EuroIntervention 2020;15:1497-1505 published online October 2019

Systematic review and meta-analysis of current risk models in predicting short-term mortality after transcatheter aortic valve replacement

Tariq Jamal Siddiqi[1], MBBS; Muhammad Shariq Usman[1], MBBS; Muhammad Shahzeb Khan[2], MD; Muhammad Arbaz Arshad Khan[1], MBBS; Haris Riaz[3], MD; Safi U. Khan[4], MD; M. Hassan Murad[5], MD; Clifford J. Kavinsky[6], MD; Rami Doukky[2,6], MD; Ankur Kalra[3]*, MD; Milind Y. Desai[3], MD; Deepak L. Bhatt[7], MD, MPH

1. Department of Medicine, Dow University of Health Sciences, Karachi, Pakistan; 2. Department of Medicine, John H. Stroger, Jr. Hospital of Cook County, Chicago, IL, USA; 3. Department of Cardiovascular Medicine, Cleveland Clinic, Cleveland, OH, USA; 4. Department of Medicine, West Virginia University, Morgantown, WV, USA; 5. Evidence-Based Practice Center, Mayo Clinic, Rochester, NY, USA; 6. Division of Cardiology, Rush University Medical Center, Chicago, IL, USA; 7. Department of Cardiovascular Medicine, Brigham and Women's Hospital Heart & Vascular Center, Boston, MA, USA

T.J. Siddiqi and M.S. Usman contributed equally to this manuscript.

This paper also includes supplementary data published online at: https://eurointervention.pcronline.com/doi/10.4244/EIJ-D-19-00636

KEYWORDS

- aortic stenosis
- death
- TAVI

Abstract

Aims: The aim of this study was to evaluate the performance of risk stratification models (RSMs) in predicting short-term mortality after transcatheter aortic valve replacement (TAVR).

Methods and results: MEDLINE and Scopus were queried to identify studies which validated RSMs designed to assess 30-day or in-hospital mortality after TAVR. Discrimination and calibration were assessed using C-statistics and observed/expected ratios (OERs), respectively. C-statistics were pooled using a random-effects inverse-variance method, while OERs were pooled using the Peto odds ratio. A good RSM is defined as one with a C-statistic >0.7 and an OER close to 1.0. Twenty-four studies (n=68,215 patients) testing 11 different RSMs were identified. Discrimination of all RSMs was poor (C-statistic <0.7); however, certain TAVR-specific RSMs such as the in-hospital STS/ACC TVT (C-statistic=0.65) and STT (C-statistic=0.66) predicted individual mortality more reliably than surgical models (C-statistic range=0.59-0.61). A good calibration was demonstrated by the in-hospital STS/ACC TVT (OER=0.99), 30-day STS/ACC TVT (OER=1.08) and STS (OER=1.01) models. Baseline dialysis (OER: 2.64 [1.88, 3.70]; p<0.001) was the strongest predictor of mortality.

Conclusions: This study demonstrates that the STS/ACC TVT model (in-hospital and 30-day) and the STS model have accurate calibration, making them useful for comparison of centre-level risk-adjusted mortality. In contrast, the discriminative ability of currently available models is limited.

**Corresponding author: Department of Cardiovascular Medicine, Heart and Vascular Institute, Cleveland Clinic, 224 West Exchange St, Suite 225, Akron, OH 44302, USA. E-mail: kalramd.ankur@gmail.com*

SUBMITTED ON 07/07/2019 - REVISION RECEIVED ON 12/09/2019 - ACCEPTED ON 23/10/2019

DOI: 10.4244/EIJ-D-19-00636

Abbreviations

AHA	American Heart Association
AUC	area under the curve
CIs	confidence intervals
ESC/EACTS	European Society of Cardiology/European Association for Cardiothoracic Surgery
EuroSCORE	European System for Cardiac Operative Risk Evaluation
FRANCE 2	FRench Aortic National CoreValve and Edwards
German AV Score	German aortic valve score
GS	Guaragna score
MeSH	medical subject heading
NYHA	New York Heart Association
OBSERVANT	OBservational Study Of Appropriateness, Efficacy and Effectiveness of AVR-TAVR Procedures For the Treatment Of Severe Symptomatic Aortic Stenosis
OERs	observed/expected ratios
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROBAST	Prediction model Risk Of Bias ASsessment Tool
RSMs	risk stratification models
SAVR	surgical aortic valve replacement
STS/ACC TVT	Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy
STS	Society of Thoracic Surgeons
STT	survival post TAVI
TAVR	transcatheter aortic valve replacement
UK TAVI CPM	UK transcatheter aortic valve implantation clinical prediction models

Introduction

The European Society of Cardiology/European Association for Cardiothoracic Surgery (ESC/EACTS) guidelines recommend transcatheter aortic valve replacement (TAVR) instead of surgical aortic valve replacement (SAVR) to improve survival and/or symptoms in patients with aortic stenosis who are at intermediate to high surgical risk[1]. Recent evidence suggests that the recommendation for TAVR might be extended to low surgical risk patients as well[2]. Although the use of TAVR is increasing, selection for TAVR of candidates in whom the expected benefits of the intervention outweigh the risks remains a challenge. Accurate risk stratification models (RSMs) can aid this process by determining the probability of a futile procedure, thereby helping to avoid hopeless procedures and simplifying treatment decisions. Initially, surgical RSMs such as the Society of Thoracic Surgeons (STS) score and the European System for Cardiac Operative Risk Evaluation (EuroSCORE) were used for this purpose[3]. However, their prognostic value has been questioned, and concerns have been raised that they tend to overestimate mortality risk.

Consequently, multiple RSMs have been developed from TAVR populations; however, their reliability is not well established, and it remains unclear which of these RSMs is optimal for clinical use[4-10]. Furthermore, the external generalisability of these models is limited given the heterogeneous patient populations, procedural and operator-specific factors. Therefore, pooling data from different validation studies can provide a more accurate assessment of the performance of the RSM compared to individual studies. The purpose of this study was to analyse systematically the clinical practicability, productiveness and discriminative performance of each RSM by conducting a meta-analysis using data from all studies validating the particular RSM. Furthermore, we aimed to assess whether TAVR-dedicated risk scores are superior to surgical risk scores in predicting survival. In addition, we sought to review the predictors used by each RSM and evaluate which patient-specific parameters were the best predictors of post-TAVR mortality.

Methods

This systematic review and meta-analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[11].

Details on search strategy **(Supplementary Table 1)**, study selection, data extraction and quality assessment are provided in **Supplementary Appendix 1**[5,10].

EFFECT SIZE ESTIMATION

Discrimination and calibration are relative and absolute measures, respectively, that are essential to have in a useful and reliable RSM. Discrimination is defined as the ability of an RSM to yield a higher "risk" for individuals who experience an event in the future, when compared with patients who do not experience the event. To evaluate discrimination, we used the C-statistic (also known as "area under the curve" [AUC]). The C-statistic ranges from 1.0 (perfect concordance between model-based risk estimates and observed events) to 0.5 (random concordance). C-statistic values have been categorised as follows: (a) 0.81-0.90 = good; (b) 0.71-0.80 = fair; (c) 0.61-0.70 = poor; and (d) 0.50-0.60 = very poor/almost no association[12]. For this meta-analysis, C-statistics and their corresponding 95% confidence intervals (CIs) were extracted from each validation study. The 95% CIs were used to compute standard errors (SEs).

Calibration is the measure of how accurately the model's predictions match overall observed events in a cohort of patients (observed/expected ratio [OER]). OERs of ~1 suggest good calibration. OERs >1 suggest underprediction, while ratios <1 suggest overprediction. From each study, we extracted the expected mortality (as predicted by the risk model) and the observed (actual) mortality. These values were then used to compute the observed – expected (O-E) value and the variance, using an online calculator (http://www.hutchon.net/peto%20vers%202.html).

STATISTICAL ANALYSIS

We performed a meta-analysis on the C-statistics and corresponding SEs using an inverse variance random-effects model to determine

the pooled discrimination. Before pooling, logit transformation of the C-statistic values was carried out. The OER and variance were measured using the Peto odds ratio method. The OERs from each study validating a particular model were pooled together for accurate estimation of the calibration of that scale. Log transformation of the OER values was done prior to pooling. We also sought to assess the association of specific predictors with short-term mortality. A covariate was selected for meta-analysis if data (odds ratios [OR] and 95% CIs) on it were provided by at least two studies. Q statistics and Higgins I^2 were used to evaluate heterogeneity across studies and a value of I^2=25%-50% was considered mild, 50%-75% as moderate, and >75% as severe. A p-value of <0.05 was considered significant for all analyses. Review Manager, Version 5.5 (Cochrane Collaboration, Oxford, UK) was used to perform the statistical analyses.

Results

SEARCH RESULTS

The initial search produced 6,099 articles; 2,930 were reviewed at title and abstract level and an additional 2,906 articles were removed based on predetermined selection criteria. Ultimately, we included 23 full articles and one abstract (Sirotina al. Utility of conventional surgical risk scores in predicting outcome after transcatheter aortic valve replacement, presented at American College of Cardiology (ACC) 2013 Scientific Sessions, 9 March 2013, San Francisco, CA, USA). A total of 68,125 patients from these studies were included in the analysis **(Figure 1)**[4-10,13-29]. These 24 studies tested 11 different RSMs (7 TAVR-specific, 3 surgical, and 1 designed for use in both TAVR and SAVR patients). **Supplementary Table 2** provides a list of all included studies along with relevant study characteristics. **Supplementary Table 3** displays the predictors that make up each included RSM. Assessment of risk of bias using the PROBAST scale revealed that all the new TAVR-specific models were developed using robust methodological methods **(Supplementary Table 4)**. Similarly, all of these models were found to have good applicability except for the UK TAVI CPM, which was adjudicated to have low applicability as it was derived from a small, selected population.

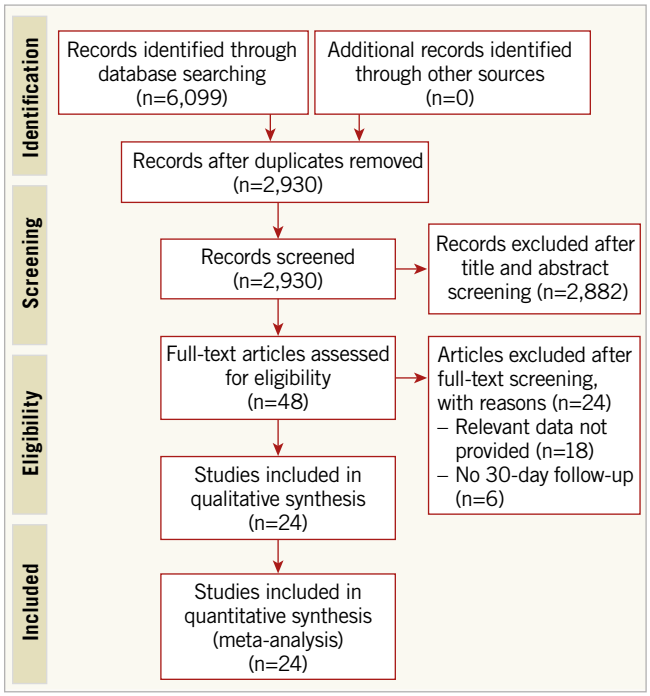


Figure 1. *PRISMA flow chart outlining the literature search.*

The summarised forest plots display the pooled discrimination **(Figure 2)** and calibration **(Figure 3)** of each RSM. The detailed forest plots are provided in **Supplementary Figure 1-Supplementary Figure 4**.

