Predicting Mortality after Transcatheter Aortic Valve Replacement: External Validation of the TVT Registry Model

Pilgrim, External validation of the TVT Registry Model

Thomas Pilgrim, MD*[1*]; Anna Franzone, MD, PhD*[1]; Stefan Stortecky, MD[1]; Fabian Nietlispach, MD, PhD[2]; Alan G. Haynes, PhD[3]; David Tueller, MD[4]; Stefan Toggweiler, MD[5]; Oliver Muller, MD[6]; Enrico Ferrari, MD [7]; Stéphane Noble, MD[8]; Francesco Maisano, MD[2]; Raban Jeger, MD[9]; Marco Roffi, MD[8]; Jürg Grünenfelder, MD[10]; Christoph Huber, MD[11]; Peter Wenaweser, MD[1,12]; Stephan Windecker, MD[1]

*the first two authors contributed equally to this manuscript

[1]Department of Cardiology, Swiss Cardiovascular Center Bern, University Hospital, Bern; [2]Department of Cardiology and Cardiovascular Surgery, University Hospital Zurich, Zurich; [3]Institute of Social and Preventive Medicine and Clinical Trials Unit, Bern University Hospital; [4]Department of Cardiology, Triemlispital, Zurich; [5]Department of Cardiology, Kantonsspital, Luzern; [6] Department of Cardiology Surgery, University Hospital, Lausanne; [7]Cardiac Surgery Unit, Cardiocentro Ticino Foundation, Lugano; [8]Division of Cardiology, University Hospital, Geneva; [9]Department of Cardiology, University Hospital, Basel; [10]Department of Cardiovascular Surgery, Hirslanden Klinik, Zurich; [11]Department of Cardiovascular Surgery, University Hospital Geneva, Geneva; [12]Department of Cardiology, Klinik im Park, Zurich

Word count: 5,680

Corresponding author:

Thomas Pilgrim, MD
Department of Cardiology
Swiss Cardiovascular Center
Bern University Hospital
University of Bern
Freiburgstrasse, 4
3010 Bern
Switzerland
Phone: +41 31 632 21 11
Fax: +41 31 632 47 70
Mail: thomas.pilgrim@insel.ch

Abstract

Background: The TVT Registry model was recently developed to predict the risk of in-hospital mortality in patients undergoing transcatheter aortic valve replacement (TAVR). We sought to externally validate the model in an independent data set of consecutively enrolled patients in the Swiss TAVI registry.

Methods and Results: The original prediction model was retrospectively applied to 3,491 consecutive patients undergoing TAVR in Switzerland between February 2011 and February 2016. We examined model performance in terms of discrimination (Harrel's c-index) and calibration (Hosmer–Lemeshow goodness-of-fit-test) for prediction of in-hospital and 30-day mortality, and compared its predictive accuracy with the Society of Thoracic Surgeons Predicted Risk of Mortality (STS-PROM) score. Rates of in-hospital and 30-day mortality in the external validation cohort were 2.9% and 3.8%, respectively. The TVT Registry model was found to have moderate discrimination (c-index 0.66, 95% confidence intervals, CI, 0.60 - 0.72 and 0.67, 95% CI 0.62 - 0.72, for in-hospital and 30-day mortality, respectively) and good calibration. Compared with the STS-PROM Score, the TVT Registry model demonstrated improved calibration for in-hospital (slope 0.83, p=0.23 vs. slope 0.24, p<0.001, respectively) and 30-day (slope 1.11, p=0.40 vs slope 0.41, p<0.001, respectively) mortality.

Conclusions: In a large, multicenter, non-US cohort of TAVR patients, the validation of the TVT Registry model demonstrated moderate discrimination and good calibration for the prediction of in-hospital and 30-day mortality. As a result, the TVT Registry model should be considered an alternative to the STS-PROM score for decision-making and assessment of early outcome in patients eligible for TAVR.

Key words: mortality, prediction, transcatheter aortic valve replacement

Introduction

Fostered by refinements in device technology, improved imaging, and streamlining of the procedure, transcatheter aortic valve replacement (TAVR) plays an increasingly important role in the treatment of severe, symptomatic aortic stenosis.[1, 2] A decline in peri-procedural complications propelled expansion of TAVR to intermediate and low risk patients, and has shifted the focus of ongoing investigations to determinants of long term outcome. Risk scoring systems are instrumental to balance the expected benefits against the probability of adverse events, and represent a useful tool to properly inform physicians, counsel patients, and optimize the allocation of health care resources. In the absence of a dedicated risk score for TAVR, the Society of Thoracic Surgeons Predicted Risk of Mortality (STS-PROM) and the System for Cardiac Operative Risk Evaluation (EuroSCORE) are routinely integrated in the heart team evaluation of patients with symptomatic severe aortic stenosis. However, both scores have been derived from cohorts of surgical patients; the extrapolation to TAVR patients remains therefore challenging and their suitability arguable.[3, 4] In recent years, several attempts to develop TAVR-specific risk models have been performed.[5-10] However, the majority of these novel scores have not been validated in external cohorts, limiting their adoption in clinical practice. Because prediction models are conceived to be applied to future patients, their value depends on the performance shown outside the development sample. To date, the TVT Registry model represents the score that has been derived from the largest cohort of TAVR patients including 13,718 participants of the Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy (STS/ACC TVT) Registry.[11]

The aim of this study was to evaluate the extent of generalizability of the TVT Registry model by quantifying its performance in an independent dataset. For this purpose, we investigated its prediction accuracy in patients included in the prospective Swiss TAVI Registry.

Methods

Participants

The external validation cohort included all patients with severe native aortic valve stenosis who were consecutively treated and entered into the Swiss TAVI Registry (NCT01368250) between February 2011 and February 2016. The details of the rationale and design of the Swiss TAVI Registry have been previously described.[12] In brief, the Swiss TAVI Registry is a nationwide registry that prospectively collects clinical and procedural data of patients undergoing TAVR with CE-marked devices in Switzerland with regular follow-up at 30 days, 1 year, and yearly thereafter. A dedicated clinical committee is responsible for the adjudication of the clinical events occurring during the index hospitalization or at follow-up according to the definitions of the Valve Academic Research Consortium (VARC)-2 criteria.[13] The registry has been approved by the local ethics committee of all recruiting centers, and all patients provided written informed consent to participate.

