, Leprince P, Leguerrier A, Lièvre M, Prat A, Teiger E, Laskar M, Vahanian A, Gilard M; FRANCE 2 Investigators. Predictive factors of early mortality after transcatheter aortic valve implantation: individual risk assessment using a simple score. *Heart*. 2014;100:1016–1023. doi: 10.1136/heartjnl-2013-305314.
9. Arnold SV, Reynolds MR, Lei Y, Magnuson EA, Kirtane AJ, Kodali SK, Zajarias A, Thourani VH, Green P, Rodés-Cabau J, Beohar N, Mack MJ, Leon MB, Cohen DJ; PARTNER Investigators. Predictors of poor outcomes after transcatheter aortic valve replacement: results from the PARTNER (Placement of Aortic Transcatheter Valve) trial. *Circulation*. 2014;129:2682–2690. doi: 10.1161/CIRCULATIONAHA.113.007477.
10. Edwards FH, Cohen DJ, O'Brien SM, Peterson ED, Mack MJ, Shahian DM, Grover FL, Tuzcu EM, Thourani VH, Carroll J, Brennan JM, Brindis RG, Rumsfeld J, Holmes DR Jr. Development and validation of a risk prediction model for in-hospital mortality after transcatheter aortic valve replacement. *JAMA Cardiol*. 2016;1:46–52.
11. Montori VM, Guyatt GH. Intention-to-treat principle. *CMAJ*. 2001;165:1339–1341.
12. Jones RH. Intraoperative crossover: the well-kept surgical secret to apparent surgical success. *J Am Coll Cardiol*. 2005;45:1529–1531. doi: 10.1016/j.jacc.2005.03.012.
13. Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med*. 2005;24:1185–1202. doi: 10.1002/sim.1970.
14. Alfredsson J, Stebbins A, Brennan JM, Matsouaka R, Afilalo J, Peterson ED, Vemulapalli S, Rumsfeld JS, Shahian D, Mack MJ, Alexander KP. Gait speed predicts 30-day mortality after transcatheter aortic valve replacement: results from the Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy Registry. *Circulation*. 2016;133:1351–1359.
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;8:e002875. doi: 10.1161/CIRCINTERVENTIONS.115.002875.
16. Cohen ME, Ko CY, Bilimoria KY, Zhou L, Huffman K, Wang X, Liu Y, Kraemer K, Meng X, Merkow R, Chow W, Matel B, Richards K, Hart AJ, Dimick JB, Hall BL. Optimizing ACS NSQIP modeling for evaluation of surgical quality and risk: patient risk adjustment, procedure mix adjustment, shrinkage adjustment, and surgical focus. *J Am Coll Surg*. 2013;217:336–46.e1. doi: 10.1016/j.jamcollsurg.2013.02.027.
17. 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.
18. Dimick JB, Staiger DO, Birkmeyer JD. Ranking hospitals on surgical mortality: the importance of reliability adjustment. *Health Serv Res*. 2010;45(6 pt 1):1614–1629. doi: 10.1111/j.1475-6773.2010.01158.x.
19. Morris CN. Parametric empirical Bayes inference: theory and applications. *J Am Stat Assoc*. 1983;78:47–55.
20. Shen W, Louis TA. Triple-goal estimates in two-stage hierarchical models. *J R Statist Soc B*. 1998;60:455–471.
21. Paddock SM, Louis TA. Percentile-based Empirical Distribution Function Estimates for Performance Evaluation of Healthcare Providers. *J R Stat Soc Ser C Appl Stat*. 2011;60:575–589. doi: 10.1111/j.1467-9876.2010.00760.x.
22. Paddock SM. Statistical benchmarks for health care provider performance assessment: a comparison of standard approaches to a hierarchical Bayesian histogram-based method. *Health Serv Res*. 2014;49:1056–1073. doi: 10.1111/1475-6773.12149.
23. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. 3rd ed, Volume 2. Boca Raton, FL: CRC Press, Taylor & Francis Group; 2014.
24. Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York, NY: Cambridge University Press; 2006.
25. Carroll JD, Edwards FH, Marinac-Dabic D, Brindis RG, Grover FL, Peterson ED, Tuzcu EM, Shahian DM, Rumsfeld JS, Shewan CM, Hewitt K, Holmes DR Jr, Mack MJ. The STS-ACC transcatheter valve therapy national registry: a new partnership and infrastructure for the introduction and surveillance of medical devices and therapies. *J Am Coll Cardiol*. 2013;62:1026–1034.

Circulation
Cardiovascular Quality and Outcomes

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

Sean M. O'Brien, David J. Cohen, John S. Rumsfeld, J. Matthew Brennan, David M. Shahian, David Dai, David R. Holmes, Rosemarie B. Hakim, Vinod H. Thourani, Eric D. Peterson and Fred H. Edwards
for the STS/ACC TVT Registry

Circ Cardiovasc Qual Outcomes. 2016;9:560-565; originally published online September 13, 2016;
doi: 10.1161/CIRCOUTCOMES.116.002756
Circulation: Cardiovascular Quality and Outcomes is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 1941-7705. Online ISSN: 1941-7713

Supplemental Material

Supplemental Methods

Details of Multivariable Analysis

Selection of Covariates. Covariates for case-mix adjustment were selected by a group of cardiologists and surgeons after reviewing the transcatheter valve therapy (TVT) data collection form and relevant prior literature, as well as performing an informal empirical analysis of bivariate associations. All candidate variables available in Version 1.1 of the TVT data collection form were individually discussed from the standpoint of clinical relevance, variation across hospitals, and data quality. Since the goal of the analysis is to risk-adjust procedural outcomes, only pre-procedural patient variables were considered for inclusion. In general, our goal was to adjust for baseline factors that were present at the beginning of the start of the transcatheter aortic valve replacement (TAVR) procedure and not to adjust for discretionary care processes during or after the initiation of the procedure. In general, all variables screened in univariable analyses were retained in the final model, regardless of their apparent association or lack of association with mortality. Exceptions were: operator reason for the procedure (inoperable vs. other), aortic stenosis, mean gradient, albumin, baseline 5-meter walk assessment, and baseline Kansas City Cardiomyopathy Questionnaire (KCCQ-12) assessment. Operator reason for the procedure was regarded as imprecisely defined and subjective. Aortic stenosis was present in >99% of patients and was excluded due to data quality considerations and the small number of patients available for estimating risk in patients coded without stenosis. Remaining variables were excluded due to high prevalence of missing data: 5-meter walk assessment (missing 60%), KCCQ-12 assessment (missing 61%), and albumin missing data (16%).

A Note on Adjustment for Processes of Care. An important principle of risk adjustment is to adjust for patient factors that are beyond the control of the entity being assessed and to avoid adjusting for factors that result from the care provided. For example, it is generally unadvisable to adjust for discretionary care processes such as intraoperative medications and procedural techniques. On the other hand, certain care processes tend to be given to patients who have a relatively serious preoperative presentation. In some cases, knowledge that a care process was delivered (e.g. preoperative intra-aortic balloon pump [IABP] or inotropes) may provide indirect information about unmeasured aspects of the patient's baseline status. For these reasons, preoperative use of a mechanical assist device and preoperative inotropes were incorporated into our risk adjustment model. Similarly, use of a non-femoral valve sheath access site was adjusted in the model. Although the decision to use a non-femoral access site is under the control of the care provider, patients receiving non-femoral access differ systematically from conventional access patients in ways that are not fully captured by other TVT data elements. For this reason, non-femoral was included as a covariate. As a result of adjusting for access site, multivariable analysis may obscure or "adjust away" some true differences in quality that are reflected in the adoption of more or less effective care processes (e.g., the decision to use femoral or alternative access). Prior to deciding to adjust for access site, a sensitivity analysis was conducted to assess the impact of adjusting versus not adjusting for this variable. The Pearson correlation between risk-adjusted mortality rates (RAMRs) calculated with (vs. without) adjustment for access site was 0.989.