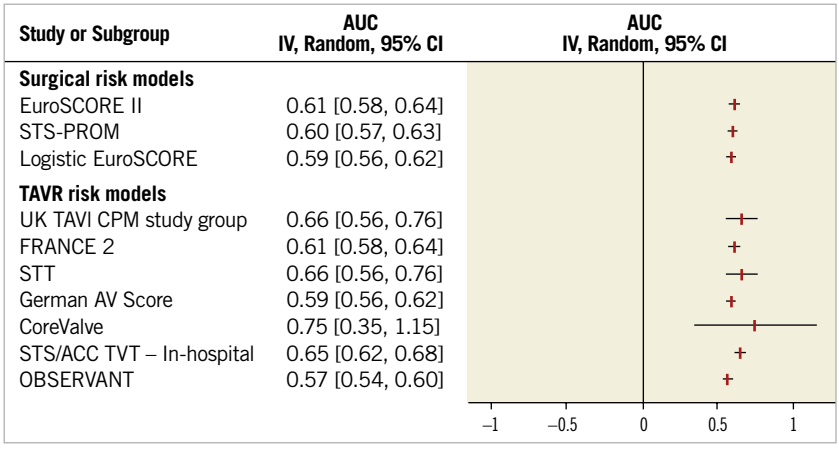


Figure 2. *Summarised forest plot displaying results of meta-analysis of discrimination of each risk stratification model. AUC: area under the curve; FRANCE 2: FRench Aortic National CoreValve and Edwards; German AV Score: German aortic valve score; OBSERVANT: Observational Study Of Appropriateness, Efficacy And Effectiveness of AVR-TAVR Procedures For the Treatment Of Severe Symptomatic Aortic Stenosis; STS/ACC TVT: Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy; STS-PROM: Society of Thoracic Surgeons Predicted Risk of Mortality; STT: survival posT TAVI; UK TAVI CPM: UK transcatheter aortic valve implantation clinical prediction models*

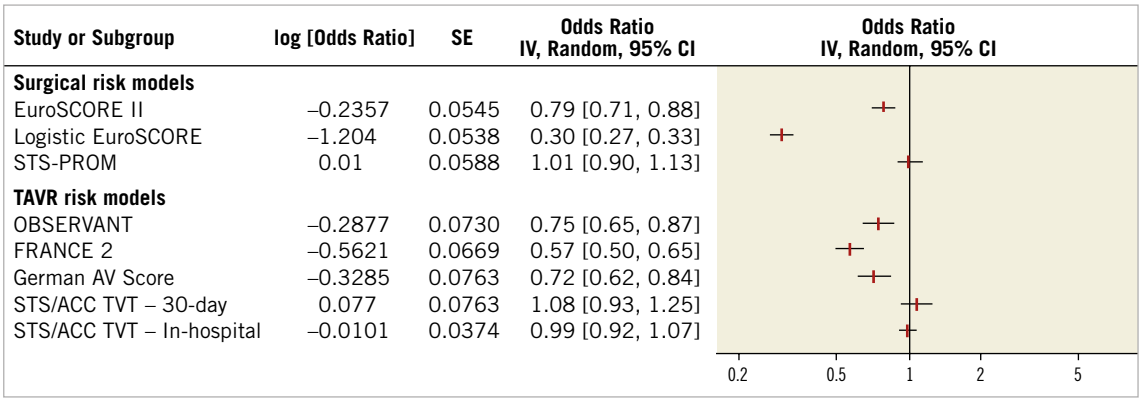


Figure 3. *Summarised forest plot displaying results of meta-analysis of calibration of each risk stratification model. AUC: area under the curve; FRANCE 2: FRench Aortic National CoreValve and Edwards; German AV Score: German aortic valve score; OBSERVANT: Observational Study Of Appropriateness, Efficacy And Effectiveness of AVR-TAVR Procedures For the Treatment Of Severe Symptomatic Aortic Stenosis; STS/ACC TVT: Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy; STS-PROM: Society of Thoracic Surgeons Predicted Risk of Mortality*

TAVR-SPECIFIC MODELS

STS/ACC TVT

Meta-analysis of 2016 and 2018 in-hospital risk models demonstrated a C-statistic of 0.65 (95% CI: 0.62-0.68; I^2=0%) and an OER of 0.99 (95% CI: 0.92-1.07; I^2=82%), indicating poor discrimination and good calibration, respectively. We could not estimate the discrimination of the 30-day model due to lack of data. The OER for this model was 1.08 (95% CI: 0.93-1.25). The 30-day mortality model has not yet been externally validated as of March 2019.

OBSERVANT

The model was found to have a poor discrimination (C-statistic: 0.57; 95% CI: 0.54-0.60; I^2=0%) and a significantly over-predictive calibration (OER: 0.75; 95% CI: 0.65, 0.87).

FRANCE 2

The pooled results demonstrated poor discrimination (C-statistic: 0.61; 95% CI: 0.59-0.64; I^2=13%). The calibration of the scale was found to be significantly over-predictive for 30-day mortality (OER: 0.57; 95% CI: 0.50-0.65; I^2=0%).

COREVALVE

This model demonstrated a fair discriminative ability (C-statistic: 0.75; 95% CI: 0.35-1.15); however, a wide confidence interval makes this result unreliable. The OER was not reported by the single study validating this model. To the best of our knowledge, this RSM has not been externally validated.

STT (SURVIVAL POST TAVI)

The STT model demonstrated poor discriminative ability (C-statistic: 0.66; 95% CI: 0.56-0.76). The OER was not reported. Our search revealed no studies which externally validated this model and met the inclusion criteria.

UK TAVI CPM

This model demonstrated a poor discriminative ability (C-statistic: 0.66; 95% CI: 0.61-0.71). The OER was not reported in the publication in which this model was derived and validated. This model has not yet been validated in an external sample.

GERMAN AV SCORE

This model showed a very poor discrimination (C-statistic: 0.59; 95% CI: 0.56-0.62) and a significantly over-predictive calibration (OER: 0.72; 95% CI: 0.62-0.82).

SAVR-SPECIFIC MODELS

STS SCORE

This surgical risk model showed a poor discrimination (C-statistic: 0.60; 95% CI: 0.58-0.64; I^2=34%); however, the calibration was good (OER: 1.01; 95% CI: 0.90-1.13; I^2=70%).

LOGISTIC EUROSCORE

This showed very poor discrimination (C-statistic: 0.59; 95% CI: 0.56-0.62; I^2=54%). Similarly, this model showed a significantly over-predictive calibration (OER: 0.30; 95% CI: 0.27-0.33; I^2=88%).

EUROSCORE II

This model showed poor discrimination (C-statistic: 0.61; 95% CI: 0.58-0.64; I^2=30%). The calibration of this model was over-predictive (OER: 0.79; 95% CI: 0.71-0.88; I^2=80%).

P-INTERACTION BETWEEN SUBGROUPS

The overall p-interactions for both discrimination (p=0.03) and calibration (p<0.001) signify significant differences between subgroups. **Supplementary Table 5** and **Supplementary Table 6** give p-interaction values between individual subgroup pairs in the discrimination and calibration analysis, respectively.

PREDICTORS OF SHORT-TERM MORTALITY

Baseline dialysis was the strongest predictor of short-term mortality (OR: 2.64 [1.88, 3.71]; p<0.001; I^2=0%). **Figure 4** displays all the predictors studied.

Discussion

This meta-analysis of 68,215 patients shows that RSMs designed specifically for TAVR patients show poor discrimination