Measurements

The TVT Registry model was applied through the automatic calculator accessible online at http://tools.acc.org/TAVRRisk/. The model includes the following variables: 1) age at admission; 2) glomerular filtration rate (eGFR), calculated on the basis of age, sex, race, pre-procedure creatinine and requirement of pre-procedure dialysis; 3) hemodialysis or peritoneal dialysis on an ongoing basis as a result of renal failure; 4) New York Heart Association (NYHA) functional class IV, defined as cardiac disease with dyspnea at rest that increases with any physical activity, resulting in inability to perform any physical activity without discomfort; 5) history of severe chronic lung disease, defined as FEV1 (forced expired volume in one second) <50% predicted and/or room air $p0_2$ <60 or room air pCO_2>50; 6) non-femoral access site; 7) acuity status 2 defined as urgent procedure status plus no pre-procedure shock, inotropes, mechanical assist device, or cardiac arrest; 8) acuity status 3 defined as elective or urgent procedure status plus pre-procedure shock, inotropes or mechanical assist device plus no prior

cardiac arrest within 24 hours of procedure; 9) acuity status 4 defined as emergency or salvage procedure or prior cardiac arrest within 24 hours of operation. Definitions used in the Swiss TAVI and the TVT Registry were similar with respect to the variables used in the model. Specifically, our registry records age at admission, dialysis status, NYHA functional class, severe chronic lung disease and femoral access. Glomerular filtration rate was calculated according to the Modification of Diet in Renal Disease (MDRD) equation and presence of dialysis. Because acuity categories are not included in the Swiss TAVI registry variables, we derived acuity status (2, 3 or 4) by matching the setting of the procedure (elective or urgent) and hemodynamic status (cardiogenic shock).

The present study complies with the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) guidelines for the reporting of studies that validate prediction scores (**Supplemental Table 1**).[14]

Statistical analysis

Patient baseline characteristics were expressed as means and standard deviation or frequencies (percentage). Validation of the TVT Registry model was performed by examining measures of discrimination and calibration. Discrimination describes the power of models to distinguish patients who have events (death) from those who have no events. It was assessed using the C-index that represents the area under the receiver operating characteristic (ROC) curve (AUC) and for which larger values are associates with better discrimination. Calibration is a measure of how closely the predicted probabilities (of death) reflect the actual risk; it was assessed by performing the Hosmer–Lemeshow goodness of fit test and was graphically depicted in the plot of observed versus predicted mortality with a value <0.05 indicating significant difference in expected versus observed mortality. Calibration was also assessed by testing for an intercept of zero and a slope of one when regressing observed proportion of deaths on predicted proportion of deaths based on the TVT. Deciles of the TVT score were used to calculate proportions. While acknowledging that the TVT Registry model was designed to predict in-hospital

mortality, we additionally tested if it could be predictive of mortality at 30 days after TAVR. Model performance in terms of calibration was also examined in pre-specified subgroups defined by age of less than 85 years or older, eGFR <60 mL/min, between 60 and 90 mL/min or greater than 90 mL/min, need for dialysis, NYHA class IV or class I to III, non-femoral access, acuity categories, and gender. The main analyses were repeated after multiple imputation of missing variables. In addition, we examined the predictive accuracy of the STS-PROM score and compared it with that of the TVT Registry Model using the DeLong method. The STS-PROM score was calculated at the time of intervention according to the models developed from the Society of Thoracic Surgeons database, available at http://riskcalc.sts.org/stswebriskcalc/.

Analyses were conducted using Stata Statistical Software: Release 14.1 (StataCorp LP, College Station, Texas) and statistical significance was defined as $p<0.05$.

Results

The validation cohort comprised 3,491 consecutive patients included into the Swiss TAVI registry between February 2011 and February 2016. In-hospital and thirty-day survival data was available for the entire cohort. Rates of in-hospital and 30-day mortality amounted to 2.9% and 3.8%, respectively. **Table 1** summarizes the baseline clinical characteristics of patients who died in hospital versus those that survived. Male and female patients were similarly represented in either group. Mean STS-PROM score was 5.8 ± 4.5 and it was significantly higher in patients who died in hospital (7.6 ± 5.9 vs. 5.8 ± 4.4, $p<0.001$). Non-survivors were older compared with survivors (84.2 ± 5.7 years vs. 82.1 ± 6.5 years, $p= 0.001$) and more often presented with renal dysfunction. In addition, non-survivors more commonly presented with cardiogenic shock (5% vs. 1%, $p<0.001$) and more often underwent urgent instead of elective TAVI (acuity category 2 or 4). Type of transcatheter heart valves used are reported in **Supplemental Table 3**. Overall, 43.6% of patients received early generation devices (Medtronic

CoreValve or Edwards Sapien XT). The comparison between validation and development cohorts in terms of demographics is reported in **Supplemental Table 2**.

Performance of the TVT Registry Model

The performance of the TVT Registry model in the Swiss TAVI cohort was assessed using the original coefficients that were obtained in the development sample. Refitted model coefficients and odds ratios with 95% CI for each covariate in the validation cohort are reported for descriptive purposes in **Supplemental Table 4.** In the Swiss TAVI Registry cohort, the TVT Registry model showed moderate discrimination, with a C-index for in-hospital mortality of 0.66, 95% CI 0.60-0.72 (**Figure 1, panel A**). Moreover, the C-index for prediction of 30-day mortality was 0.67, 95% CI 0.65-0.69 (**Figure 1, panel C**). The results were consistent when analyzing the performance of the Model among patients included in the Swiss TAVI Registry during the same period of patients included in the derivation cohort (**Supplemental Table 5**).

Calibration plots are shown in **Figure 2, panels A and C.** A close agreement between predicted versus observed mortality was documented for both in-hospital and 30-day outcome. Model calibration was preserved across several pre-specified subgroups; we recorded however an overestimation of in-hospital and 30-day mortality for patients on hemodialysis (**Figure 3**).

Performance of the STS-PROM score

As shown in **Figure 1, panel B and D**, the STS-PROM score achieved moderate discriminative ability for prediction of in-hospital (C-index: 0.61, 95% CI 0.56-0.67) and 30-day (0.63, 95% CI 0.59-0.68) mortality. **Figure 2, panel B and D** displays a separation between observed and predicted mortality rates, especially for the higher values of estimated risk.