Functional Form of Predictors. Graphical exploratory analyses were used to determine the functional form of continuous variables and to decide whether categorical variables with several

categories could be collapsed into fewer categories. In a preliminary marginal logistic regression model in which flexible regression splines were used for modeling continuous variables, plots of the variables body surface area (BSA), glomerular filtration rate (GFR), hemoglobin, and ejection fraction revealed an approximately linear association with the log-odds of mortality except for an attenuation of the slope in the extreme tails of each variable's distribution. To account for this attenuation in the tails of the distribution, each of these variables was modeled as linear across a designated interval and as a constant (i.e., zero slope) outside of this interval. For example, ejection fraction was modeled as linear between 15% and 60% and as constant in the range <15% or >60%. Truncation intervals for other variables were: BSA 1.4–2.6 m^2; GFR 30–90 mL/min/1.73 m^2; and hemoglobin 8–15g/dL. For modeling renal function, patients on dialysis were represented by an indicator variable for dialysis without further adjustment for the patient's estimated GFR. For patients not on dialysis, the relationship between estimated GFR and mortality was modeled as linear between 30 and 90 mL/min/1.73 m^2, and constant below 30 or above 90 mL/min/1.73 m^2. Surgery date displayed an apparent linear association with mortality and was, therefore, modeled as linear. For other continuous variables, the association with mortality was relatively more complicated. Age was modeled as piecewise linear with a change of slope at 75 years and no slope below 50 years or above 100 years. Finally, platelet count was modeled as piecewise linear with a change of slope at 200k and no slope when below 50kμL or above 400kμL.

Interactions. The ability to identify interaction effects empirically based on the data was limited due to the large number of candidate interactions (e.g., more than 1000 pairwise interactions; more than 15,000 three-way interactions) in relation to the number of endpoint events. Since

automated variable selection methods can be relatively sensitive to chance variation, we instead opted to develop a main effects model approximation. The single exception was the inclusion of an interaction between sex and BSA, which was selected a priori based on our prior experience with the development of the Society of Thoracic Surgeons (STS) risk model for surgical aortic valve replacement.

Modeling of Acuity Status. For modeling acuity status, a derived status variable was created which combines information from the following variables:

- procedure status (#6055)
- prior cardiac arrest within 24 hours (#5035)
- prior cardiac shock within 24 hours (#5030)
- pre-procedure inotropes (#5400)
- pre-procedure mechanical assist device (#6095)

For the derived variable, the following categories were created:

Category	**Definition**
Category 1	Defined as meeting both of the following:
	1. Procedure status is elective
	2. No pre-procedure shock, inotropes, mechanical assist device, or cardiac arrest
Category 2	Defined as meeting both of the following:
	1. Procedure status is urgent
	2. No pre-procedure shock, inotropes, mechanical assist device, or cardiac arrest
Category 3	Defined as meeting all three of the following:
	1. Procedure status is elective or urgent
	2. Patient had pre-procedure shock, inotropes, or mechanical assist device
	3. No prior cardiac arrest within 24 hours of operation
Category 4	Defined as either one or both of the following:
	1. Procedure status is emergency or salvage; or
	2. Patient had prior cardiac arrest within 24 hours of operation

Modeling of Access Site. TAVR access sites were femoral (58.8%), transapical (31.3%), transaortic (7.3%), transiliac (0.5%), subclavian (0.4%), axillary (0.1%), transcarotic (0.1%), and other (1.4%). To reduce the potential for noisy estimates in rare categories, access site was dichotomized as femoral versus non-femoral. The decision to dichotomize access site was guided by prior clinical experience and was made prior to observing the association of access site with mortality. To assess whether modeling results were impacted by our decision to treat access site as dichotomous, we subsequently re-estimated hospital-specific risk-adjusted mortality rates and regression coefficients using a modified version of the model with access site modeled as a three-category variable (femoral versus transapical versus other). We found that odds ratio estimates in the modified model were nearly identical for transapical versus femoral and "other" versus femoral (OR = 1.92, OR = 1.89, respectively) and were also nearly identical to the odds ratio for non-femoral versus femoral in the final 2-category model (OR = 1.90). The surprisingly similar odds ratios suggest that minimal information was lost in the decision to collapse access site into 2 rather than 3 categories. The Pearson correlation between point estimates of hospital-specific risk-adjusted mortality rates from the final and modified models was >0.999.

Mathematical Form of Regression Model. The goal of multivariable analysis was to estimate participant-specific standardized mortality ratios (SMRs) and to quantify variation in SMRs across participants. Variation in SMRs was assumed to be described by a two-level hierarchical logistic regression model with participant-specific random intercept parameters. At the first level (within participants), a patient's probability of in-hospital mortality was modeled as:

$$\pi_{ji} = expit(x_{ji}\beta + \alpha_j)$$

where π_{ji} denotes the probability of mortality for the i-th patient treated by the j-th participant; α_j is a (random effect) intercept parameter for participant j; x_{ji} is a set of patient baseline covariates (e.g., age, weight, etc.) for the i-th patient treated by the j-th participant; β is a set of unknown regression parameters; and $expit(x) = 1/[1 + \exp(-x)]$. Outcomes of individual patients were assumed to be conditionally independent conditional upon $(\alpha_1, \ldots\ \alpha_N)$, β and the x_{ji}'s. At the second level (between hospitals), α_j's were assumed to be independent and identically distributed according to normal distribution with unknown mean and variance, that is,

$$\alpha_j \sim N(\mu, \sigma^2),$$

where μ and σ^2 denote unknown parameters to be estimated from the data.

Calculation of SMRs and RAMRs. The unknown SMR of the j-th participant is defined as the quantity

$$\text{SMR}_j = \frac{\sum_{i=1}^{n_j} expit\left(x_{ji}\beta + \alpha_j\right)}{\sum_{i=1}^{n_j} expit\left(x_{ji}\beta + \mu\right)},$$

where n_j is the number of patients at the j-th participant. The SMR is interpreted as the ratio of a participant's actual expected mortality rate for a given set of patients relative to what the expected mortality rate would be hypothetically if the participant's performance was "average." An SMR >1 implies worse than expected mortality, whereas SMR <1 implies better than expected mortality. A participant's RAMR was calculated by the formula $\text{RAMR}_j = \text{SMR}_j \times$ (overall mortality rate) where (overall mortality rate) denotes the proportion of patients dying prior to discharge in the overall study sample.

Computation. Model parameters were initially estimated via maximum likelihood with a Laplace approximation of the marginal log-likelihood as implemented in SAS version 9.2 software. Subsequently, model parameters were re-estimated in a Bayesian statistical framework by specifying a prior probability distribution for unknown parameters. Advantages of Bayesian methodology include the ability to make probability statements about each hospital's SMR and the ability to draw exact statistical inferences about complicated functions of unknown model parameters. Since our prior knowledge was limited, we specified a vague proper prior distribution for unknown parameters. The prior consisted of independent normal distributions for the elements of β and μ, and an independent inverse Gamma distribution for σ^2. Similar results were obtained when using three alternative prior distributions for σ^2 (results not shown). Posterior means and 95% credible intervals (CrIs) were calculated using Markov Chain Monte Carlo (MCMC) simulations as implemented in WinBUGS version 1.4 software. Posterior summaries were based on 25,000 sets of simulated parameter values which were generated after discarding 4 out of 5 simulation iterations to reduce autocorrelation and a long burn-in period to ensure convergence.