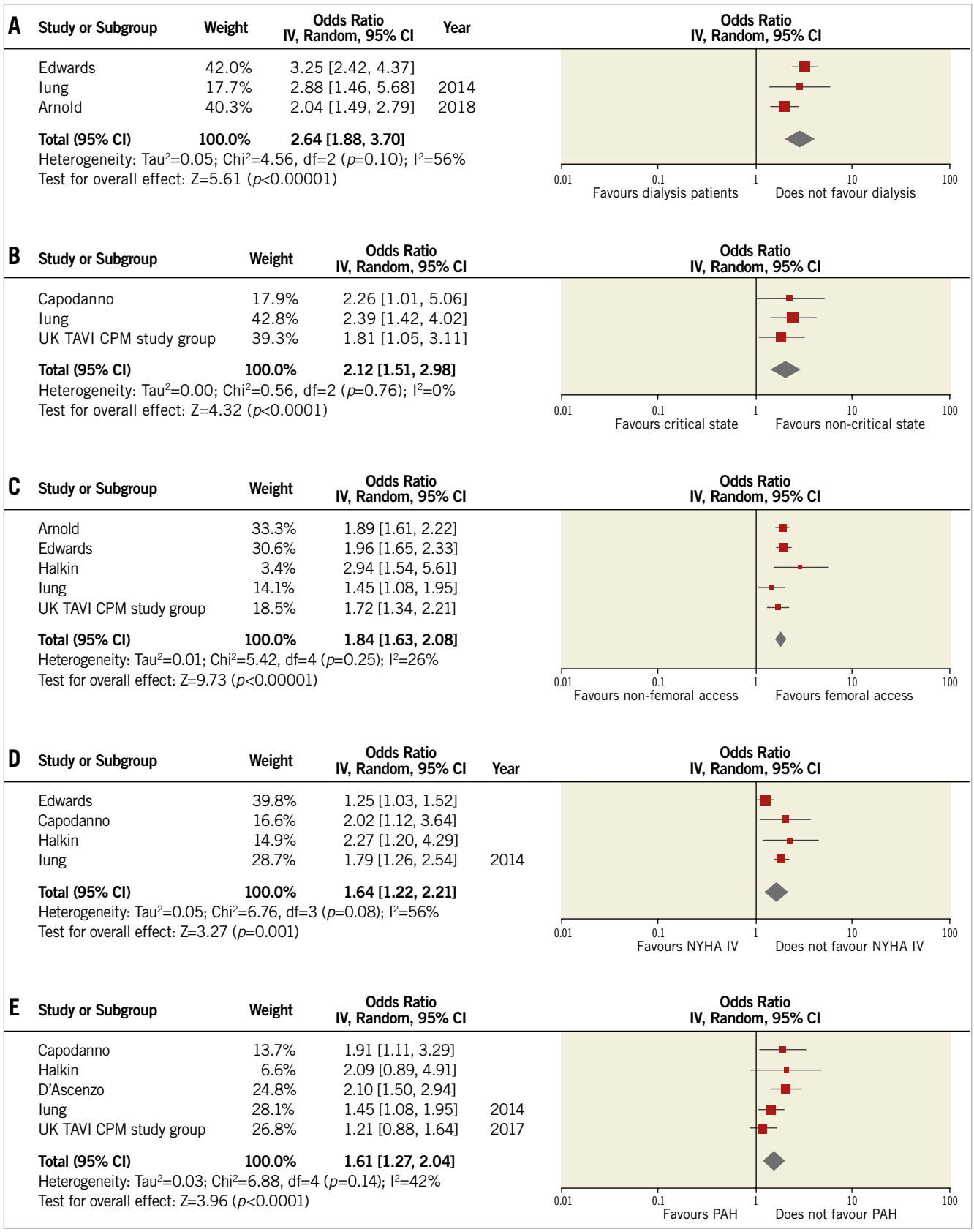


Figure 4. *Forest plots displaying the association of each predictor with short-term mortality. Baseline dialysis (A) was the strongest predictor of short-term mortality, followed by critical preoperative state (B), non-femoral access site (C), NYHA Class IV (D), pulmonary hypertension (E), home oxygen use (F), age greater than 85 (G), and GFR (per 5-unit decrease) (H). GFR: glomerular filtration rate; NYHA: New York Heart Association; PAH: pulmonary arterial hypertension*

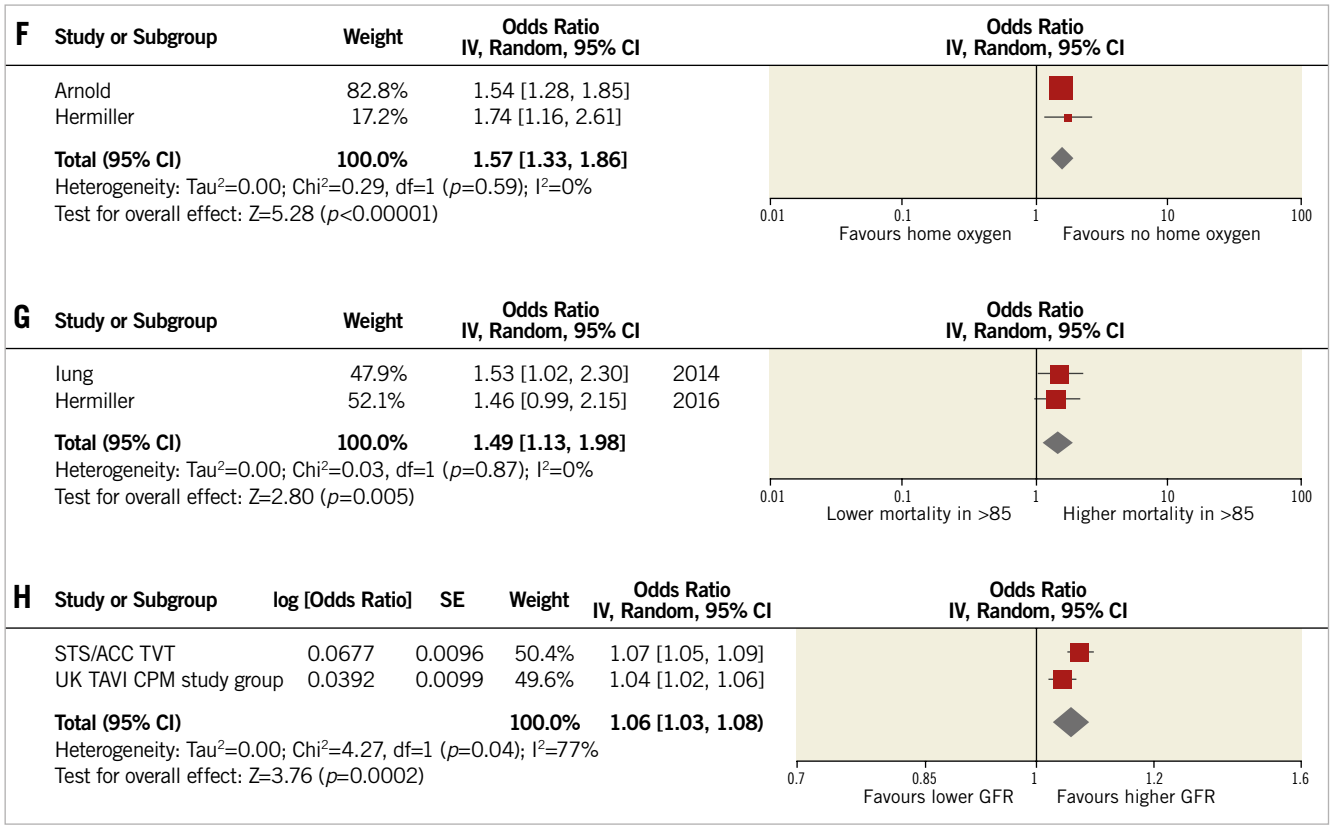


Figure 4. (cont'd) *Forest plots displaying the association of each predictor with short-term mortality. Baseline dialysis (A) was the strongest predictor of short-term mortality, followed by critical preoperative state (B), non-femoral access site (C), NYHA Class IV (D), pulmonary hypertension (E), home oxygen use (F), age greater than 85 (G), and GFR (per 5-unit decrease) (H). GFR: glomerular filtration rate; NYHA: New York Heart Association; PAH: pulmonary arterial hypertension*

(C-statistic range: 0.57-0.66); however, some of these models, such as the in-hospital STS/ACC TVT (C-statistic=0.65), STT (C-statistic=0.66), and UK TAVI CPM (C-statistic=0.66) predicted individual mortality more reliably than surgical models (C-statistic range: 0.59-0.61). Amongst the new TAVR-specific models that reported data on calibration, the STS/ACC TVT (both the in-hospital as well as the 30-day mortality versions) had the best performance. When both discrimination and calibration were considered together, the in-hospital STS/ACC TVT was the best performing RSM. Amongst the individual parameters analysed, baseline dialysis and non-femoral access site were the strongest predictors of 30-day mortality.

Globally, in the last few years, TAVR has been performed in more than 400,000 patients and indications keep growing at a rate of 40% annually[27]. This has presented the need for RSMs that can predict 30-day mortality, thereby allowing patient selection and provider comparisons[27]. Due to the lack of TAVR-specific models initially, several investigators tested the usefulness of surgical RSMs in assessing the risk of mortality in patients undergoing TAVR. However, valid concerns were raised about the limitations of surgical models. For example, these models do not include crucial factors that are strongly believed to affect candidacy for TAVR, such as home oxygen use, access site, assessments of frailty, and consideration of functional disabilities. Since 2014, several TAVR-specific models have emerged. However, reports concerning the applicability of these TAVR-specific RSMs have varied markedly in their findings.

A model with a discriminative capacity of C>0.80 provides strong support to guide medical decision making and can reliably dictate whether a patient will experience an event. Strongly discriminative models can also be relevant for research purposes, such as covariate adjustment in RCTs. Unfortunately, our study found that neither surgical nor TAVR-specific risk models currently meet the threshold of C>0.80. The highest C-statistic was of the CoreValve model (C-statistic=0.75), but it was unreliable due to a wide 95% CI (0.35-1.15). This unreliability may be because only a single, relatively small-sized study developed and validated this RSM, and due to the lack of external validation studies. The discriminative ability of the CoreValve model will become clearer as additional studies validate it. When both the C-statistic and 95% CI are considered, the in-hospital STS/ACC TVT model currently appears to have the best discrimination (C-statistic: 0.65; 95% CI: 0.62-0.68). We were only able to perform a meta-analysis on the C-statistics from an older version of this model; an

updated version demonstrated an even better C-statistic reaching up to 0.70 for in-hospital mortality and 0.71 for 30-day mortality[10]. However, there is still room for improvement. For example, other cardiovascular risk models, such as the ones for the management of heart failure and percutaneous coronary intervention, demonstrate C-statistics >0.80 for 30-day mortality[30]. There could be a couple of explanations as to why the TAVR-specific risk models do not currently achieve this level of discrimination. First, it could be due to limitations in the model, such as an insufficient number of predictors or due to predictors being dichotomised for simplicity. Additionally, relatively small and homogenous derivation cohorts, and absence of validation in external data sets could also be responsible. If this is the case, additional data (for example, from the continuously growing TVT registry), along with periodic model refinements will probably improve the discrimination. Regular model updates using the most recent outcome data are particularly important in a rapidly evolving field such as TAVR, where device and procedural advances have been shown to reduce periprocedural complications significantly, as reflected by a large heterogeneity of reported outcomes across major studies[22]. A second reason for the weak discrimination could be the inherent inability to discriminate between patients who will or will not die post TAVR. However, a poorly discriminating model (e.g., C~0.6), may be useful (when used in conjunction with clinical judgement) in a situation that does not have one outcome or choice that is clearly better or more likely than another.

RSMs with a good calibration (OER ~1) are useful for benchmarking and comparison of centre-level risk-adjusted outcome. This can be used by providers and sites to spur quality improvement, resulting in improved outcomes in patients with different risk profiles. According to our study, both the STS/ACC TVT (in-hospital and 30-day versions) and STS models demonstrate good calibration and may be used for this purpose. Our study demonstrates that there is considerable heterogeneity in the covariates incorporated in the TAVR-specific risk prediction models. This underscores the need for combining these covariates to form an RSM that outperforms the currently available RSMs.

Limitations

This meta-analysis has limitations that need to be considered when interpreting the results. First, this meta-analysis is based only on retrospective observational studies and some bias may be present as not all parameters may have been available for calculation in the risk models. In the future, large prospective validation cohorts are needed to assess the accuracy of such RSMs and validate our results. Second, some validation studies had to be excluded from our analysis as relevant data were not provided, which could have contributed to bias. Third, these estimates are derived from individual studies as we did not have access to the individual patient data. Fourth, most of these models were derived from patient populations with high to intermediate risk. Amongst the low-risk patient population, comorbidities are a less relevant part of risk scores to predict outcomes; other factors such as anatomical and procedural variables may be more important but are traditionally not included in RSMs. The publication of studies in lower-risk populations (such as the PARTNER 3 and Evolut trials) is likely to shift the TAVR use to lower-risk patients; the applicability of these scales in a lower-risk population is currently not known. While the focus of this manuscript is short-term mortality, it must be noted that it is not the only outcome driving clinical decisions. Long-term efficacy, functional outcomes and quality of life are also important and must be considered.