Comparative performance of the TVT Registry Model and the STS-PROM score in the SwissTAVI Registry

Tables 2 and 3 report the comparison between the predictive accuracy of the TVT Registry model and the STS-PROM score in our population. C-index for the prediction of in-hospital and 30-day mortality were 0.66 vs. 0.61 (p= 0.14) and 0.67 vs. 0.63 (p=0.12) for the TVT Registry model and STS-PROM score, respectively. The Hosmer-Lemeshow statistics showed a better calibration ability of the TVT Registry model compared with the STS-PROM score for in-hospital (slope 0.83, p= 0.23 vs. slope 0.24, p <0.001, respectively) and 30-day (slope 1.11, p= 0.40 vs. slope 0.41, p <0.001, respectively) mortality. Discrimination of the TVT Registry model and STS-PROM score after multiple imputation of missing variables yielded comparable results (**Supplemental Table 6**).

Discussion

The main findings of our study validating the performance of the TVT Registry model in a large cohort of patients undergoing TAVR at multiple centers in Switzerland can be summarized as follows: (1) The TVT Registry model showed moderate discrimination and adequate calibration for the prediction of in-hospital mortality after TAVR; (2) its predictive accuracy was maintained for mortality at 30-days; (3) the TVT Registry model showed significantly better predictive accuracy in terms of calibration as compared to the STS-PROM score, while discrimination was comparable.

The TVT Registry model has been recently developed to predict in-hospital mortality in a cohort of more than 13,000 patients undergoing TAVR in the United States between 2011 and 2014. The internal validation cohort comprised more than 6,000 patients treated between March and October 2014. The model showed moderate discrimination with a C-index of 0.67 (95% CI 0.65-0.69) in the development group and 0.66 (95%CI 0.62-0.69) in the validation group, respectively, and good calibration. While alternative scores have been both derived and validated in relatively small cohorts, the TVT registry

model has been derived and validated in a cohort surpassing the next largest cohort used to build a risk score by a factor of 5. The time interval of patients included in the present analysis largely corresponded with the time interval of the STS/ACC TVT registry. In our study, we found a discrimination of the TVT registry model for the prediction of in-hospital mortality comparable to the original report; moreover, discrimination was maintained at 30 days after the procedure. This clearly defined time window allows for a better assessment of early outcomes after TAVR as in-hospital length of stay may be highly variable across different centers.

A risk-benefit–analysis is integral part of the Heart Team assessment for the selection of the optimal treatment strategy for patients with severe aortic valve stenosis. Clinical and anatomic characteristics complement the multidisciplinary evaluation of the patient, and are consolidated in specific scores quantifying peri-procedural risk. Risk scores allow for the possibility of comparing health across different populations. Several risk scores have proven instrumental for surgical procedures and are regularly harmonized with updated information on contemporary event rates. In the absence of a tailored risk score for TAVR, risk models originally derived from surgical cohorts have been used for the definition of risk categories and patient selection in randomized trials of TAVR versus surgical aortic valve replacement.[15-17] However, there is a large body of evidence demonstrating a suboptimal performance of such scores in TAVR cohorts. Indeed, in the Placement of Aortic Transcatheter Valves trial (PARTNER) I and continued access registry both the STS-PROM score and the Logistic EuroSCORE overestimated the mortality occurring in-hospital or at 30-days after TAVR. [3] Along the same line, in a retrospective analysis of patients treated with the Medtronic CoreValve prosthesis at two European centers, both the Logistic EuroSCORE and the STS-PROM algorithm were found to have suboptimal discriminatory power and calibration.[4] Consistently, in our cohort, the STS-PROM score showed poorer calibration among patients with higher estimated mortality risk. This finding does not only pertain to the field of TAVR but has

already been reported in surgical series.[18]Arguably, such suboptimal calibration in high-risk categories may stem from very high mortality rates in the original derivation cohort of the STS-PROM score.

More recently, several TAVR-specific risk scores have been suggested, as summarized in **Table 4**. Most scores have been validated for 30-day mortality and were found to have a C-index ranging from 0.57 to 0.75. While applicability of both the STS-PROM score and the EuroSCORE has been repeatedly questioned in view of their derivation and validation in patients with surgical access, the TVT risk model is the first among the specific TAVR risk scores to differentiate between transfemoral and alternative (surgical) approach for TAVR.

Currently available risk scores for TAVR are limited by a number of factors. Time is an important covariable rarely accounted for in conventional risk scores. A discount in risk over time has been observed for the STS-PROM score resulting in a reclassification of more than half of patients originally deemed to be high risk to intermediate risk in a repeated analysis 6 to 7 years.[19] Sensitive scores work bi-directionally: they inform about anticipated risk, while regularly being updated by most recent outcome data. This may be particularly important in a rapidly evolving field such as TAVR where device iterations have been shown to substantially reduce peri-procedural complications as reflected by a large heterogeneity of reported outcomes across major studies. Moreover, deficiencies of standard modelling methods, relatively small and homogenous derivation cohorts, and absence of validation in external datasets further hamper the robustness of existing TAVR risk scores. So far, there were no studies assessing the reproducibility and transportability of the TVT Registry model. Geographical variability in performance is mainly related to variation in case-mix, that is dissimilarity between patients in different countries.[20]In our study, the predictive accuracy of the TVT registry model was confirmed in an unselected cohort of consecutive patients treated in Switzerland. The reproducibility of the results observed in the development cohort is an important finding in view of the expected differences between the two sides of the Atlantic in terms of patient features, devices, procedural characteristics, and post-procedural care. Some concerns may arise about model performance as its discrimination was

only moderate in the original and current cohorts. However, this limitation is counterbalanced at least in part by the good calibration that was confirmed in this external cohort and preserved across several subgroups of patients. This property ensures high reliability in counseling patients and their relatives about the risk of death early after the procedure. At this regard, the considerable gain in terms of calibration of the TVT Registry model over the STS-PROM score could have important clinical implications especially when dealing with patients at the extremes of risk categories where the reliability of the STS-PROM score is poorer.