Missing Data. Missing data for model covariates were below 1% for most covariates. Only one covariate had more than 3% missing data (number of prior cardiac operations; missing 3.6%). Missing data rates for other covariates were as follows: age (0.0%); BSA (0.1%); GFR (0.6%); ejection fraction (2.9%); hemoglobin (0.3%); platelet count (0.7%); dialysis (0.3%); Hispanic ethnicity (2.2%); left main disease (2.4%); proximal left anterior descending (2.6%); prior myocardial infarction (0.6%); infectious endocarditis (0.3%); prior stroke (0.3%); prior transient ischemic attack (0.7%); carotid stenosis (2.6%); prior peripheral artery disease (0.5%); smoker

(0.4%); diabetes (0.3%); New York Heart Association class (1.8%); atrial fibrillation or flutter (0.4%); conduction defect (0.8%); chronic lung disease (1.8%); home oxygen (1.0%); hostile chest (1.0%); porcelain aorta (0.8%); access site (1.0%); pacemaker (0.3%); previous internal cardioverter defibrillator (0.5%); prior percutaneous coronary intervention(0.5%); prior coronary artery bypass grafting (0.3%); number of prior cardiac operations (3.6%); prior aortic valve procedure (0.3%); prior non-aortic procedure (0.8%); aortic etiology (0.7%); valve morphology (1.7%); aortic insufficiency (1.0%); mitral insufficiency (0.8%); tricuspid insufficiency (1.1%); status (0.4%); acuity category (2.3%); prior cardiac arrest (0.3%); shock (0.3%); pre-procedural inotropes (1.7%); mechanical assist device (0.4%)

Missing data were imputed to the most common category of categorical variables and to the median or subgroup-specific median of continuous variables. More computationally intensive missing data strategies, such as multiple imputation, were not used for this project because of the low rate of missing data and because it would be resource-intensive to implement them in combination with the computationally intensive MCMC procedure described above.

Model Summary. For summarizing the association between model covariates and mortality, odds ratios were estimated by maximum likelihood and presented with 95% confidence intervals and Wald-type p-values. Maximum likelihood was only used for the purpose of summarizing the model and obtaining conventional p-values for the online supplement. Fully Bayesian analysis (rather than maximum likelihood) was used for all analyses reported in the main text of the manuscript.

Assessment of Model Calibration and Discrimination. Prior to estimating hospital-specific performance, we first performed graphical analyses to assess the model's ability to fit the data. To assess calibration, we compared observed versus average expected mortality rates within ssubgroups of patients based on quantiles of predicted risk. Large discrepancies between observed and expected probabilities would suggest that the functional form of the model was misspecified. In addition to assessing calibration, we also estimated the C-index (i.e., discrimination) of the model. The C-index quantifies the ability of a classification algorithm to separate the target population into a group of patients who will have the endpoint of interest and a group of patients who will not have the endpoint of interest. A low C-index does not imply that the model is misspecified or that hospital comparisons will be biased. Nonetheless, the C-index is widely reported and was, therefore, calculated for the sake of completeness.

After assessing calibration and discrimination in the original full sample, the model was subsequently re-estimated using a 70% random sample of records and assessed for calibration and discrimination in the remaining 30% of records. Calibration graphs and C statistics from the original internal validation ("overall sample") and the split sample ("validation sample") are presented side-by-side below.

Estimation of Histogram of True RAMR's for 318 Hospitals (Technical Details). To create Figure 2, the range of possible RAMR values (0 to 100 percent) was partitioned into 201 equally-sized categories with cutpoints 0, 0.5, 1.0, etc. Let κ_c denote the number of true RAMR's (out of 318) falling in the interval $\left(\frac{c-1}{2}, \frac{c}{2}\right)$, for $c = 1,2, \ldots, 200$, and let $\bar{\kappa}_c = 100 \times \kappa_j /$ 318 denote the corresponding percentage. The posterior mean of each $\bar{\kappa}_c$ was calculated as $E[\bar{\kappa}_c \ |\text{data}] = \frac{1}{M}\sum_{i=1}^{M} \bar{\kappa}_c^{(i)}$ where M denotes the number of MCMC iterations and $\bar{\kappa}_c^{(i)}$ denotes the

value of $\bar{\kappa}_c$ on the i-th MCMC iteration. Figure 2 was obtained by plotting bars of width $\left(\frac{c-1}{2}, \frac{c}{2}\right)$ and height $E[\bar{\kappa}_c \ |\text{data}]$ over the range for which $E[\bar{\kappa}_c \ |\text{data}] > 0$.

Estimation of Average RAMRs by RAMR Decile (Technical Details). Let $\theta_1, \theta_2, \ldots, \theta_{318}$ denote the collection of RAMR's for all 318 hospitals and let $\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(318)}$ denote the same RAMRs after sorting them in order from smallest to largest. The quantities

$$\bar{\theta}_{10} = \frac{1}{32}\sum_{j=1}^{32} \theta_{(j)} \quad \text{and} \quad \bar{\theta}_{90} = \frac{1}{32}\sum_{j=1}^{32} \theta_{(318-j+1)}$$

are the average RAMR's among the 10 percent of hospitals with the lowest true RAMRs and the average RAMRs among the 10 percent of hospitals with the highest true RAMR's, respectively. The posterior means of $\bar{\theta}_{10}$ and $\bar{\theta}_{90}$ were calculated as

$$E[\bar{\theta}_{10} \,|\text{data}] = \tfrac{1}{M}\textstyle\sum_{i=1}^{M} \bar{\theta}_{10}^{(i)} \text{ and } E[\bar{\theta}_{90} \,|\text{data}] = \tfrac{1}{M}\textstyle\sum_{i=1}^{M} \bar{\theta}_{90}^{(i)}$$

where M denotes the number of MCMC iterations and $\bar{\theta}_{10}^{(i)}$ and $\bar{\theta}_{90}^{(i)}$ denote the value of $\bar{\theta}_{10}$ and $\bar{\theta}_{90}$on the i-th MCMC iteration, that is,

$$\bar{\theta}_{10}^{(i)} = \frac{1}{32}\sum_{j=1}^{32} \theta_{(j)}^{(i)} \quad \text{and} \quad \bar{\theta}_{90}^{(i)} = \frac{1}{32}\sum_{j=1}^{32} \theta_{(318-j+1)}^{(i)}$$

where $\theta_{(1)}^{(i)}, \theta_{(2)}^{(i)}, \ldots, \theta_{(318)}^{(i)}$ denotes the simulated values of $\theta_1, \theta_2, \ldots, \theta_{318}$ on the i-th MCMC iteration after sorting them in order from smallest to largest. Analogously, posterior means for the difference $\bar{\theta}_{90} - \bar{\theta}_{10}$ and ratio $\bar{\theta}_{90}/\bar{\theta}_{10}$ were calculated as

$$E[\bar{\theta}_{90} - \bar{\theta}_{10} \,|\text{data}] = \frac{1}{M}\sum_{i=1}^{M} \bar{\theta}_{90}^{(i)} - \bar{\theta}_{10}^{(i)} \quad \text{and} \quad E[\bar{\theta}_{90}/\bar{\theta}_{10} \,|\text{data}] = \frac{1}{M}\sum_{i=1}^{M} \bar{\theta}_{90}^{(i)}/\bar{\theta}_{10}^{(i)}.$$

For each quantity described above, a 95% Bayesian probability interval was obtained via MCMC by calculating the 2.5$^{\text{th}}$ and 97.5$^{\text{th}}$ percentiles of the set of simulated values.