Conclusions

In conclusion, our study demonstrates that the in-hospital STS/ACC TVT model, the 30-day STS/ACC TVT model, and the STS model have accurate calibration in predicting short-term mortality. This makes these models useful for comparison of centre-level risk-adjusted mortality. In contrast, the discriminative ability of currently available models is limited, and room for improvement exists before wide clinical implementation.

Impact on daily practice

This study demonstrates that the STS/ACC TVT models (in-hospital and 30-day) and the STS model have accurate calibration and can therefore help physicians and administrators to compare centre-level risk-adjusted mortality. Discrimination of all RSMs was poor, and room for improvement exists before these can be used to predict the risk of individual patient mortality reliably. This study also reviews the predictors that make up each RSM and highlights the strongest predictors of mortality, which can assist in the development of new, better-performing models.

Conflict of interest statement

D.L. Bhatt discloses the following relationships – Advisory Board: Cardax, Elsevier Practice Update Cardiology, Medscape Cardiology, PhaseBio, Regado Biosciences; Board of Directors: Boston VA Research Institute, Society of Cardiovascular Patient Care, TobeSoft; Chair: American Heart Association Quality Oversight Committee; Data Monitoring Committees: Baim Institute for Clinical Research (formerly Harvard Clinical Research Institute, for the PORTICO trial, funded by St. Jude Medical, now Abbott), Cleveland Clinic (including for the ExCEED trial, funded by Edwards), Duke Clinical Research Institute, Mayo Clinic, Mount Sinai School of Medicine (for the ENVISAGE trial, funded by Daiichi Sankyo), Population Health Research Institute; Honoraria: American College of Cardiology (Senior Associate Editor, Clinical Trials and News, http://ACC.org; Vice-Chair, ACC Accreditation Committee), Baim Institute for Clinical Research (formerly Harvard Clinical Research Institute; RE-DUAL PCI clinical trial steering committee funded by Boehringer Ingelheim), Belvoir Publications (Editor in Chief, Harvard Heart Letter), Duke Clinical Research Institute (clinical trial steering committees), HMP Global (Editor in Chief, Journal of Invasive Cardiology),

Journal of the American College of Cardiology (Guest Editor; Associate Editor), Medtelligence/ReachMD (CME steering committees), Population Health Research Institute (for the COMPASS operations committee, publications committee, steering committee, and USA national co-leader, funded by Bayer), Slack Publications (Chief Medical Editor, Cardiology Today's Intervention), Society of Cardiovascular Patient Care (Secretary/Treasurer), WebMD (CME steering committees); Other: Clinical Cardiology (Deputy Editor), NCDR-ACTION Registry Steering Committee (Chair), VA CART Research and Publications Committee (Chair); Research Funding: Abbott, Amarin, Amgen, AstraZeneca, Bayer, Boehringer Ingelheim, Bristol-Myers Squibb, Chiesi, Eisai, Ethicon, Forest Laboratories, Idorsia, Ironwood, Ischemix, Lilly, Medtronic, PhaseBio, Pfizer, Regeneron, Roche, Sanofi Aventis, Synaptic, The Medicines Company; Royalties: Elsevier (Editor, Cardiovascular Intervention: A Companion to Braunwald's Heart Disease); Site Co-Investigator: Biotronik, Boston Scientific, St. Jude Medical (now Abbott), Svelte; Trustee: American College of Cardiology; Unfunded Research: FlowCo, Fractyl, Merck, Novo Nordisk, PLx Pharma, Takeda. The other authors have no conflicts of interest to declare.

References

1. Baumgartner H, Falk V, Bax JJ, De Bonis M, Hamm C, Holm PJ, Iung B, Lancellotti P, Lansac E, Rodriguez Muñoz D, Rosenhek R, Sjögren J, Tornos Mas P, Vahanian A, Walther T, Wendler O, Windecker S, Zamorano JL; ESC Scientific Document Group. 2017 ESC/EACTS Guidelines for the management of valvular heart disease. *Eur Heart J.* 2017;38:2739-91.

2. Siontis GC, Praz F, Pilgrim T, Mavridis D, Verma S, Salanti G, Søndergaard L, Jüni P, Windecker S. Transcatheter aortic valve implantation vs. surgical aortic valve replacement for treatment of severe aortic stenosis: a meta-analysis of randomized trials. *Eur Heart J.* 2016;37:3503-12.

3. 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:S23-42.

4. 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;100:1016-23.

5. 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; Steering Committee of the Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy Registry. Development and Validation of a Risk Prediction Model for In-Hospital Mortality After Transcatheter Aortic Valve Replacement. *JAMA Cardiol.* 2016;1:46-52.

6. Hermiller JB Jr, Yakubov SJ, Reardon MJ, Deeb GM, Adams DH, Afilao J, Huang J, Popma J; CoreValve United States Clinical Investigators. Predicting Early and Late Mortality After Transcatheter Aortic Valve Replacement. *J Am Coll Cardiol.* 2016;68:343-52.

7. D'Ascenzo F, Capodanno D, Tarantini G, Nijhoff F, Ciuca C, Rossi ML, Brambilla N, Barbanti M, Napodano M, Stella P, Saia F, Ferrante G, Tamburino C, Gasparetto V, Agostoni P, Marzocchi A, Presbitero P, Bedogni F, Cerrato E, Omedè P, Conrotto F, Salizzoni S, Biondi Zoccai G, Marra S, Rinaldi M, Gaita F, D'Amico M, Moretti C. Usefulness and validation of the survival posT TAVI score for survival after transcatheter aortic valve implantation for aortic stenosis. *Am J Cardiol.* 2014;114:1867-74.

8. 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, Möllmann H, Reichenspurner H, Nickenig G, Blankenberg S, Diemert P, Treede H, Walther T, Werner N, Schnabel RB. Development of a risk score for outcome after transcatheter aortic valve implantation. *Clin Res Cardiol.* 2014;103:631-40.

9. Martin GP, Sperrin M, Ludman PF, de Belder MA, Redwood SR, Townend JN, Gunning M, Moat NE, Banning AP, Buchan I, Mamas MA. Novel United Kingdom prognostic model for 30-day mortality following transcatheter aortic valve implantation. *Heart.* 2018;104:1109-16.

10. Arnold SV, O'Brien SM, Vemulapalli S, Cohen DJ, Stebbins A, Brennan JM, Shahian DM, Grover FL, Holmes DR, Thourani VH, Peterson ED, Edwards FH; STS/ACC TVT Registry. Inclusion of Functional Status Measures in the Risk Adjustment of 30-Day Mortality After Transcatheter Aortic Valve Replacement: A Report From the Society of Thoracic Surgeons/American College of Cardiology TVT Registry. *JACC Cardiovasc Interv.* 2018;11:581-9.

11. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ.* 2009;339:b2700.

12. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, McGinn T, Guyatt G. Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. *JAMA.* 2017;318:1377-84.

13. Martin GP, Sperrin M, Ludman PF, de Belder MA, Gale CP, Toff WD, Moat NE, Trivedi U, Buchan I, Mamas MA. Inadequacy of existing clinical prediction models for predicting mortality after transcatheter aortic valve implantation. *Am Heart J.* 2017;184:97-105.

14. Arai T, Lefèvre T, Hayashida K, Watanabe Y, O'Connor SA, Hovasse T, Romano M, Garot P, Bouvier E, Chevalier B, Morice MC. Usefulness of a Simple Clinical Risk Prediction Method, Modified ACEF Score, for Transcatheter Aortic Valve Implantation. *Circ J.* 2015;79:1496-503.

15. Durand E, Borz B, Godin M, Tron C, Litzler PY, Bessou JP, Dacher JN, Bauer F, Cribier A, Eltchaninoff H. Performance analysis of EuroSCORE II compared to the original logistic EuroSCORE and STS scores for predicting 30-day mortality after transcatheter aortic valve replacement. *Am J Cardiol.* 2013;111:891-7.

16. Haensig M, Holzhey DM, Borger MA, Schuler G, Shi W, Subramanian S, Rastan AJ, Mohr FW. Is the new EuroSCORE II a better predictor for transapical aortic valve implantation? *Eur J Cardiothorac Surg.* 2013;44:302-8.

17. 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, 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.

18. Sedaghat A, Sinning JM, Vasa-Nicotera M, Ghanem A, Hammerstingl C, Grube E, Nickenig G, Werner N. The revised EuroSCORE II for the prediction of mortality in patients undergoing transcatheter aortic valve implantation. *Clin Res Cardiol.* 2013;102:821-9.

19. Silva LS, Caramori PR, Nunes Filho AC, Katz M, Guaragna JC, Lemos P, Lima V, Abizaid A, Tarasoutchi F, Brito FS Jr. Performance of surgical risk scores to predict mortality after transcatheter aortic valve implantation. [Article in English, Portuguese]. *Arq Bras Cardiol.* 2015;105:241-7.

20. Watanabe Y, Hayashida K, Lefèvre T, Chevalier B, Hovasse T, Romano M, Garot P, Farge A, Donzeau-Gouge P, Bouvier E, Cormier B, Morice MC. Is EuroSCORE II better than EuroSCORE in predicting mortality after transcatheter aortic valve implantation? *Catheter Cardiovasc Interv.* 2013;81:1053-60.

21. Wendt D, Thielmann M, Kahlert P, Kastner S, Price V, Al-Rashid F, Patsalis P, Erbel R, Jakob H. Comparison between different risk scoring algorithms on isolated conventional or transcatheter aortic valve replacement. *Ann Thorac Surg.* 2014;97:796-802.

22. Yatsynovich Y, Khattak H, Ali M, Schwartz B, Pak S, Chen T. Comparison of transcatheter aortic valve replacement risk score against currently accepted surgical risk models as predictors of 30-day mortality in transcatheter aortic valve replacement. *J Interv Cardiol.* 2017;30:595-603.

23. Ben-Dor I, Gaglia MA Jr, Barbash IM, Maluenda G, Hauville C, Gonzalez MA, Sardi G, Laynez-Carnicero A, Torguson R, Okubagzi P, Xue Z, Goldstein SA, Suddath WO, Kent KM, Lindsay J, Satler LF, Pichard AD, Waksman R. Comparison between Society of Thoracic Surgeons Score and logistic EuroSCORE for predicting mortality in patients referred for transcatheter aortic valve implantation. *Cardiovasc Revasc Med.* 2011;12:345-9.