Limitations

We acknowledge the following limitations of our study: 1) although we were able to include a large contemporary TAVR population with excellent documentation of baseline and follow-up status, the TVT Registry model was validated in a retrospective manner; 2) we were unable to assess the added value of indices of frailty and measures of quality of life that were not included in the original model, as they are not systematically collected in our database; 3) the results of our validation analysis may be affected by the impossibility to quantify the case mix differences between development and validation samples as, with the exception of age and gender, no other baseline clinical characteristics of the original cohort were available; 4) although predicted versus observed mortality was consistent for both in-hospital and 30-day outcomes across several subgroups, an overestimation of in-hospital and 30-day mortality for patients on hemodialysis was observed. This should be carefully interpreted in view of multiple testing and the small number of patients included in this subgroup; 5) we were unable to assess the comparative performance of the TVT Registry Model and other risk scores as measures such as frailty, mini-mental status examination, 6-min walk test distance, assisted living, home oxygen use and Charlson Comorbidity Index are not systematically collected in our database; 6)in view of the ongoing expansion of TAVR adoption in lower risk patients, further studies are needed to validate the accuracy of this model in low-risk populations.

Conclusions

In a large, multicenter, non-US cohort of TAVR patients, the validation of the TVT Registry model demonstrated moderate discrimination and good calibration for the prediction of in-hospital and 30-day mortality. As a result, the TVT Registry model should be considered an alternative to the STS-PROM score for decision-making and assessment of early outcome in patients eligible for TAVR.

Sources of Funding

The Swiss TAVI registry is supported by a study grant from the Swiss Heart Foundation and the Swiss Working Group of Interventional Cardiology, and is sponsored by unrestricted funds from Boston Scientific, Edwards Lifesciences, Medtronic, St. Jude Medical, and Symetis. The study sponsors had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

Disclosures

Dr Pilgrim serves as a consultant to St. Jude Medical, received speakers' fees from Biotronik and Edwards Lifesciences, and received research contracts to the institution from Edwards Lifesciences and Symetis. Dr Nietlispach serves as consultant to Edwards Lifesciences and St. Jude Medical. Dr Tueller received speakers' fees from Edwards Lifesciences and travel expenses from Medtronic. Dr Toggweiler serves as a consultant for Symetis and NVT, and received speakers' fees from Symetis, Edwards Lifesciences, and Medtronic. Dr Jeger serves as a consultant to St. Jude Medical and has received reimbursement for travel expenses from Medtronic, Boston Scientific, and Edwards Lifesciences. Dr Ferrari is a proctor and consultant for Edwards Lifesciences. Dr Noble serves as consultant for Medtronic. Dr Roffi received institutional research grants from Abbott Vascular, Boston Scientific,

Biotronik, Terumo, and Medtronic. Dr Huber is a proctor for Edwards Lifesciences and Symetis, and has received speaker fees from Edwards Lifesciences, Symetis and Medtronic. Dr Windecker has received research contracts to the institution from Abbott, Boston Scientific, Biosensors, Cordis, Medtronic, and St. Jude. Dr Wenaweser serves as proctor for Medtronic, Edwards Lifesciences, and Boston Scientific and has received an unrestricted grant from Medtronic to the institution (University of Bern). All of the other authors have no conflicts of interest to declare.

References

1. Abdelghani M and Serruys PW. Transcatheter Aortic Valve Implantation in Lower-Risk Patients With Aortic Stenosis: Is It Justified to Be the Preferred Treatment? *Circ Cardiovasc Interv*. 2016; 9(4):e002944.
2. Eggebrecht H and Mehta RH. Transcatheter aortic valve implantation (TAVI) in Germany 2008-2014: on its way to standard therapy for aortic valve stenosis in the elderly? *EuroIntervention*. 2016;11:1029-33.
3. Beohar N, Whisenant B, Kirtane AJ, Leon MB, Tuzcu EM, Makkar R, Svensson LG, Miller DC, Smith CR, Pichard AD, Herrmann HC, Thourani VH, Szeto WY, Lim S, Fischbein M, Fearon WF, O'Neill W, Xu K, Dewey T and Mack M. The relative performance characteristics of the logistic European System for Cardiac Operative Risk Evaluation score and the Society of Thoracic Surgeons score in the Placement of Aortic Transcatheter Valves trial. *J Thorac Cardiovasc Surg*. 2014;148:2830-7 e1.
4. Piazza N, Wenaweser P, van Gameren M, Pilgrim T, Tzikas A, Otten A, Nuis R, Onuma Y, Cheng JM, Kappetein AP, Boersma E, Juni P, de Jaegere P, Windecker S and Serruys PW. Relationship between the logistic EuroSCORE and the Society of Thoracic Surgeons Predicted Risk of Mortality score in patients implanted with the CoreValve ReValving system--a Bern-Rotterdam Study. *Am Heart J*. 2010;159:323-9.
5. Seiffert M, Sinning JM, Meyer A, Wilde S, Conradi L, Vasa-Nicotera M, Ghanem A, Kempfert J, Hammerstingl C, Ojeda FM, Kim WK, Koschyk DH, Schirmer J, Baldus S, Grube E, Mollmann H, Reichenspurner H, Nickenig G, Blankenberg S, Diemert P, Treede H, Walther T, Werner N and Schnabel RB. Development of a risk score for outcome after transcatheter aortic valve implantation. *Clin Res Cardiol*. 2014;103:631-40.
6. Iung B, Laouenan C, Himbert D, Eltchaninoff H, Chevreul K, Donzeau-Gouge P, Fajadet J, Leprince P, Leguerrier A, Lievre M, Prat A, Teiger E, Laskar M, Vahanian A, Gilard M and Investigators F. Predictive factors of early mortality after transcatheter aortic valve implantation: individual risk assessment using a simple score. *Heart*. 2014;100:1016-23.
7. Debonnaire P, Fusini L, Wolterbeek R, Kamperidis V, van Rosendael P, van der Kley F, Katsanos S, Joyce E, Tamborini G, Muratori M, Gripari P, Bax JJ, Marsan NA, Pepi M and Delgado V. Value of the "TAVI2-SCORe" versus surgical risk scores for prediction of one year mortality in 511 patients who underwent transcatheter aortic valve implantation. *Am J Cardiol*. 2015;115:234-42.
8. Capodanno D, Barbanti M, Tamburino C, D'Errigo P, Ranucci M, Santoro G, Santini F, Onorati F, Grossi C, Covello RD, Capranzano P, Rosato S, Seccareccia F and Group OR. A simple risk tool (the OBSERVANT score) for prediction of 30-day mortality after transcatheter aortic valve replacement. *Am J Cardiol*. 2014;113:1851-8.
9. Arnold SV, Reynolds MR, Lei Y, Magnuson EA, Kirtane AJ, Kodali SK, Zajarias A, Thourani VH, Green P, Rodes-Cabau J, Beohar N, Mack MJ, Leon MB, Cohen DJ and Investigators P. Predictors of poor outcomes after transcatheter aortic valve replacement: results from the PARTNER (Placement of Aortic Transcatheter Valve) trial. *Circulation*. 2014;129:2682-90.
10. Hermiller JB, Jr., Yakubov SJ, Reardon MJ, Deeb GM, Adams DH, Afilalo J, Huang J, Popma JJ and CoreValve United States Clinical I. Predicting Early and Late Mortality After Transcatheter Aortic Valve Replacement. *J Am Coll Cardiol*. 2016;68:343-52.
11. Edwards FH, Cohen DJ, O'Brien SM, Peterson ED, Mack M, Shahian DM, Grover F-L, Tuzcu EM, Thourani VH, Carroll J, Brennan JM, Brindis RG, Rumsfeld J and Holmes DR. Development and validation of a risk prediction model for in-hospital mortality after transcatheter aortic valve replacement. *JAMA Cardiol 2016;1(1):46-52*.
12. Wenaweser P, Stortecky S, Heg D, Tueller D, Nietlispach F, Falk V, Pedrazzini G, Jeger R, Reuthebuch O, Carrel T, Raber L, Amann FW, Ferrari E, Toggweiler S, Noble S, Roffi M, Gruenenfelder J, Juni P, Windecker S and Huber C. Short-term clinical outcomes among patients undergoing