A 95% Bayesian probability interval for $\bar{\theta}_{10}$ was obtained by calculating the 2.5th and 97.5th percentiles of the set of numbers $\bar{\theta}_{10}^{(1)}, \bar{\theta}_{10}^{(2)}, \ldots, \bar{\theta}_{10}^{(M)}$. Analogous calculations were used for inferences about $\bar{\theta}_{90}$, $\bar{\theta}_{90} - \bar{\theta}_{10}$, and $\bar{\theta}_{90}/\bar{\theta}_{10}$.

Estimation of Signal-to-Noise Ratio (Technical Details). Let θ_j denote the true unknown SMR for the j-th of 318 participants. Prior to estimating the signal-to-noise ratio, the numerical value of θ_j was estimated for each hospital. Estimation was done using MCMC simulations and involved the following steps: First, for each j, we randomly generated a large number M of possible numerical values of θ_j by sampling from the Bayesian posterior probability distribution of θ_j via MCMC sampling. Let $\theta_j^{(i)}$ denote the i-th of these M randomly sampled numerical values for the j-th participant. Second, for each j, the posterior mean $\hat{\theta}_j$ of θ_j was calculated as the arithmetic average of the randomly sampled values $\theta_j^{(1)}, \ldots, \theta_j^{(M)}$; in other words $\hat{\theta}_j = \frac{1}{M}\sum_{i=1}^{M} \theta_j^{(i)}$. Our signal-to-noise ratio measure was defined as the squared correlation between the set of hospital-specific estimates $\hat{\theta}_1, \ldots, \hat{\theta}_J$ and the corresponding unknown true values $\theta_1, \ldots, \theta_J$. Let ρ^2 denote the unknown true squared correlation of interest and let $\hat{\rho}^2$ denote an estimate of this quantity. The estimate was calculated as

$$\hat{\rho}^2 = \frac{1}{M}\sum_{i=1}^{M} \rho_{(i)}^2$$

where

$$\rho_{(i)}^2 = \frac{\left[\sum_{j=1}^{J}\left(\theta_j^{(i)} - \bar{\theta}^{(i)}\right)\left(\hat{\theta}_j - \bar{\theta}\right)\right]^2}{\sum_{j=1}^{J}\left(\theta_j^{(i)} - \bar{\theta}^{(i)}\right)^2 \sum_{j=1}^{J}\left(\hat{\theta}_j - \bar{\theta}\right)^2} \quad \bar{\theta} = \frac{1}{JM}\sum_{j=1}^{J}\sum_{i=1}^{M} \theta_j^{(i)} \quad \text{and} \quad \bar{\theta}^{(i)} = \frac{1}{J}\sum_{j=1}^{J} \theta_j^{(i)}.$$

A 95% Bayesian probability interval for ρ^2 was obtained calculating the 2.5th and 97.5th percentiles of the set of numbers $\rho^2_{(1)},\ldots,\rho^2_{(M)}$.

Supplemental Tables

Supplemental Table 1. Covariates Discussed and Examined for Inclusion in Multivariable Analysis

✓ Age	✓ Prior peripheral artery disease	✓ # prior cardiac operations
✓ BSA	✓ Current/recent smoker	✓ Prior aortic procedure
✓ Sex	✓ Diabetes	✓ Prior other valve procedure
✓ Race/ethnicity	✓ NYHA class	✓ Aortic etiology
✓ eGFR	✓ Atrial fibrillation/flutter	✓ Valve morphology
✓ Dialysis	✓ Conduction defect	✗ Aortic stenosis
✓ Ejection fraction	✓ Chronic lung disease	✗ Aortic valve mean gradient
✓ Hemoglobin	✓ Home oxygen	✓ Aortic insufficiency
✓ Platelet count	✓ Hostile chest	✓ Mitral insufficiency
✓ Procedure date	✓ Porcelain aorta	✓ Tricuspid insufficiency
✓ LMD ≥50%	✗ Operator reason for procedure	✓ Acuity status
✗ # of diseased vessels	✓ Access site	✓ Cardiogenic shock
✓ Proximal LAD ≥70%	✓ Pacemaker	✓ Cardiac arrest w/in 24 hours
✓ Prior MI	✓ Previous ICD	✓ Pre-procedure inotropes
✓ Endocarditis	✓ Prior PCI	✓ Mechanical assist device
✓ Prior TIA/stroke	✓ Prior CABG	✗ Albumin
✓ Carotid stenosis		

✓ = covariates retained in final model (alone or as part of a derived variable)
✗ = covariates excluded from final model
BSA indicates body surface area; CABG, coronary artery bypass grafting; eGFR, estimated glomerular filtration rate; ICD, implantable cardioverter defibrillator; LAD, left anterior descending; LMD, left main disease; MI, myocardial infarction; NYHA, New York Heart Association; PCI, percutaneous coronary intervention; TIA, transient ischemic attack

Supplemental Table 2. Estimated ORs and 95% CIs from Hierarchical Logistic Regression

Variable	OR (95% CI)	p-value
Age (per 5 years when ≤75)	0.95 (0.86 - 1.05)	0.32
Age (per 5 years when >75)	1.17 (1.09 - 1.25)	<0.001
BSA (per m^2 for male)	0.41 (0.25 - 0.67)	<0.001
BSA (per m^2 for female)	0.47 (0.30 - 0.74)	<0.001
Sex (male vs. female)	0.84 (0.73 - 0.97)	0.01
Race (White non-Hispanic vs. other)	0.87 (0.71 - 1.08)	0.21
GFR (per 5 unit)	0.94 (0.92 - 0.96)	<0.001
Ejection fraction (per 5 unit)	0.98 (0.95 - 1.01)	0.19
Hemoglobin (per g/dL)	1.03 (0.99 - 1.07)	0.19
Platelet count (per 10k when ≤200k)	0.97 (0.95 - 0.99)	0.003
Platelet count (per 10k when >200k)	1.01 (0.99 - 1.02)	0.32
Procedure date (per 30 day)	0.99 (0.98 - 0.99)	<0.001
Left main ≥50%	1.08 (0.87 - 1.33)	0.47
Proximal LAD	1.09 (0.92 - 1.30)	0.32
Prior MI	1.17 (1.01 - 1.36)	0.03
Current dialysis	2.41 (1.80 - 3.23)	<0.001
Endocarditis	0.98 (0.49 - 1.96)	0.95
Prior TIA w/o stroke, or prior stroke	1.10 (0.95 - 1.29)	0.20
Carotid stenosis one or both	1.32 (1.12 - 1.55)	<0.001
Prior PAD	1.16 (1.01 - 1.33)	0.03
Smoker	0.69 (0.50 - 0.95)	0.02
Diabetes	1.00 (0.87 - 1.15)	0.99
NYHA class 4	1.20 (1.04 - 1.40)	0.02
Afib/flutter	1.11 (0.98 - 1.27)	0.11
Conduction defect	1.04 (0.91 - 1.19)	0.57
CLD (severe)	1.45 (1.21 - 1.73)	<0.001
Home oxygen	1.21 (1.01 - 1.46)	0.04
Hostile chest	1.10 (0.87 - 1.38)	0.43
Porcelain aorta	0.98 (0.77 - 1.24)	0.84
Non-femoral access	1.90 (1.66 - 2.17)	<0.001
Pacemaker	0.95 (0.80 - 1.12)	0.53
Previous ICD	1.15 (0.85 - 1.54)	0.36
Prior PCI	0.99 (0.87 - 1.13)	0.87
Prior CABG	0.82 (0.64 - 1.06)	0.12
Prior cardiac operators (1 vs. 0)	0.96 (0.76 - 1.23)	0.76
Prior cardiac operators (2+ vs. 0)	1.01 (0.70 - 1.46)	0.95
Prior aortic procedure	1.11 (0.94 - 1.30)	0.22
Prior non-aortic procedure	0.84 (0.56 - 1.27)	0.42
Aortic etiology (degenerative vs. other)	1.16 (0.85 - 1.58)	0.34
Valve morphology (tricuspid)	0.91 (0.71 - 1.16)	0.46
Aortic insufficiency (moderate/severe)	0.99 (0.85 - 1.16)	0.93
Mitral insufficiency (moderate/severe)	1.02 (0.89 - 1.17)	0.79
Tricuspid insufficiency (moderate/severe)	1.22 (1.05 - 1.41)	0.009