24. Pilgrim T, Franzone A, Stortecky S, Nietlispach F, Haynes A, Tueller D, Toggweiler S, Muller O, Ferrari E, Noble S, Maisano F, Jeger R, Roffi M, Grüenfelder J, Huber C, Wenaweser P, Windecker S. Predicting Mortality After Transcatheter Aortic Valve Replacement: External Validation of the Transcatheter Valve Therapy Registry Model. *Circ Cardiovasc Interv.* 2017 Nov;10(11).

25. Halkin A, Steinvil A, Witberg G, Barsheshet A, Barkagan M, Assali A, Segev A, Fefer P, Guetta V, Barbash IM, Kornowski R, Finkelstein A. Mortality prediction following transcatheter aortic valve replacement: A quantitative comparison of risk scores derived from populations treated with either surgical or percutaneous aortic valve replacement. The Israeli TAVR Registry Risk Model Accuracy Assessment (IRRMA) study. *Int J Cardiol.* 2016;215:227-31.

26. Silaschi M, Conradi L, Seiffert M, Schnabel R, Schön G, Blankenberg S, Reichenspurner H, Diemert P, Treede H. Predicting Risk in Transcatheter Aortic Valve Implantation: Comparative Analysis of EuroSCORE II and Established Risk Stratification Tools. *Thorac Cardiovasc Surg.* 2015;63:472-8.

27. Carmo J, Teles RC, Madeira S, Ferreira A, Brito J, Nolasco T, Gonçalves PA, Gabriel HM, Raposo L, Vale N, Ribeiras R, Abecasis M, Almeida MS, Neves JP, Mendes M. Comparison of multiparametric risk scores for predicting early mortality after transcatheter aortic valve implantation. [Article in English, Portuguese]. *Rev Port Cardiol.* 2018;37:585-90.

28. Codner P, Malick W, Kouz R, Patel A, Chen CH, Terre J, Landes U, Vahl TP, George I, Nazif T, Kirtane AJ, Khalique OK, Hahn RT, Leon MB, Kodali S. Mortality risk after transcatheter aortic valve implantation: analysis of the predictive accuracy of the Transcatheter Valve Therapy registry risk assessment model. *EuroIntervention.* 2018;14:e405-12.

29. Zbroński K, Huczek Z, Puchta D, Paczwa K, Kochman J, Wilimski R, Scisło P, Rymuza B, Filipiak KJ, Opolski G. Outcome prediction following transcatheter aortic valve implantation: Multiple risk scores comparison. *Cardiol J.* 2016;23:169-77.

30. Matheny ME, Ohno-Machado L, Resnic FS. Discrimination and calibration of mortality risk prediction models in interventional cardiology. *J Biomed Inform.* 2005;38:367-75.

Supplementary data

Supplementary Appendix 1. Methods.

Supplementary Figure 1. Discrimination of each TAVR-specific risk stratification model.

Supplementary Figure 2. Discrimination of each surgical risk stratification model.

Supplementary Figure 3. Calibration of each TAVR-specific risk stratification model.

Supplementary Figure 4. Calibration of each surgical risk stratification model.

Supplementary Table 1. Detailed search strategy used in each database.

Supplementary Table 2. Characteristics of included studies.

Supplementary Table 3. Covariates included in each RSM.

Supplementary Table 4. Risk of bias assessment of TAVR-specific risk models using the PROBAST scale.

Supplementary Table 5. P-interaction values for differences in discrimination between individual pairs of risk stratification models.

Supplementary Table 6. P-interaction values for differences in calibration between individual pairs of risk stratification models.

The supplementary data are published online at: https://eurointervention.pcronline.com/doi/10.4244/EIJ-D-19-00636

Supplementary data

Supplementary Appendix 1. Methods

Data sources and search strategy

Two reviewers (M.A.A. Khan and M.S. Usman) independently queried MEDLINE and Scopus databases up till June 2019. No time or language restrictions were placed. The search strategy involved using MeSH to determine the different keywords for the RSMs and TAVR coupled with Boolean operators AND and OR. A detailed search strategy for each database is provided in the supplementary files **(Supplementary Table 1)**. In order to cast a broad net, our search was conducted using the "keywords, abstract and title" filter. Other data sources included bibliographies of editorials and relevant reviews from major medical journals, conference proceedings for indexed abstracts, and databases of grey/unpublished literature.

Study selection

The predefined eligibility criteria were: (1) studies that sought to validate RSMs to be used in TAVR patients; (2) RSMs designed to predict short-term (30-day or in-hospital) mortality; and (3) reported C-statistics (also known as area under the curve [AUC]) with respective 95% confidence intervals (CIs) and/or expected and observed mortality rates.

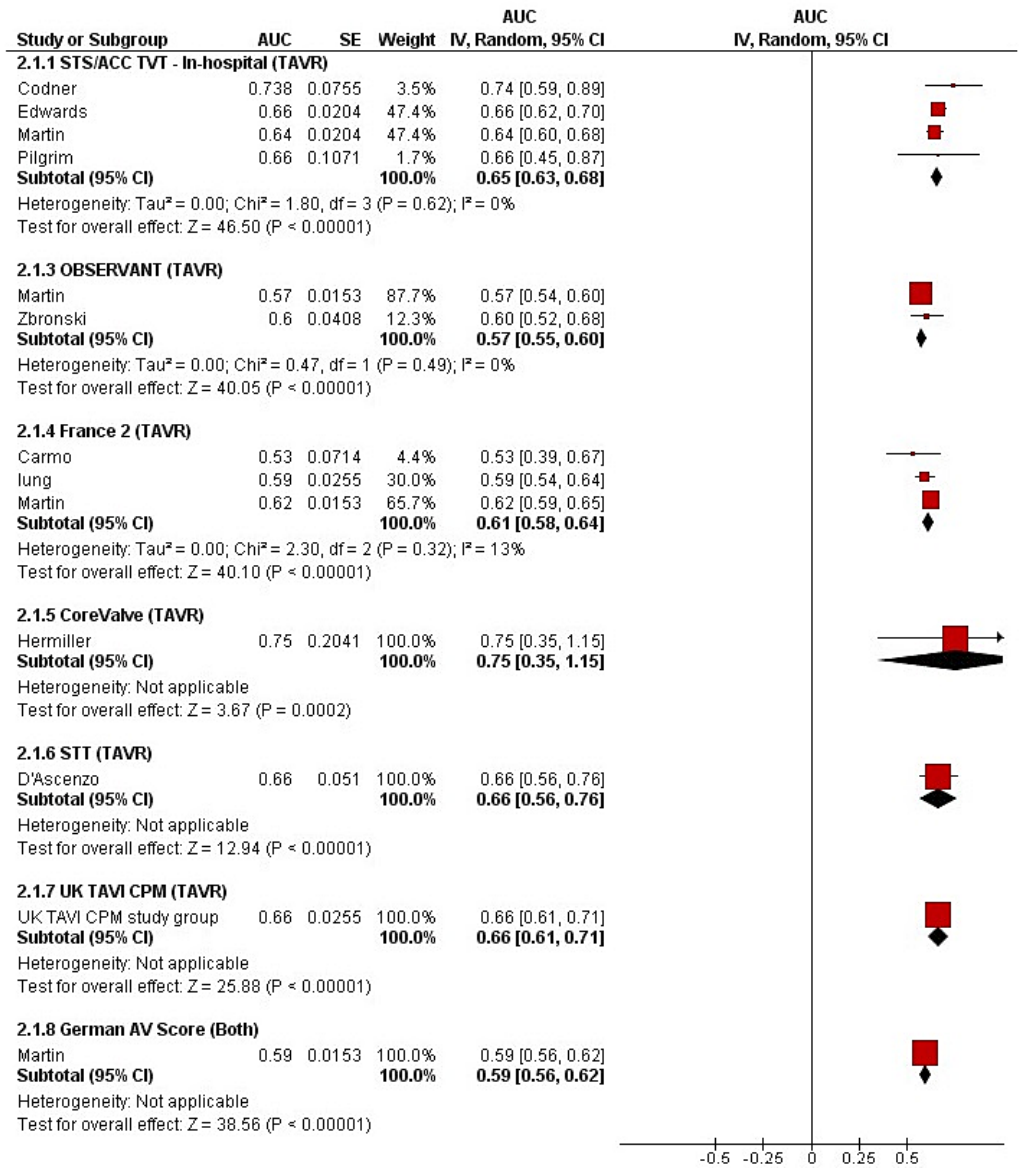
All articles retrieved from the systematic search were exported to Endnote Reference Library (Version X8.1; Clarivate Analytics, Philadelphia, PA, USA) software, where duplicates were removed. Remaining articles were initially short-listed at title and abstract level, after which the full-text articles were reviewed based on predefined criteria. Two reviewers (M.A.A. Khan and M.S. Usman) independently carried out this process under the supervision of a third reviewer (T.J. Siddiqi).