transcatheter aortic valve implantation in Switzerland: the Swiss TAVI registry. *EuroIntervention*. 2014;10:982-9.
13. Kappetein AP, Head SJ, Genereux 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, Rodes-Cabau J, Vranckx P, Webb JG, Windecker S, Serruys PW and Leon MB. Updated standardized endpoint definitions for transcatheter aortic valve implantation: the Valve Academic Research Consortium-2 consensus document. *J Am Coll Cardiol*. 2012;60:1438-54.
14. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF and Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162:W1-73.
15. Smith CR, Leon MB, Mack MJ, Miller DC, Moses JW, Svensson LG, Tuzcu EM, Webb JG, Fontana GP, Makkar RR, Williams M, Dewey T, Kapadia S, Babaliaros V, Thourani VH, Corso P, Pichard AD, Bavaria JE, Herrmann HC, Akin JJ, Anderson WN, Wang D, Pocock SJ and Investigators PT. Transcatheter versus surgical aortic-valve replacement in high-risk patients. *N Engl J Med*. 2011;364:2187-98.
16. Makkar RR, Fontana GP, Jilaihawi H, Kapadia S, Pichard AD, Douglas PS, Thourani VH, Babaliaros VC, Webb JG, Herrmann HC, Bavaria JE, Kodali S, Brown DL, Bowers B, Dewey TM, Svensson LG, Tuzcu M, Moses JW, Williams MR, Siegel RJ, Akin JJ, Anderson WN, Pocock S, Smith CR, Leon MB and Investigators PT. Transcatheter aortic-valve replacement for inoperable severe aortic stenosis. *N Engl J Med*. 2012;366:1696-704.
17. Herrmann HC, Thourani VH, Kodali SK, Makkar RR, Szeto WY, Anwaruddin S, Desai N, Lim S, Malaisrie SC, Kereiakes DJ, Ramee S, Greason KL, Kapadia S, Babaliaros V, Hahn RT, Pibarot P, Weissman NJ, Leipsic J, Whisenant BK, Webb JG, Mack MJ, Leon MB and Investigators P. One-Year Clinical Outcomes With SAPIEN 3 Transcatheter Aortic Valve Replacement in High-Risk and Inoperable Patients With Severe Aortic Stenosis. *Circulation*. 2016;134:130-40.
18. Thourani VH, Suri RM, Gunter RL, Sheng S, O'Brien SM, Ailawadi G, Szeto WY, Dewey TM, Guyton RA, Bavaria JE, Babaliaros V, Gammie JS, Svensson L, Williams M, Badhwar V and Mack MJ. Contemporary real-world outcomes of surgical aortic valve replacement in 141,905 low-risk, intermediate-risk, and high-risk patients. *Ann Thorac Surg*. 2015;99:55-61.
19. Rogers T, Koifman E, Patel N, Gai J, Torguson R, Corso P and Waksman R. Society of Thoracic Surgeons Score Variance Results in Risk Reclassification of Patients Undergoing Transcatheter Aortic Valve Replacement. *JAMA Cardiol*. 2017;2:455-456.
20. Altman DG, Vergouwe Y, Royston P and Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605.

Table 1. Baseline characteristics of the entire validation cohort and stratified according to in-hospital mortality

	All patients n= 3,491	**Survivors** n= 3,390	**Died in-hospital** n= 101	**p value**
Model covariates				
Age (years)	82.1 ± 6.5	82.1 ± 6.5	84.2 ± 5.7	0.001
STS-PROM score	5.8 ± 4.5	5.8 ± 4.4	7.6 ± 5.9	0.001
Gender				0.10
Male	1,760 (50%)	1,701 (50%)	59 (58%)	
Female	1,731 (50%)	1,689 (50%)	42 (42%)	

Dialysis				0.27
No	3,406 (98%)	3,309 (98%)	97 (96%)	
Yes	81 (2%)	77 (2%)	4 (4%)	
Severe chronic lung disease				0.52
No	3045 (87%)	2959 (87%)	86 (85%)	
Yes	445 (13%)	430 (13%)	15 (15%)	
NYHA functional class				0.088
I	313 (9%)	307 (9%)	6 (6%)	
II	852 (25%)	835 (25%)	17 (17%)	
III	1,848 (54%)	1,790 (54%)	58 (59%)	
IV	401 (12%)	384 (12%)	17 (17%)	
Cardiogenic shock (class Killip 4)				<0.001
No	3,457 (99%)	3,361 (99%)	96 (95%)	
Yes	34 (1%)	29 (1%)	5 (5%)	
eGFR (mL/min)	63.7 ± 26.0	63.9 ± 26.0	55.4 ± 24.8	0.001
Access				<0.001
Femoral	3,045 (87%)	2,971 (88%)	74 (73%)	
Transapical	357 (10%)	337 (10%)	20 (20%)	
Subclavian	34 (1%)	32 (1%)	2 (2%)	
Direct aortic	34 (1%)	30 (1%)	4 (4%)	
Other	21 (1%)	20 (1%)	1 (1%)	
Acuity category				<0.001
1	3,370 (97%)	3,281 (97%)	89 (88%)	
2	87 (2%)	80 (2%)	7 (7%)	
4	34 (1%)	29 (1%)	5 (5%)	
TVT score	3.9 ± 3.1	3.9 ± 2.9	6.1 ± 5.7	<0.001
STS-PROM score	4.4 (3.0 - 7.0)	4.4 (2.9 - 7.0)	5.4 (3.6 - 10.0)	<0.001

Values are mean ±SD or medians (25%-75% interquartile ranges).. eGFR, Estimated glomerular filtration rate. Definition of acuity categories is provided in the text.