Acuity category 2	1.53 (1.25 - 1.87)	<0.001
Acuity category 3	2.50 (1.98 - 3.16)	<0.001
Acuity category 4	3.70 (2.12 - 6.46)	<0.001

Afib/flutter indicates atrial fibrillation/atrial flutter; BSA, body surface area; CABG, coronary artery bypass grafting; CI, confidence interval; CLD, chronic lung disease; GFR, glomerular filtration rate; ICD, implantable cardioverter defibrillator; LAD, left anterior descending; MI, myocardial infarction; NYHA, New York Heart Association; OR, odds ratio; PAD, peripheral artery disease; PCI, percutaneous coronary intervention; TIA, transient ischemic attack

Supplemental Table 3. Definitions of Variables Used in Multivariable Analysis

Variable	Label
Age function 1	= min (max [age, 50],100)
Age function 2	= max (age function 1 – 75, 0)
BSA, among males	= min (max [BSA, 1.4], 2.6) if patient is male, = 0 otherwise
BSA, among females	= min (max [BSA, 1.4], 2.6) if patient is female, = 0 otherwise
EF function	= min (max [LVEF, 15], 60)
Hemoglobin function	= min (max [hemoglobin, 8], 15)
Platelet function 1	= min (max [platelet count, 50000], 400000)
Platelet function 2	= max (platelets function 1 – 200000, 0)
Procedure date	= # of days from November 1, 2011 until date of operation
eGFR function	= min (max [eGFR, 30], 90) if patient is not on dialysis, = 0 otherwise
Variables coded as 1=yes, 0=no:	Dialysis; Female; Non-white or Hispanic; Left main ≥50%; Proximal LAD ≥70%; Prior MI; Endocarditis; Prior stroke or TIA; Carotid stenosis; Prior PAD; Current/recent smoker; Diabetes; NYHA class IV; Afib/flutter; Conduction defect; Chronic lung disease, severe; Home oxygen; Hostile chest; Porcelain aorta; Non-femoral access site; Pacemaker; Previous ICD; Prior PCI; Prior CABG; Prior cardiac operations, 1; Prior cardiac operations, 2 or more; Prior aortic valve procedure; Prior non-aortic valve procedure; Aortic etiology, degenerative; Valve morphology, tricuspid; Aortic insufficiency, ≥moderate; Mitral insufficiency, ≥moderate; Tricuspid insufficiency, ≥moderate; Acuity category 2; Acuity category 3; Acuity category 4

BSA indicates body surface area; CABG, coronary artery bypass grafting; EF, ejection fraction; eGFR, estimated glomerular filtration rate; ICD, implantable cardioverter defibrillator; LAD, left anterior descending; LVEF, left ventricular ejection fraction; NYHA, New York Heart Association; TIA, transient ischemic attack; PAD, peripheral artery disease; PCI, percutaneous coronary intervention

Supplemental Table 4. Model Discrimination Overall and By Subgroup

Internally Estimated C-Statistic (Overall Sample)	Cross-Validated C-Statistic (Validation Sample)
0.71	0.67

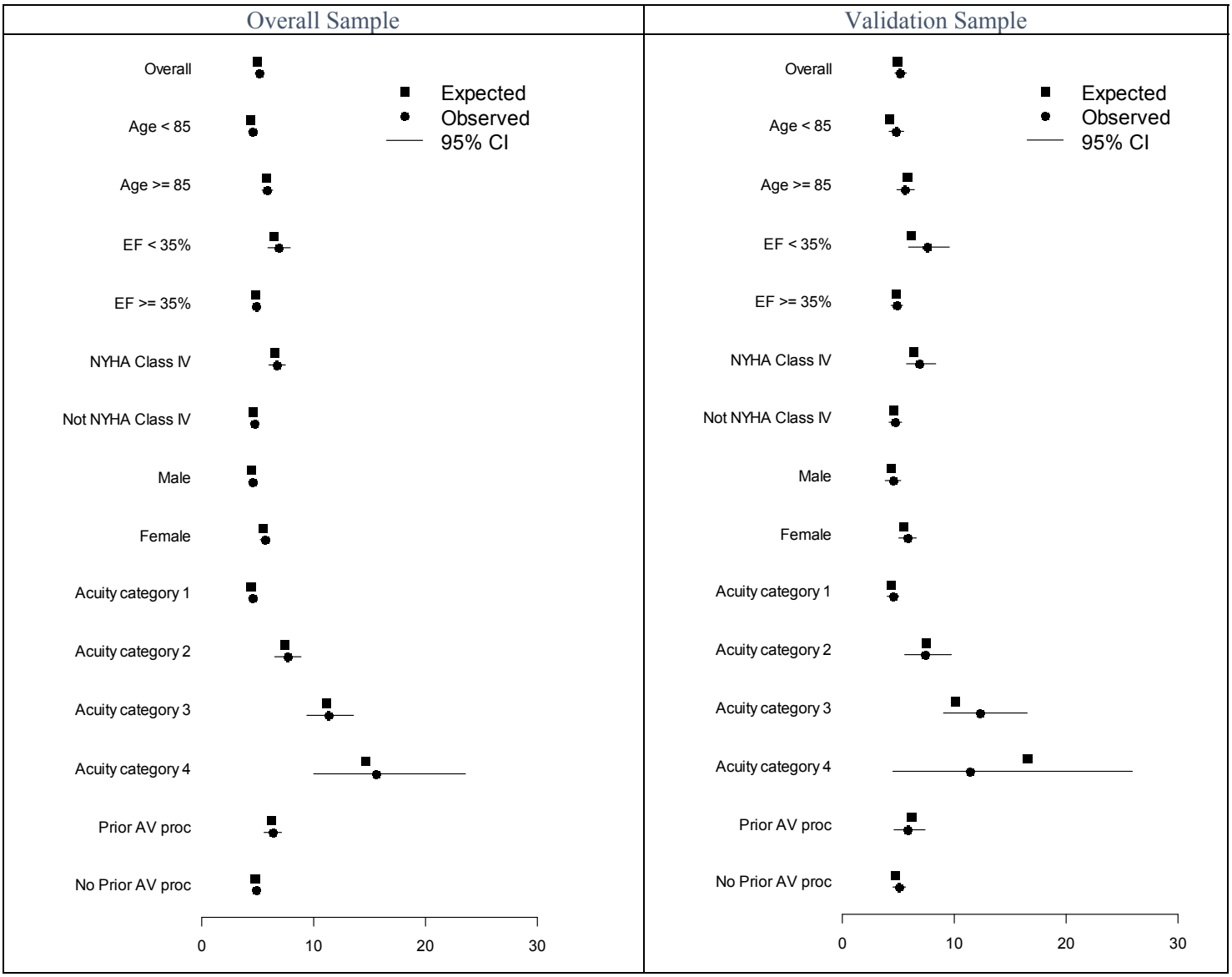
Supplemental Figures

Supplemental Figure Legends

Supplemental Figure 1. Comparison of Observed versus Average Predicted In-hospital Mortality Rates by Subgroup
Displayed is a comparison of observed versus average predicted in-hospital mortality rates by subgroup
AV indicates aortic valve; EF, ejection fraction; NYHA, New York Heart Association