Data extraction and quality assessment

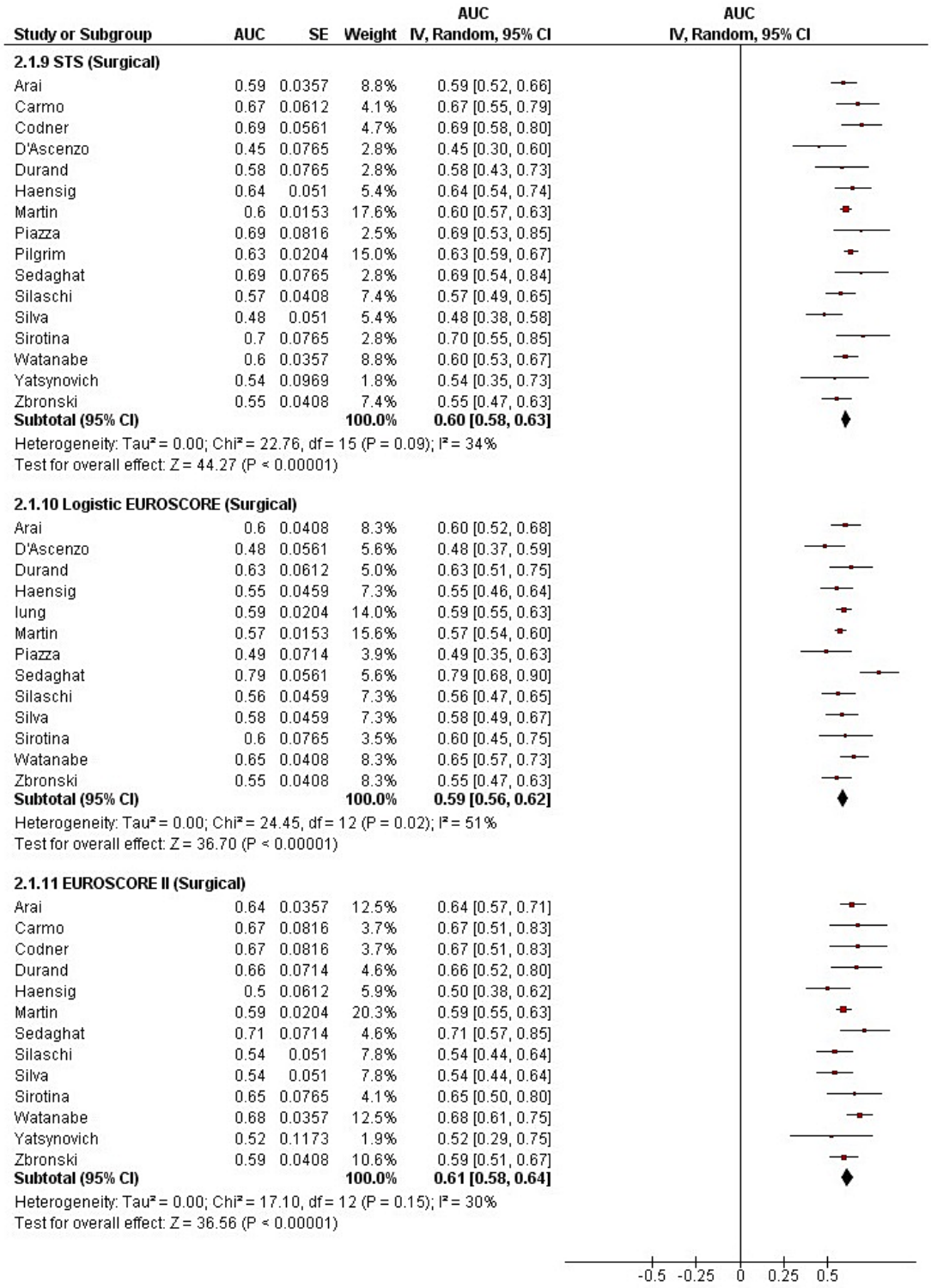
Data were abstracted on a standardised data collection from the short-listed articles and verified by two reviewers (M.A.A. Khan and M.S. Usman). In case of any discrepancy, the original reference article was reviewed again. Discrimination and calibration data were extracted from each study. The following information was abstracted: study characteristics, sample size, models derived and/or validated, follow-up duration, data registry and type of RSM (i.e., surgical or TAVR-specific). Additionally, the predictors used in each RSM were recorded.

It is important to note that different studies compared different subsets of risk models. We extracted data relevant to the following TAVR-specific models: (a) STS/ACC TVT (Society for Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy; this model was developed in 2016 by Arnold et al to predict in-hospital mortality [10]. It was then updated in 2018 and a new 30-day mortality risk model was also designed [15]. For the purposes of this study, information on the 2016 and 2018 in-hospital risk models was considered as the same model). (b) OBSERVANT (Observational Study Of Appropriateness, Efficacy And Effectiveness of AVR-TAVR Procedures For the Treatment Of Severe Symptomatic Aortic Stenosis); (c) FRANCE 2; (d) CoreValve; (e) STT and (f) UK TAVI models. Data were extracted on the following SAVR-specific models: (a) STS score, (b) logistic EuroSCORE, and (c) EuroSCORE II models. CPM German AV Score is used for both TAVR and SAVR.

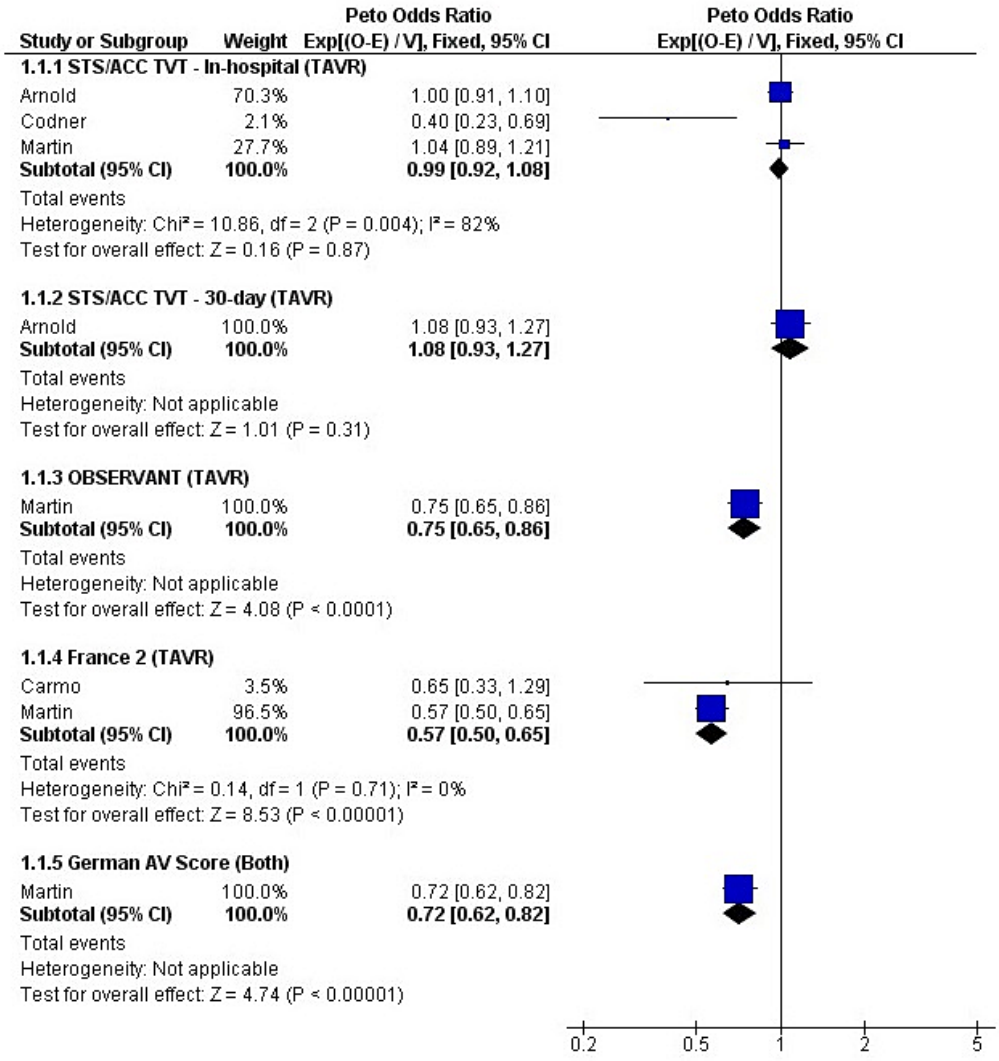
The Prediction model Risk of Bias Assessment Tool (PROBAST) was used to assess the risk of bias of the new TAVR-specific risk models [18]. This scale enables critical appraisal of a particular RSM by assessment of four domains: participants, predictors, outcome, and analysis. A total of 20 signalling questions within these domains help to assess the structured judgement of risk of bias.



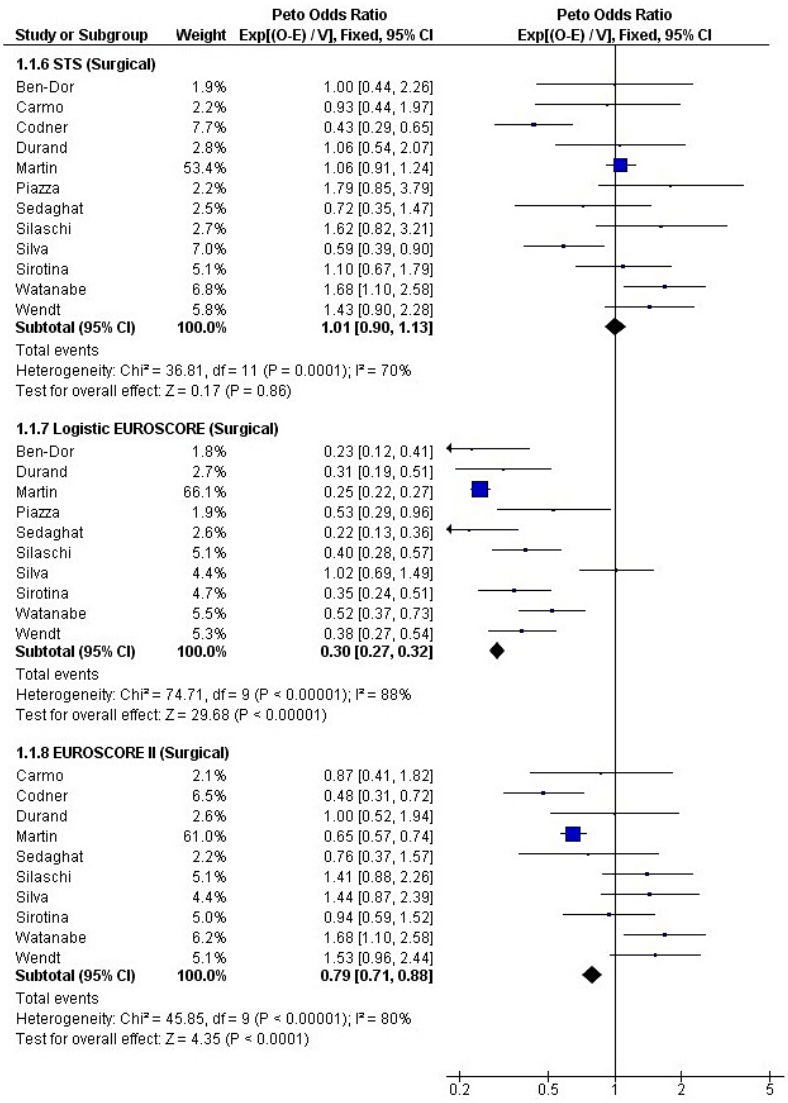
Supplementary Figure 1. Discrimination of each TAVR-specific risk stratification model.



Supplementary Figure 2. Discrimination of each surgical risk stratification model.



Supplementary Figure 3. Calibration of each TAVR-specific risk stratification model.



Supplementary Figure 4. Calibration of each surgical risk stratification model.

Supplementary Table 1. Detailed search strategy used in each database.