Table 2. Discrimination of the TVT Registry model and the STS-PROM score

	AUC (95% CI)	TVT Registry model vs. STS-PROM score p value
In-hospital mortality		0.14
TVT Registry model	0.66 (0.60 - 0.72)	
STS-PROM score	0.61 (0.56 - 0.67)	
30-day mortality		0.12
TVT Registry model	0.67 (0.62 - 0.72)	
STS-PROM score	0.63 (0.59 - 0.68)	

Table 3. Calibration of the TVT Registry model and the STS-PROM score

	Hosmer-Lemeshow test (p-value)	Intercept	p-value*	Slope	p-value**
TVT Registry model					
In-hospital mortality	0.15	-0.00 (-0.02 - 0.01)	0.517	0.83 (0.58 - 1.08)	0.23
30-day mortality	0.36	-0.01 (-0.02 - 0.01)	0.323	1.11 (0.87 - 1.34)	0.40
STS-PROM score					
In-hospital mortality	0.58	0.01 (0.01 - 0.02)	0.006	0.24 (0.14 - 0.35)	< 0.001
30-day mortality	0.58	0.01 (0.01 - 0.02)	0.003	0.41 (0.32 - 0.50)	< 0.001

*Null hypothesis, calibration plot intercept = 0. **Null hypothesis, calibration plot slope = 1.

Table 4. Main features of the currently available TAVI risk scores							
Score (Author, Year)	**FRANCE-2 (Iung, 2014)**	**TARIS (Seiffert 2014)**	**OBSERVANT (Capodanno, 2014)**	**Predictor of poor outcomes (Arnold et al, 2014)**	**$TAVI_2$ (Debonnaire, 2015)**	**CoreValve U.S. Program (Hermiller, 2016)**	**TVT Registry model (Edwards, 2015)**
Population	FRANCE-2 Registry Derivation cohort, n= 2,552 Validation cohort, n= 1,281	GARY Registry Derivation cohort, n= 845 Validation cohort, n= 333	OBSERVANT Study Derivation cohort, n= 1,256 Validation cohort, n= 622	PARTNER program Derivation cohort, n= 1,420 Validation cohort, n= 717	Patients treated at two centers (The Netherlands and Italy) Derivation cohort, n= 511	Medtronic CoreValve U.S. Pivotal Trial Derivation cohort, n= 2,482 Validation cohort, n= 1,205	STS/ACC TVT Registry Derivation cohort, n= 13,718 Validation cohort, n= 6,868
Variables	BMI <30; NYHA Class IV; Respiratory insufficiency; Pulmonary hypertension; ≥ 2 Episodes of pulmonary edema during past year; Critical hemodynamic state; Dialysis	BMI; Estimated GFR; Hemoglobin; Pulmonary hypertension; Mean transvalvular gradient; LVEF	GFR <45 ml/min; Critical pre-operative state; NYHA class IV; Pulmonary artery hypertension; Diabetes; Prior BAV; LVEF < 40%	Male sex; Diabetes; Major arrhythmia; Serum creatinine; Mean arterial pressure; Body mass index; Oxygen-dependent lung disease; Mean aortic valve gradient; Mini-Mental Status examination; 6-Min Walk Test Distance	Age >85 yrs; Male; Porcelain Aorta; Recent MI (<90 days); CrCl <30 ml/kg/min; Hemoglobin <10 g/dl; LVEF <35%; Baseline AVMG ≥70 mm Hg	Albumin ≤3.3 g/dl; Assisted living; Home oxygen; Age >85 yrs Albumin ≤3.3 g/dl; Seve Charlson score; Home oxygen; STS >7%	Age; NYHA class IV; Chronic lung disease (severe); Acuity (3 levels); Dialysis or glomerular filtration rate; Nonfemoral approach
Predicted Outcomes	30-day mortality	30-day mortality	30-day mortality	Death, KCCQ-OS score <45, or ≥10-point decrease in KCCQ-OS score compared with baseline at 6-month and 1-year	1-year mortality	30-day mortality 1-year mortality	In-hospital mortality

C-index	0.67	0.57	0.71	0.66	0.71	0.75 (30-day); 0.79 (1-year)	0.66

Figure Legend

Figure 1. Receiving operating curve (ROC) for prediction of in-hospital and 30-day mortality of the TVT Registry model (A and C) and the STS-PROM score (B and D).

Figure 2. Calibration plots showing the predicted (*x*-axis) probability versus observed (*y*-axis) in-hospital and 30-day mortality after transcatheter aortic valve replacement for the TVT Registry model (A and C, respectively) and the STS-PROM score (B and D). The diagonal line represents the perfect calibration (observed = calibration). Observed mortality is represented with 95% CI (error bars).

Figure 3. Observed (closed circle) versus predicted (open circle) in-hospital and 30-day mortality across pre-specified subgroups.