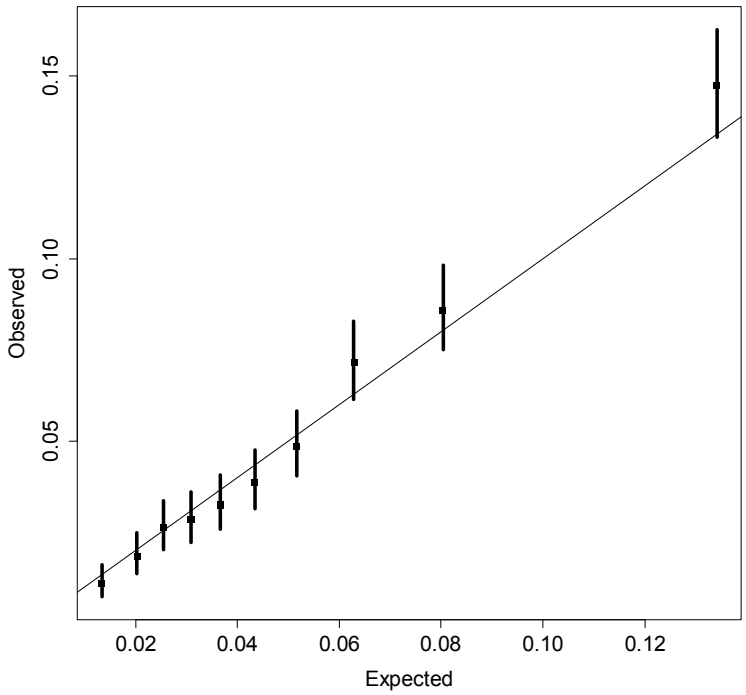
Supplemental Figure 2. Comparison of Observed Versus Expected Mortality Rates
Displayed is a comparison of observed versus expected mortality rates in: a) overall; b) by prior aortic valve status; c) sex; d) NYHA class; e) ejection fraction; f) age; and g) acuity status.
AV, aortic valve; EF, ejection fraction; NYHA, New York Heart Association