Database	Search strategy	Articles retrieved
MEDLINE	(("transcatheter aortic valve replacement"[MeSH Terms] OR ("transcatheter"[All Fields] AND "aortic"[All Fields] AND "valve"[All Fields] AND "replacement"[All Fields]) OR "transcatheter aortic valve replacement"[All Fields]) OR TAVR[All Fields] OR ("transcatheter aortic valve replacement"[MeSH Terms] OR ("transcatheter"[All Fields] AND "aortic"[All Fields] AND "valve"[All Fields] AND "replacement"[All Fields]) OR "transcatheter aortic valve replacement"[All Fields] OR ("transcatheter"[All Fields] AND "aortic"[All Fields] AND "valve"[All Fields] AND "implantation"[All Fields]) OR "transcatheter aortic valve implantation"[All Fields]) OR TAVI[All Fields] OR (Percutaneous[All Fields] AND ("aortic valve"[MeSH Terms] OR ("aortic"[All Fields] AND "valve"[All Fields]) OR "aortic valve"[All Fields]) AND ("replantation"[MeSH Terms] OR "replantation"[All Fields] OR "replacement"[All Fields])) OR (Percutaneous[All Fields] AND ("aortic valve"[MeSH Terms] OR ("aortic"[All Fields] AND "valve"[All Fields]) OR "aortic valve"[All Fields]) AND ("embryo implantation"[MeSH Terms] OR ("embryo"[All Fields] AND "implantation"[All Fields]) OR "embryo implantation"[All Fields] OR "implantation"[All Fields]))) AND ((("risk"[MeSH Terms] OR "risk"[All Fields]) AND Model[All Fields]) OR (("risk"[MeSH Terms] OR "risk"[All Fields]) AND Prediction[All Fields]) OR (("risk"[MeSH Terms] OR "risk"[All Fields]) AND Stratification[All Fields]) OR (("risk"[MeSH Terms] OR "risk"[All Fields]) AND Score[All Fields]) OR ("risk assessment"[MeSH Terms] OR ("risk"[All Fields] AND "assessment"[All Fields]) OR "risk assessment"[All Fields])) AND (("mortality"[Subheading] OR "mortality"[All Fields] OR "mortality"[MeSH Terms]) OR ("mortality"[Subheading] OR "mortality"[All Fields] OR "survival"[All Fields] OR "survival"[MeSH Terms]) OR ("death"[MeSH Terms] OR "death"[All Fields]))	1,545
Scopus	((TITLE-ABS-KEY (transcatheter AND aortic AND valve AND replacement) OR TITLE-ABS-KEY (tavr) OR TITLE-ABS-KEY (transcatheter AND aortic AND valve AND implantation) OR TITLE-ABS-KEY (tavi) OR TITLE-ABS-KEY (percutaneous AND aortic AND valve AND replacement) OR TITLE-ABS-KEY (percutaneous AND aortuc AND valve AND implantation))) AND ((TITLE-ABS-KEY (risk AND model) OR TITLE-ABS-KEY (risk AND prediction) OR TITLE-ABS-KEY (risk AND stratificatiom) OR TITLE-ABS-KEY (risk AND score) OR TITLE-ABS-KEY (risk AND assessment))) AND ((TITLE-ABS-KEY (mortality) OR TITLE-ABS-KEY (survival) OR TITLE-ABS-KEY (death)))	2,662
EMBASE	(Transcatheter aortic valve replacement OR TAVR OR Transcatheter Aortic Valve Implantation OR TAVI OR Percutaneous Aortic Valve Replacement OR Percutaneous Aortic Valve Implantation) AND (Risk Model OR Risk Prediction OR Risk Stratification OR Risk Score OR Risk Assessment) AND (Mortality OR Survival OR Death)	1,892

Supplementary Table 2. Characteristics of included studies.

Author	Year	Study population	Models compared	Follow-up used for analysis	Sample size	Type of study	Inclusion criteria	Data sample (validation)	Country
D'Ascenzo	2014	Single centre	STT, LES, STS	30 days	180	Derivation (STT) + validation	All consecutive patients with severe symptomatic aortic stenosis referred for transcatheter aortic valve implantation	Bologna	Italy
Hermiller	2016	Single centre	CoreValve	30 days	1,205	Derivation (CoreValve) +validation	Patients with New York Heart Association functional Class II or greater symptoms related to aortic valve disease were eligible for the trial from which data were used to derive and validate the model.	Core Valve US Pivotal trial (USA)	USA
Iung	2014	Single centre	FRANCE 2, LES	30 days	1,281	Derivation (FRANCE 2) +validation	Patients were selected for transcatheter aortic valve implantation if they had severe, symptomatic aortic stenosis and if surgery was contraindicated or judged to be high risk by a multidisciplinary team.	Data from the French Aortic National CoreValve and Edwards (FRANCE 2)	Monaco and France
Arnold	2018	Single centre	STS/ACC TVT (30-day)	30 days	26,687	Derivation (STS/ACC TVT – 30-day) +validation	-	ACC/TVT	USA

UK TAVI CPM study group	2017	Single centre	UK TAVI CPM	30 days	6,339	Derivation (UK TAVI CPM) +validation	-	UK-TAVI	UK
Edwards	2016	Multicentre	STS/ACC TVT	In-hospital mortality	13,718	Derivation (STS/ACC TVT - in-hospital) +validation	The appropriate clinical indication for transcatheter aortic valve implantation was determined by at least 2 cardiothoracic surgeons. In general, the patients undergoing TAVR were considered to be unsuitable for or at extreme risk with surgical aortic valve replacement.	ACC/TVT	USA
Arai	2015	Multicentre	STS, LES, LES-II, ACEF	12 months	703	Validation	From October 2006, all consecutive high-risk patients with severe symptomatic aortic stenosis treated with transcatheter aortic valve implantation were prospectively enrolled	-	France, Japan
Durand	2013	Multicentre	STS, LES, LES-II	30 days	250	Validation	The patients were considered candidates for transcatheter aortic valve replacement when the logistic EuroSCORE was >20%, in case of frailty (by agreement between cardiologists and cardiac surgeons), or in case of comorbidities contraindicating surgical aortic valve replacement (porcelain aorta, chest irradiation, or deformation)	University Hospital of Rouen, Hospital Charles Nicolle, INSERM UMR 1096, Rouen, France.	France

Haesnig	2013	Single centre	STS, LES, LES-II	30 days	360	Validation	Clinical inclusion criteria were age ≥75 years, New York Heart Association functional Class II or higher, written informed consent and comorbidities leading to a logistic EuroSCORE ≥15%.	-	Germany
Piazza	2009	Multicentre	STS, LES	30 days	168	Validation	Patients were referred for transcatheter aortic valve implantation after a team of physicians (typically including interventional cardiologists and cardiac surgeons) agreed that surgical replacement would be associated with either high or prohibitive risk.	Bern University Hospital, Erasmus Medical Center	Switzerland, Netherlands
Sedaghat	2013	Multicentre	STS, LES, LES-II	1 year	206	Validation	-	Universitatsklinikum Bonn, Med. Klinik und Poliklinik II	Germany
Silva	2015	Multicentre	STS, LES, LES-II, AG, GS	30 days	418	Validation	-	Brazilian Society of Interventional Cardiology	Brazil
Watanabe	2013	Single centre	STS, LES, LES-II	30 days	453	Validation	Patients with severe symptomatic aortic stenosis (valve area <1.0 cm^2) were considered candidates for transcatheter aortic valve implantation if they had an LES >20%	Institut Cardiovasculaire Paris Sud	France

Sirotina	2013	Multicentre	STS, LES, LES-II	30 days	450	Validation	-	-	Germany
Ben-Dor	2011	Single centre	STS, LES	30 days	718	Validation	-	Washington Hospital Center	USA
Yatsynovich	2016	Single centre	STS, LES, LES-II, TAVR-RS	30 days	182	Validation	-	Kettering Medical Center	USA
Wendt	2014	Single centre	STS, ACEF, LES-II	30 days	1,512	Validation	Patients undergoing reoperation, emergency procedures, myectomy, aortic root enlargement to prevent patient-prosthesis mismatch, or simple wrapping/plication of the ascending aorta were included.	West-German Heart Center Essen	Germany
Martin	2016	Multicentre	German AV, FRANCE 2, OBSERVANT, STS/ACC TVT, LES, ES-II, STS	30 days	6,676	Validation	-	UK-TAVI	UK

Pilgrim	2017	Multicentre	STS, STS/ACC TVT	30 days	3,491	Validation	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.	Swiss TAVI registry (NCT01368250)	Switzerland
Halkin	2016	Multicentre	STS, LES, LES-II, OBSERVANT, FRANCE 2, German AV Score	30 days	1,327	Validation	Eligibility for transcatheter aortic valve replacement was established by a multidisciplinary Heart Team based on the calculated STS or EuroSCORE, or, for cases with an STS score <8, surgical risk was considered high based on other factors and comorbidities absent from the surgical risk scores (e.g., frailty measures).	Israeli TAVR Registry Risk Model Accuracy Assessment (IRRMA) study	Israel
Silaschi	2014	Single centre	STS, LES, LES-II	30 days	457	Validation	Patients were allocated to transcatheter aortic valve implantation when deemed unsuitable for conventional surgery due to contraindications or high risk by the local interdisciplinary Heart Team consisting of cardiologists and cardiac surgeons.	University Heart Center Hamburg	Germany

Codner	2018	Single centre	STS, STS/ACC TVT, LES II	30 days and in-hospital mortality	1,038	Validation	Severe AS was defined as a valvular orifice area <1.0 cm^2 or <0.6 cm^2/m^2 and/or mean pressure gradient >40 mmHg and/or jet velocity >4.0 m/s. Selected patients with discordant echocardiographic findings underwent dobutamine echocardiography.	NewYork-Presbyterian Hospital/Columbia University Medical Center	USA
Carmo	2018	Single centre	FRANCE 2, EuroSCORE II and STS scores	30-day mortality	240	Validation	-	Department of Cardiology, Hospital of Santa Cruz	Portugal
Zbroński	2016	Single centre	OBSERVANT, ACEF, SURTAVI, LES-II, STS	30-day mortality	156	Validation	-	Department of Cardiology, Medical University of Warsaw	Poland

ACEF: age, creatinine, and ejection fraction; AS: Ambler score; FRANCE 2: FRench Aortic National CoreValve and Edwards; GS: Guaragna score; German AV Score: German aortic valve score; LES: logistic EuroSCORE; LES II: logistic EuroSCORE 2; OBSERVANT: Observational Study Of Appropriateness, Efficacy And Effectiveness of AVR-TAVR Procedures For the Treatment Of Severe Symptomatic Aortic Stenosis; STS/ACC TVT: Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy; STT: survival posT TAVI; UK TAVI CPM: UK transcatheter aortic valve implantation clinical prediction models

Supplementary Table 3. Covariates included in each RSM.