Figure 1

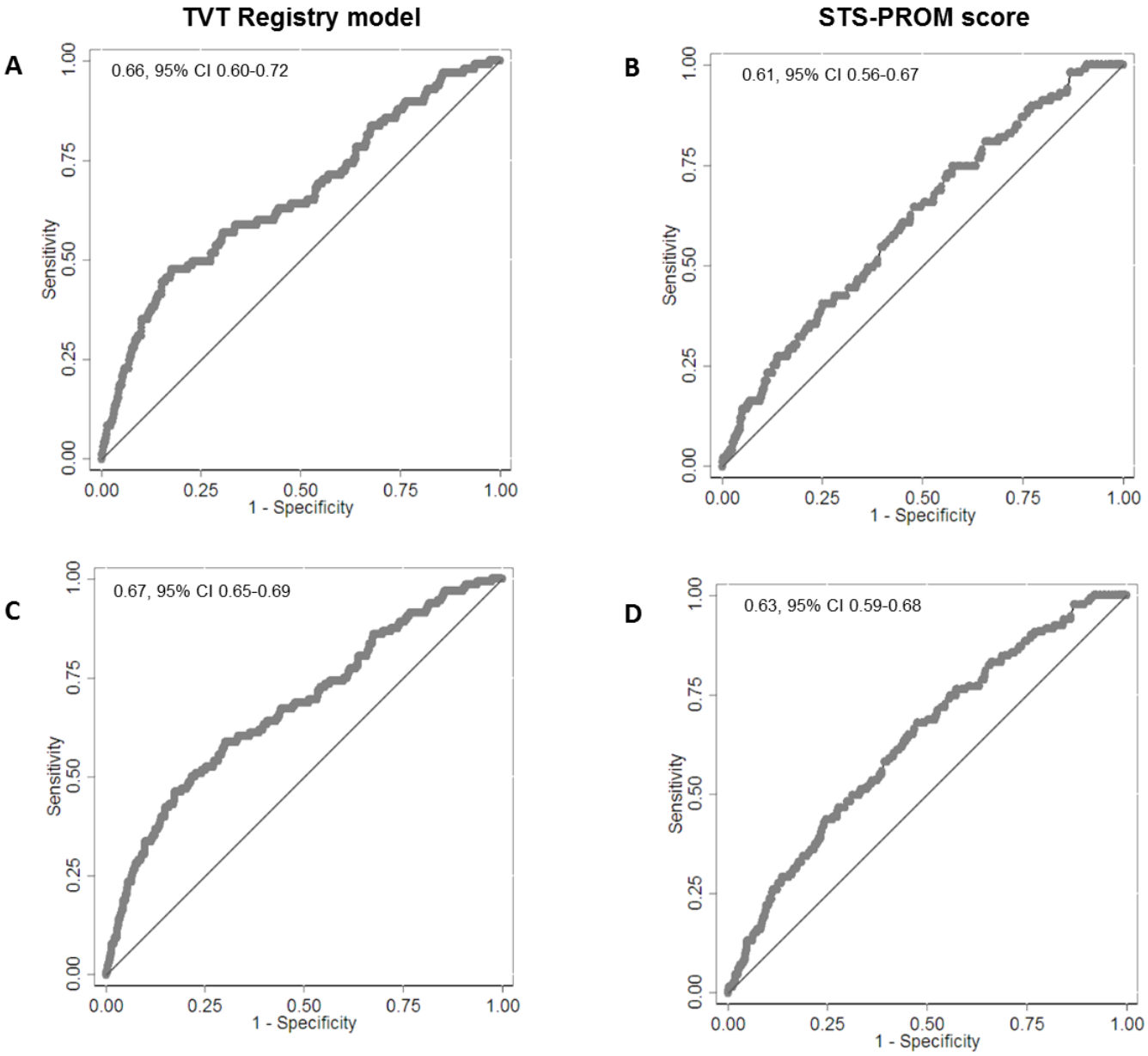

TVT Registry model
A
0.66, 95% CI 0.60-0.72
Sensitivity
1 - Specificity
0.00
0.25
0.50
0.75
1.00
STS-PROM score
B
0.61, 95% CI 0.56-0.67
C
0.67, 95% CI 0.65-0.69
D
0.63, 95% CI 0.59-0.68