Supplemental Figure 1

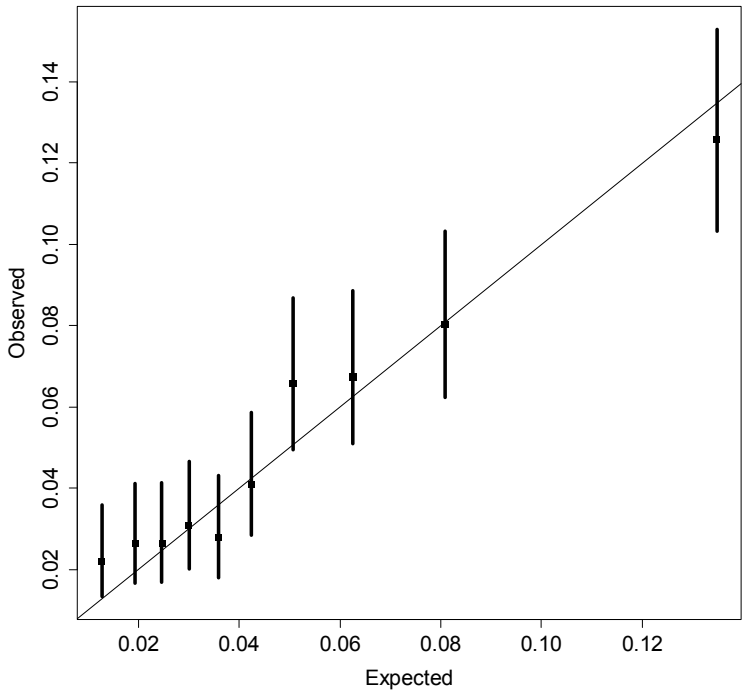


Supplemental Figure 2a

Overall Sample

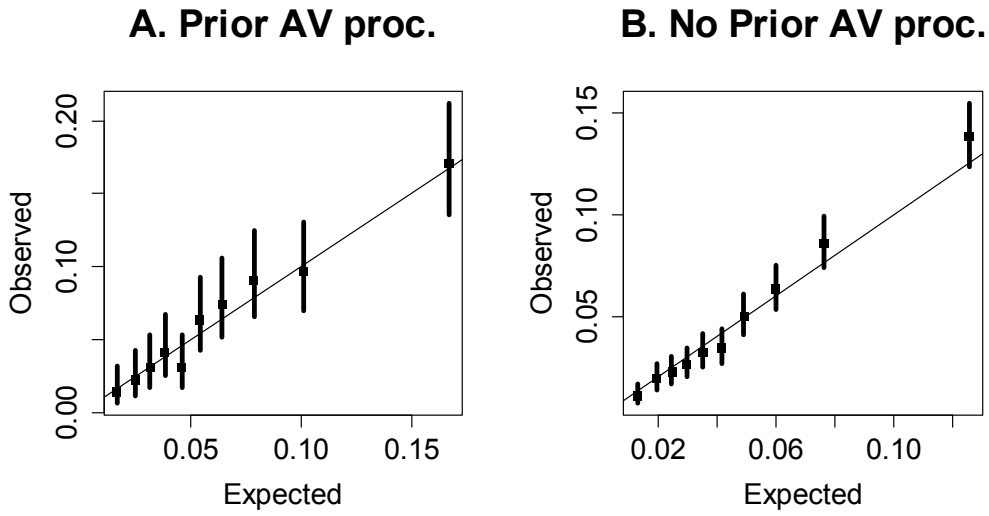


Validation Sample

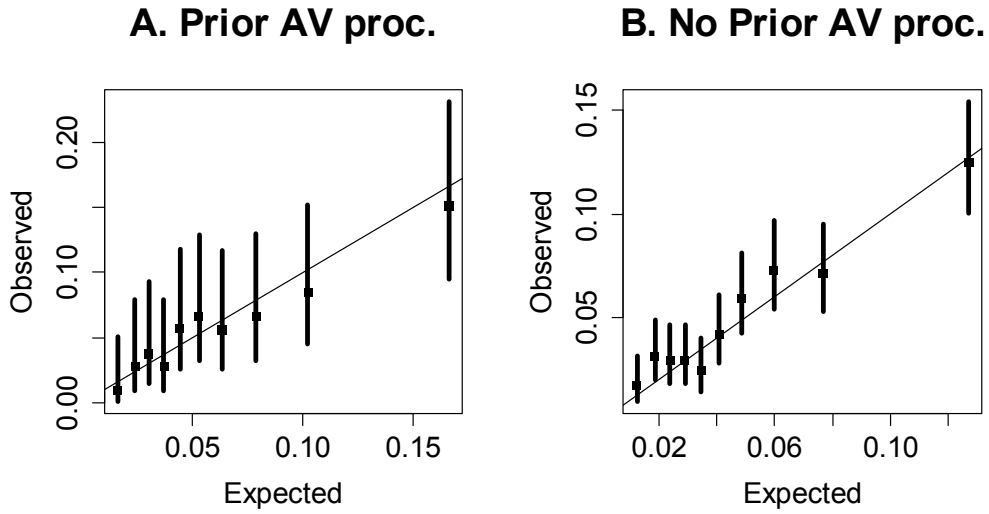


Supplemental Figure 2b

Overall Sample

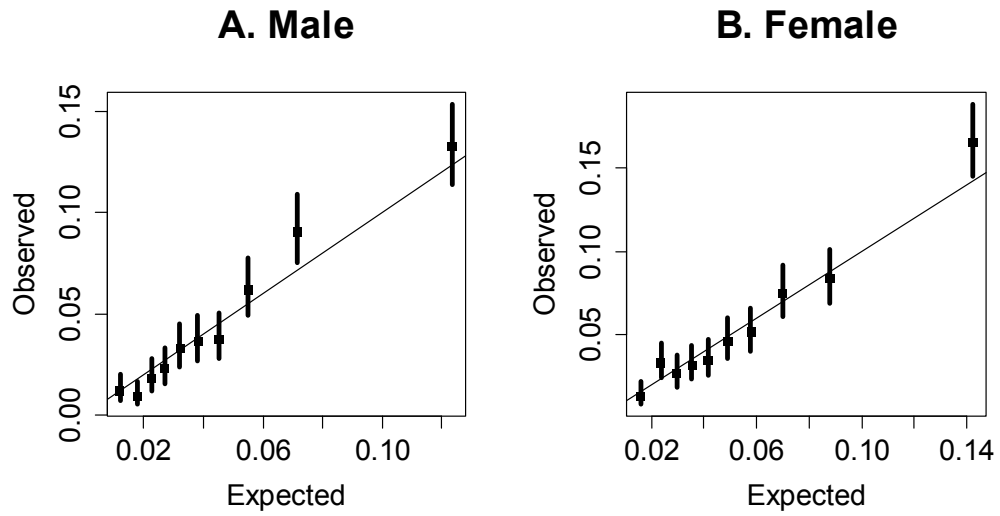


Validation Sample

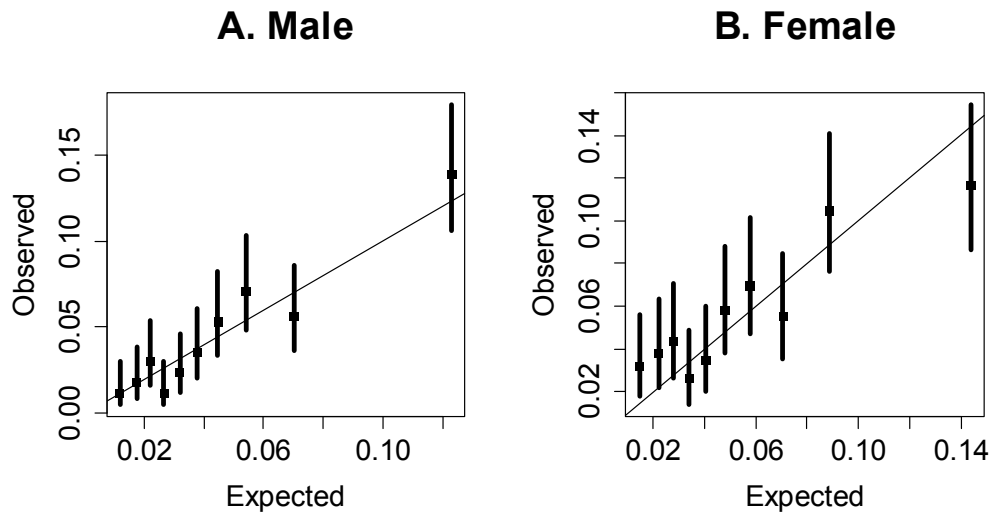


Supplemental Figure 2c

Overall Sample

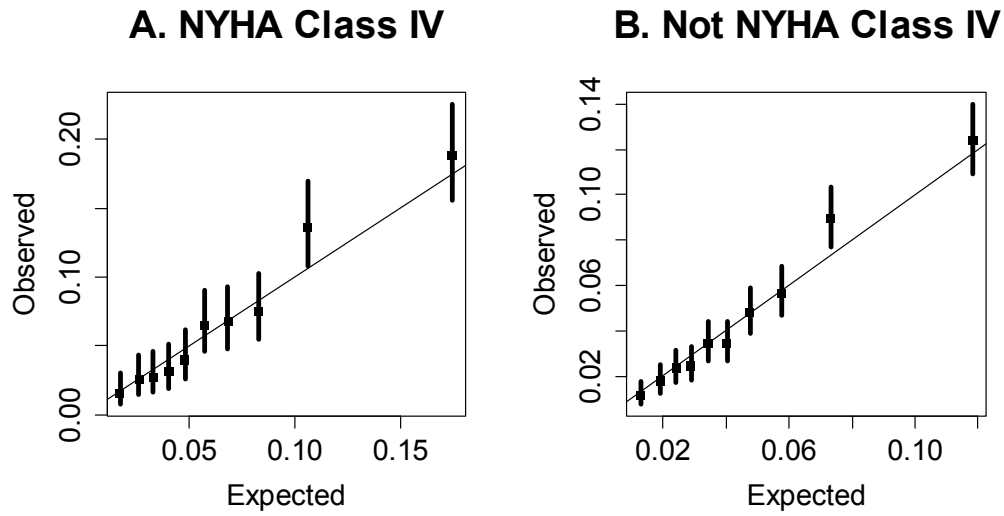


Validation Sample