Variables	German AV score	STT	STS/ACC TVT (in-hospital - updated)	STS/ACC TVT (30-day)	CoreValve	FRANCE 2	OBSERVANT	UK TAVI CPM	STS	LES	ESII
ACE inhibitor use									*		
Active endocarditis									*	*	*
ADP inhibitor use									*		
Age	*		*	*	*	*		*	*	*	*
Albumin level <3.3 G/Dl					*						
Alcohol									*		
Anaemia											
Aortic insufficiency									*		
Aortic stenosis									*		
Approach (transfemoral, transapical, etc.)						*					
Aortic valve disease aetiology									*		
Body mass index (kg/m^2)	*					*			*		
Cardiac surgery	*									*	*
Cardiogenic shock									*		
Clinical preoperative state	*							*		*	*
Concomitant surgery									*		
Coronary artery disease			*								
Critical preoperative state	*					*	*	*			
Diabetes on insulin									*		*
Dialysis						*					

Ejection fraction									*		
Emergency	*									*	*
Extracardiac arteriopathy	*							*	*		*
Female sex	*							*	*	*	*
Forced expiratory volume of 1 breath					*						
Glomerular filtration rate			*				*	*			
GP IIb/IIIa inhibitor use									*		
Heart block									*		
Haematocrit									*		
Haemodialysis			*	*					*		
Haemoglobin											
Home oxygen use				*	*				*		
Illicit drug use									*		
Immunocompromise									*		
Inability to complete 5 m walk test			*	*							
Inverse renal function		*									
KATZ index of activities of daily living					*			*			
Last creatinine level									*		
Left ventricular ejection fraction	*						*	*			
Liver dysfunction									*		
Logistic EuroSCORE					*						
Lower body surface area			*	*							
Lower KCCQ scores			*	*							
Left ventricle dysfunction										*	*

Male sex											
Mean transvalvular gradient											
Mitral insufficiency									*		
Mitral stenosis									*		
Neurological dysfunction										*	
New York Heart Association (NYHA) functional Class IV	*		*		*	*	*		*		*
Non-elective procedure								*			
Non-femoral access site			*	*							
Pulmonary artery systolic pressure >50 mmHg		*									
Pulmonary artery systolic pressure >60 mmHg								*			
Percent stenosis									*		
Peripheral artery disease		*	*								
Platelet count			*						*		
Pneumonia									*		
Poor mobility								*			
Porcelain thoracic aorta											
Post-infarct septal rupture										*	
Previous myocardial infarction	*		*						*	*	*
Previous stroke		*									
Prior balloon aortic valvuloplasty							*				
Procedural acuity			*	*							
Pulmonary hypertension	*					*	*			*	*
Pulmonary oedema						*		*			
Race									*		

Renal dysfunction	*			*					*		*
Residence in an assisted living facility					*						
Serum creatinine										*	
Severe chronic lung disease	*		*						*	*	*
Steroid use									*		
STS PROM					*						
STS severe lung disease					*						
Surgery on thoracic aorta										*	*
Syncope									*		
Tricuspid insufficiency			*	*					*		
Unplanned weight loss					*						
Unstable angina										*	*
Ventricular dysfunction											
WBC count									*		
Weight of the intervention											*

FRANCE 2: FRench Aortic National CoreValve and Edwards; German AV Score: German aortic valve score; KCCQ: Kansas City Cardiomyopathy Questionnaire Score; LES: logistic EuroSCORE; LES II: logistic EuroSCORE 2; OBSERVANT: Observational Study Of Appropriateness, Efficacy And Effectiveness of AVR-TAVR Procedures For the Treatment Of Severe Symptomatic Aortic Stenosis; STS/ACC TVT: Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy; STS-PROM: Society of Thoracic Surgeons Predicted Risk of Mortality; STT: survival posT TAVI; UK TAVI CPM: UK transcatheter aortic valve implantation clinical prediction models; WBC: white blood cell

Supplementary Table 4. Risk of bias assessment of TAVR-specific risk models using the PROBAST scale.

Study	Risk of bias				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB	Applicability
German AV Score	-	-	-	-	-	-	-	-	-
STT	-	-	+	+	-	-	-	+	-
STS/ACC TVT (in-hospital)	+	-	-	-	-	-	-	-	-
STS/ACC TVT (30-day)	?	-	-	+	-	-	-	-	-
CoreValve	-	-	-	+	+	+	-	-	-
FRANCE 2	+	-	+	-	-	-	-	+	-
OBSERVANT	?	-	?	-	-	-	-	?	-
UK TAVI CPM	+	-	-	-	+	+	+	-	+

(+) High risk of bias; (-) low risk of bias; (?) unclear risk of bias

TAVR: transcatheter aortic valve replacement; FRANCE 2: FRench Aortic National CoreValve and Edwards; German AV Score: German aortic valve score; OBSERVANT: Observational Study Of Appropriateness, Efficacy And Effectiveness of AVR-TAVR Procedures For the Treatment Of Severe Symptomatic Aortic Stenosis; ROB: risk of bias; STS/ACC TVT: Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy; STT: survival posT TAVI; UK TAVI CPM: UK transcatheter aortic valve implantation clinical prediction models

Supplementary Table 5. P-interaction values for differences in discrimination between individual pairs of risk stratification models.

Comparison	*p*-interaction	Comment
STS/ACC TVT vs OBSERVANT	$p<0.001$	Favours STS/ACC TVT
STS/ACC TVT vs FRANCE 2	$p=0.03$	Favours STS/ACC TVT
STS/ACC TVT vs CoreValve	$p=0.64$	No difference
STS/ACC TVT vs STT	$p=0.90$	No difference
STS/ACC TVT vs UK TAVI CPM	$p=0.82$	No difference
STS/ACC TVT vs German AV Score	$p=0.002$	Favours STS/ACC TVT
STS/ACC TVT vs STS	$p=0.009$	Favours STS/ACC TVT
STS/ACC TVT vs logistic EuroSCORE	$p=0.002$	Favours STS/ACC TVT
STS/ACC TVT	$p=0.049$	Favours STS/ACC TVT
OBSERVANT vs FRANCE 2	$p=0.11$	No difference
OBSERVANT vs CoreValve	$p=0.39$	No difference
OBSERVANT vs STT	$p=0.10$	No difference
OBSERVANT vs UK TAVI CPM	$p=0.003$	Favours UK TAVI CPM
OBSERVANT vs German AV score	$p=0.44$	No difference
OBSERVANT vs STS	$p=0.15$	No difference
OBSERVANT vs logistic EuroSCORE	$p=0.53$	No difference
OBSERVANT vs EuroSCORE II	$p=0.10$	No difference
FRANCE 2 vs CoreValve	$p=0.48$	No difference
FRANCE 2 vs STT	$p=0.32$	No difference
FRANCE 2 vs UK TAVI CPM	$p=0.07$	No difference
FRANCE 2 vs German AV Score	$p=0.43$	No difference
FRANCE 2 vs STS	$p=0.81$	No difference
FRANCE 2 vs logistic EuroSCORE	$p=0.36$	No difference
FRANCE 2 vs logistic EuroSCORE II	$p=0.89$	No difference
CoreValve vs STT	$p=0.67$	No difference
CoreValve vs UK TAVI CPM	$p=0.66$	No difference

CoreValve vs German AV Score	p=0.43	No difference
CoreValve vs STS	p=0.47	No difference
CoreValve vs logistic EuroSCORE	p=0.43	No difference
CoreValve vs EuroSCORE II	p=0.49	No difference
STT vs UK TAVI CPM	p=1.00	No difference
STT vs German AV Score	p=0.19	No difference
STT vs STS	p=0.27	No difference
STT vs logistic EuroSCORE	p=0.17	No difference
STT vs logistic EuroSCORE II	p=0.35	No difference
UK TAVI CPM vs German AV Score	p=0.02	Favours UK TAVI CPM
UK TAVI CPM vs STS	p=0.049	Favours UK TAVI CPM
UK TAVI CPM vs logistic EuroSCORE	p=0.02	Favours UK TAVI CPM
UK TAVI CPM vs EuroSCORE II	p=0.10	No difference
German AV Score vs STS	p=0.56	No difference
German AV Score vs logistic EuroSCORE	p=0.89	No difference
German AV Score vs logistic EuroSCORE II	p=0.38	No difference
STS vs logistic EuroSCORE	p=0.47	No difference
STS vs logistic EuroSCORE II	p=0.71	No difference

Supplementary Table 6. P-interaction values for differences in calibration between individual pairs of risk stratification models.

Comparison	*p*-interaction	Comment
STS/ACC TVT in-hospital vs STS/ACC 30-day	$p=0.33$	No difference
STS/ACC TVT in-hospital vs OBSERVANT	$p<0.001$	Favours STS/ACC TVT in-hospital
STS/ACC TVT in-hospital vs FRANCE 2	$p<0.001$	Favours STS/ACC TVT in-hospital
STS/ACC TVT in-hospital vs German AV Score	$p<0.001$	Favours STS/ACC TVT in-hospital
STS/ACC TVT in-hospital vs STS	$p=0.82$	No difference
STS/ACC TVT in-hospital vs logistic EuroSCORE	$p<0.001$	Favours STS/ACC TVT in-hospital
STS/ACC in-hospital vs EuroSCORE II	$p<0.001$	Favours STS/ACC TVT in-hospital
STS/ACC 30-day vs OBSERVANT	$p<0.001$	Favours STS/ACC TVT in-hospital
STS/ACC 30-day vs FRANCE 2	$p<0.001$	Favours STS/ACC TVT in-hospital
STS/ACC 30-day vs German AV Score	$p<0.001$	Favours STS/ACC TVT in-hospital
STS/ACC 30-day vs STS	$p=0.47$	No difference
STS/ACC 30-day vs logistic EuroSCORE	$p<0.001$	Favours STS/ACC TVT in-hospital
STS/ACC 30-day vs EuroSCORE II	$p<0.001$	Favours STS/ACC TVT in-hospital
OBSERVANT vs FRANCE 2	$p<0.001$	Favours OBSERVANT
OBSERVANT vs German AV Score	$p=0.66$	No difference
OBSERVANT vs STS	$p<0.001$	Favours OBSERVANT
OBSERVANT vs logistic EuroSCORE	$p<0.001$	Favours OBSERVANT
OBSERVANT vs EuroSCORE II	$p=0.53$	No difference

FRANCE 2 vs German AV Score	p=0.02	Favours German AV Score
FRANCE 2 vs STS	p<0.001	Favours STS
FRANCE 2 vs logistic EuroSCORE	p<0.001	Favours FRANCE 2
FRANCE 2 vs EuroSCORE II	p<0.001	Favours EuroSCORE II
German AV Score vs STS	p<0.001	Favours STS
German AV Score vs logistic EuroSCORE	p<0.001	Favours German AV Score
German AV Score vs EuroSCORE II	p=0.26	No difference
STS vs logistic EuroSCORE	p<0.001	Favours STS
STS vs EuroSCORE II	p<0.002	Favours STS
Logistic EuroSCORE vs EuroSCORE II	p<0.001	Favours EuroSCORE II