Figure 2

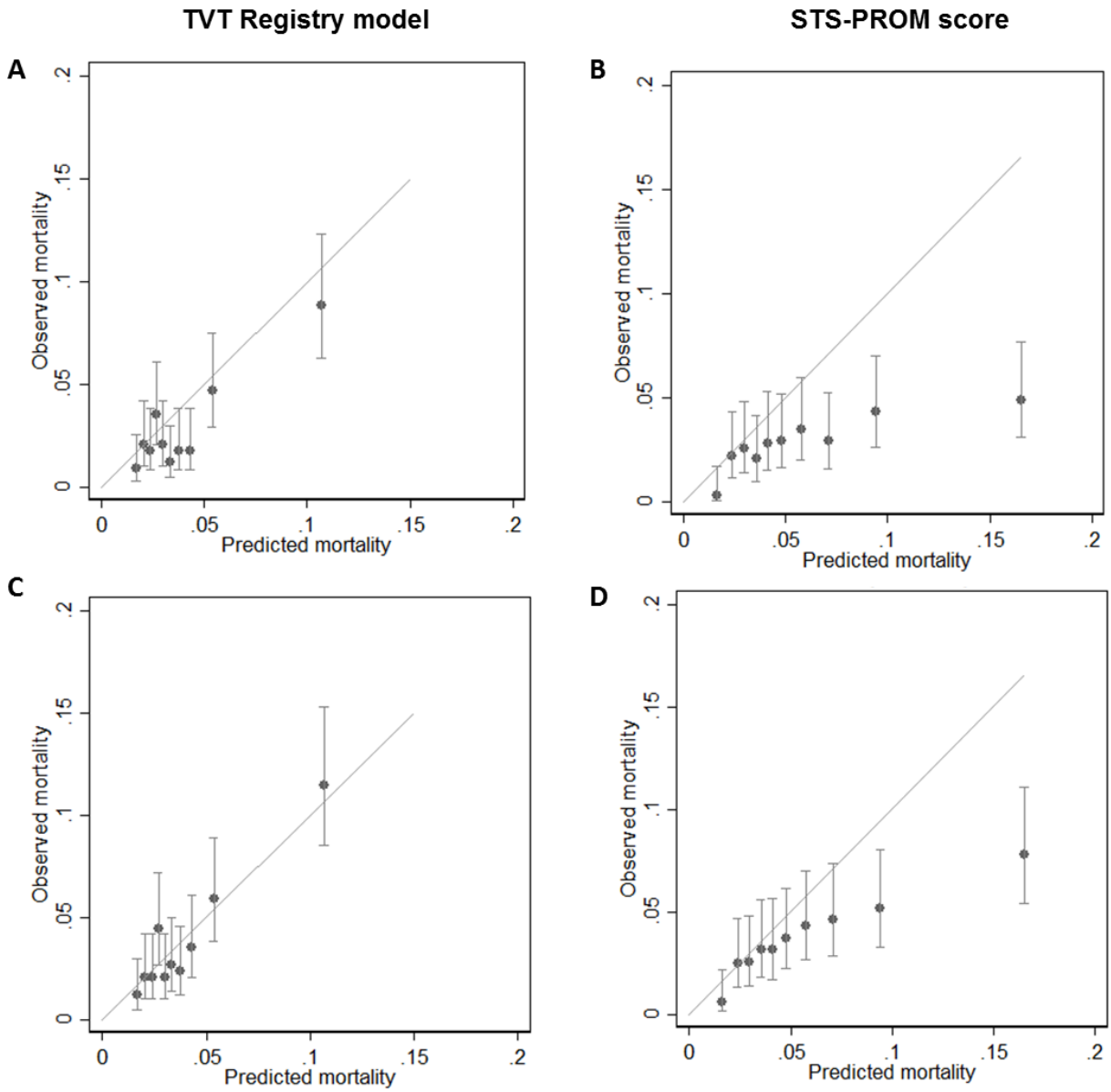
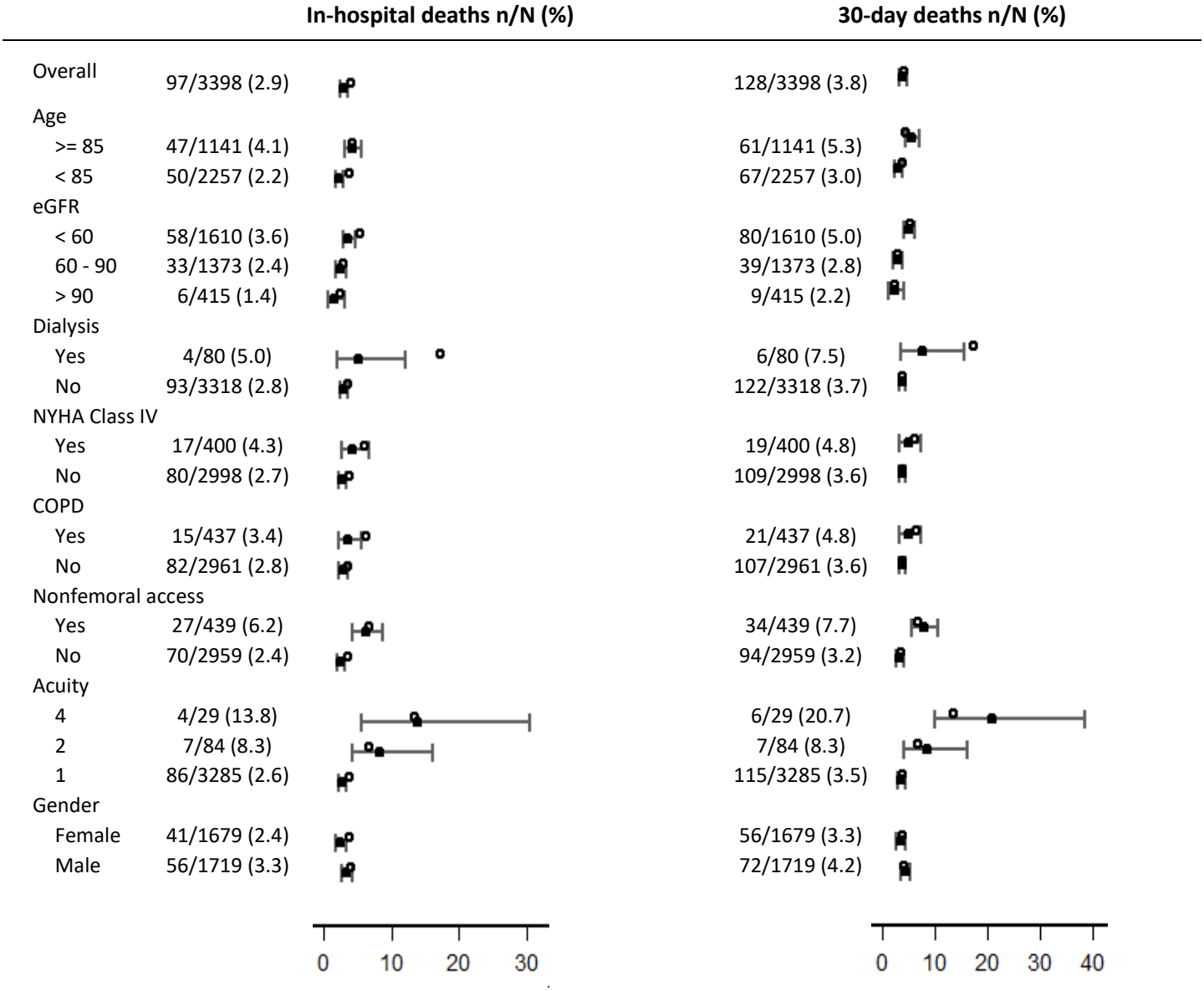

TVT Registry model
STS-PROM score
A
B
C
D
Observed mortality
Predicted mortality
0
.05
.1
.15
.2

Figure 3


In-hospital deaths n/N (%)
30-day deaths n/N (%)
Overall
97/3398 (2.9)
128/3398 (3.8)
Age
>= 85
47/1141 (4.1)
61/1141 (5.3)
< 85
50/2257 (2.2)
67/2257 (3.0)
eGFR
< 60
58/1610 (3.6)
80/1610 (5.0)
60 - 90
33/1373 (2.4)
39/1373 (2.8)
> 90
6/415 (1.4)
9/415 (2.2)
Dialysis
Yes
4/80 (5.0)
6/80 (7.5)
No
93/3318 (2.8)
122/3318 (3.7)
NYHA Class IV
Yes
17/400 (4.3)
19/400 (4.8)
No
80/2998 (2.7)
109/2998 (3.6)
COPD
Yes
15/437 (3.4)
21/437 (4.8)
No
82/2961 (2.8)
107/2961 (3.6)
Nonfemoral access
Yes
27/439 (6.2)
34/439 (7.7)
No
70/2959 (2.4)
94/2959 (3.2)
Acuity
4
4/29 (13.8)
6/29 (20.7)
2
7/84 (8.3)
7/84 (8.3)
1
86/3285 (2.6)
115/3285 (3.5)
Gender
Female
41/1679 (2.4)
56/1679 (3.3)
Male
56/1719 (3.3)
72/1719 (4.2)
0
10
20
30
0
10
20
30
40