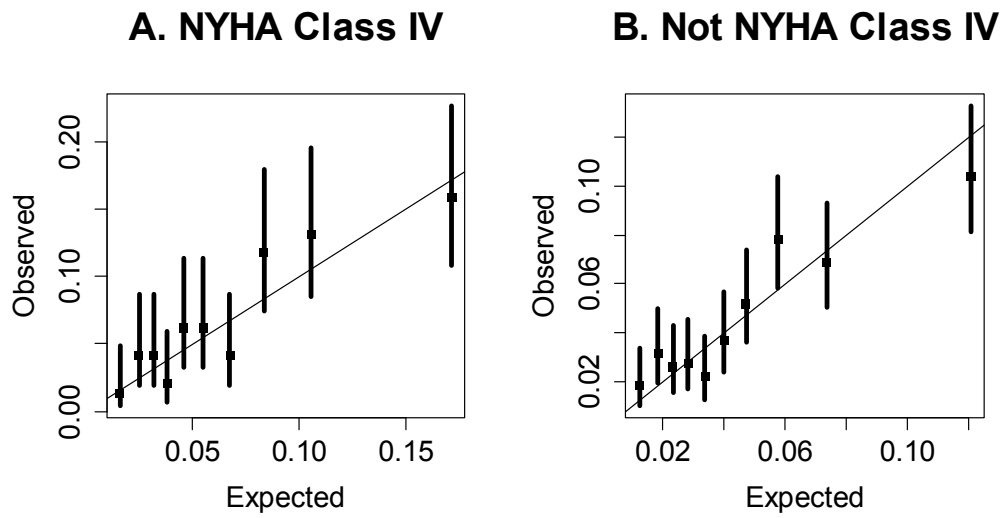


Supplemental Figure 2d

Overall Sample

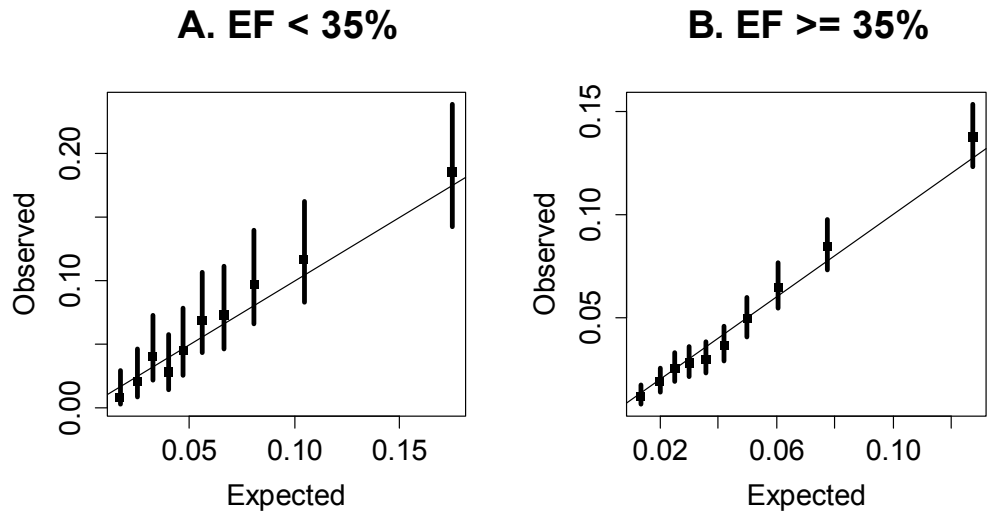


Validation Sample

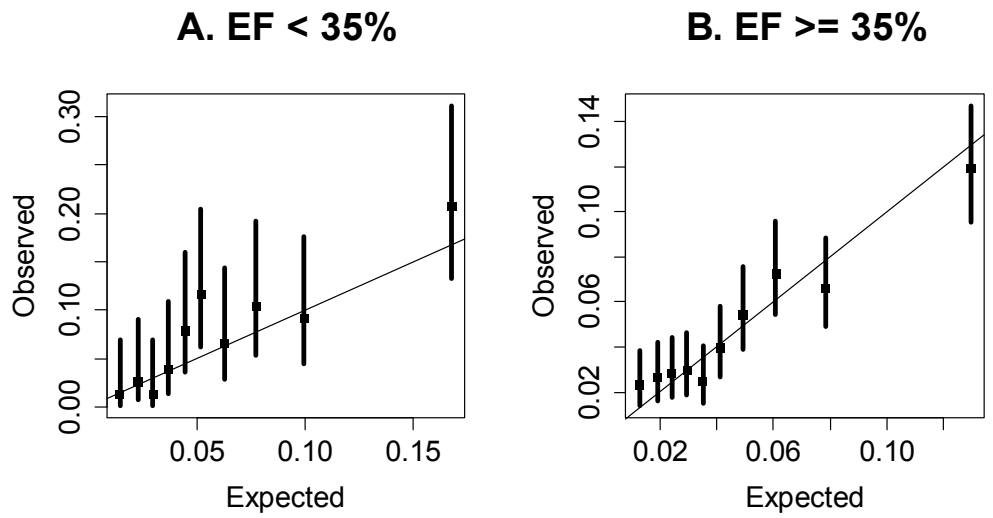


Supplemental Figure 2e

Overall Sample

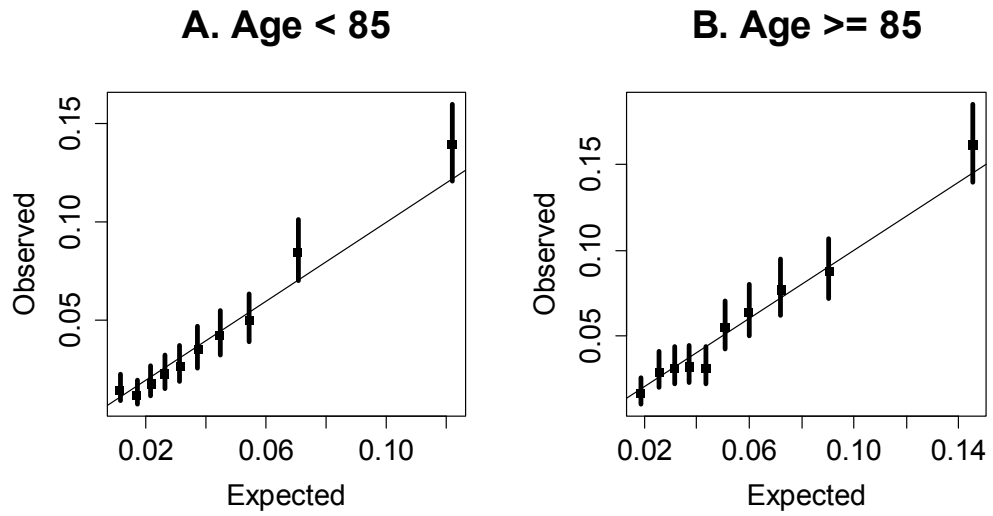


Validation Sample

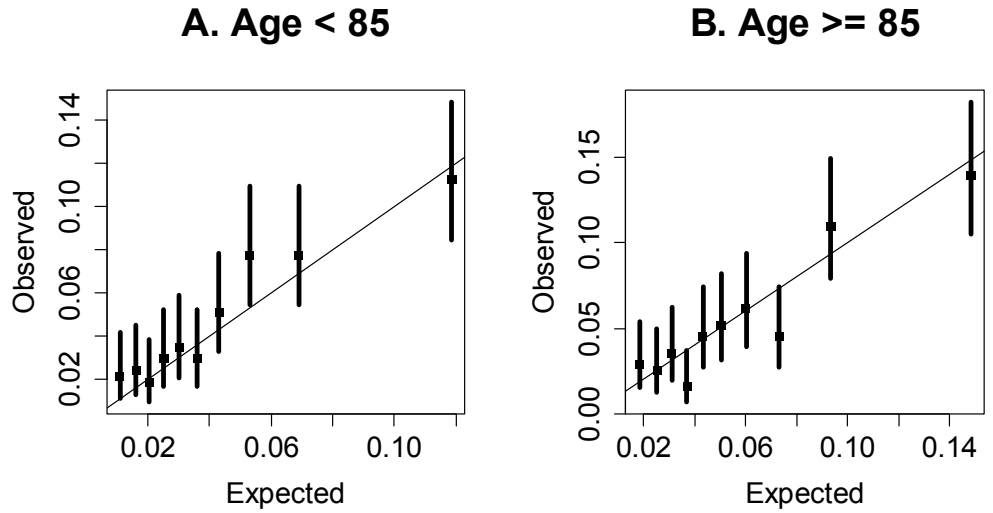


Supplemental Figure 2f

Overall Sample

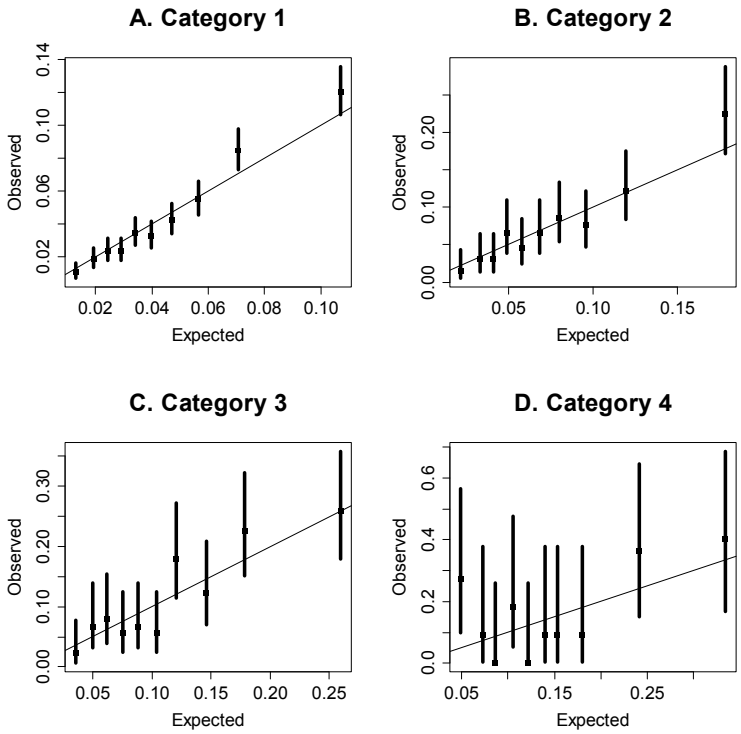


Validation Sample



Supplemental Figure 2g

Overall Sample



Validation Sample