



Risk modeling in transcatheter aortic valve replacement remains unsolved: an external validation study in 2946 German patients

Georg Wolff¹ · Jasmin Shamekhi² · Baravan Al-Kassou² · Noriaki Tabata² · Claudio Parco¹ · Kathrin Klein¹ · Oliver Maier¹ · Alexander Sedaghat² · Amin Polzin¹ · Atsushi Sugiura² · Christian Jung¹ · Eberhard Grube² · Ralf Westenfeld¹ · Andrea Icks⁴ · Tobias Zeus¹ · Jan-Malte Sinning^{2,5} · Stephan Baldus^{3,5} · Georg Nickenig^{2,5} · Malte Kelm^{1,5} · Verena Veulemans¹

Received: 13 June 2020 / Accepted: 12 August 2020 / Published online: 26 August 2020
© The Author(s) 2020

Abstract

Background Surgical risk prediction models are routinely used to guide decision-making for transcatheter aortic valve replacement (TAVR). New and updated TAVR-specific models have been developed to improve risk stratification; however, the best option remains unknown.

Objective To perform a comparative validation study of six risk models for the prediction of 30-day mortality in TAVR

Methods and results A total of 2946 patients undergoing transfemoral (TF, $n = 2625$) or transapical (TA, $n = 321$) TAVR from 2008 to 2018 from the German Rhine Transregio Aortic Diseases cohort were included. Six surgical and TAVR-specific risk scoring models (LogES I, ES II, STS PROM, FRANCE-2, OBSERVANT, GAVS-II) were evaluated for the prediction of 30-day mortality. Observed 30-day mortality was 3.7% (TF 3.2%; TA 7.5%), mean 30-day mortality risk prediction varied from $5.8 \pm 5.0\%$ (OBSERVANT) to $23.4 \pm 15.9\%$ (LogES I). Discrimination performance (ROC analysis, c -indices) ranged from 0.60 (OBSERVANT) to 0.67 (STS PROM), without significant differences between models, between TF or TA approach or over time. STS PROM discriminated numerically best in TF TAVR (c -index 0.66; range of c -indices 0.60 to 0.66); performance was very similar in TA TAVR (LogES I, ES II, FRANCE-2 and GAVS-II all with c -index 0.67). Regarding calibration, all risk scoring models—especially LogES I—overestimated mortality risk, especially in high-risk patients.

Conclusions Surgical as well as TAVR-specific risk scoring models showed mediocre performance in prediction of 30-day mortality risk for TAVR in the German Rhine Transregio Aortic Diseases cohort. Development of new or updated risk models is necessary to improve risk stratification.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00392-020-01731-9>) contains supplementary material, which is available to authorized users.

✉ Georg Wolff
Georg.Wolff@med.uni-duesseldorf.de

¹ Division of Cardiology, Pulmonology and Vascular Medicine, Department of Internal Medicine, Heinrich Heine University, Medical Faculty, Moorenstr. 5, 40225 Düsseldorf, Germany

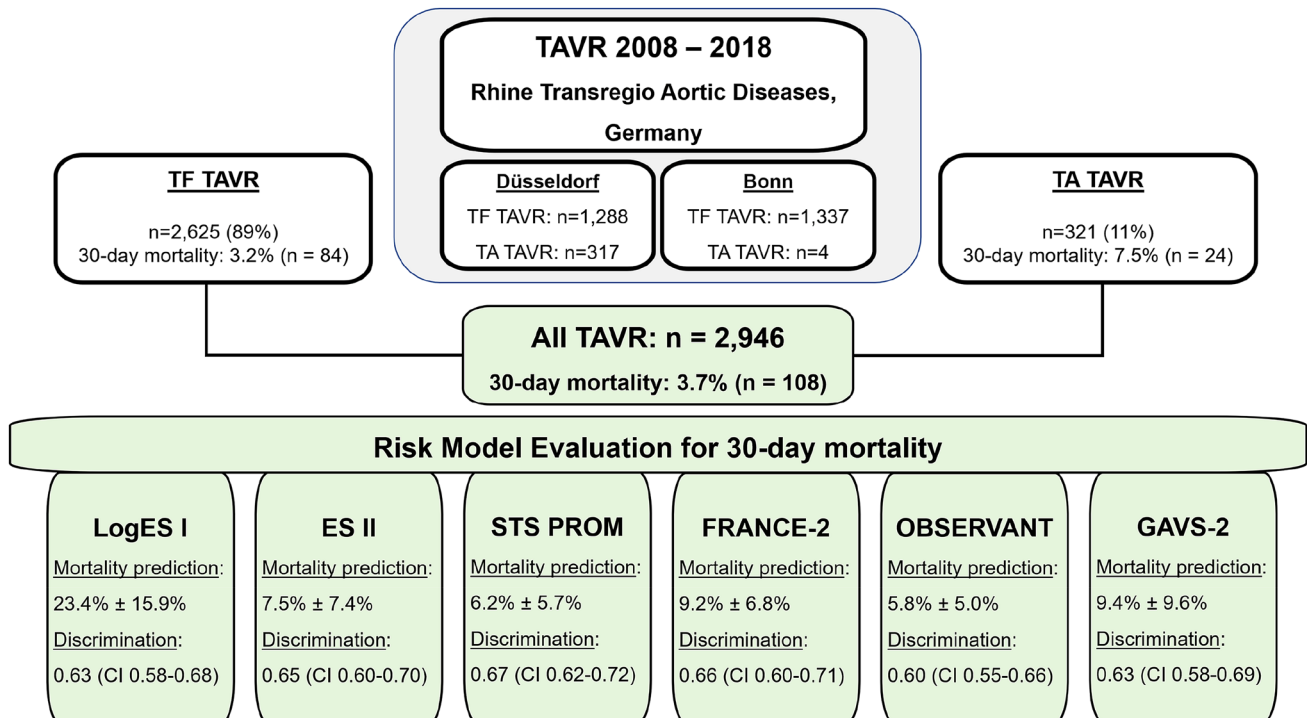
² Department of Medicine II, Heart Center Bonn, University Hospital Bonn, Bonn, Germany

³ Division of Cardiology, Pneumology, Angiology and Intensive Care, Department of Internal Medicine III, University of Cologne, Cologne, Germany

⁴ Institute for Health Services Research and Health Economics, Centre for Health and Society, Medical Faculty, Heinrich-Heine-University, Düsseldorf, Germany

⁵ Transregio 259: Aortic Diseases-Scientific Network of University Heart Centers in Düsseldorf/Bonn/Cologne, Düsseldorf/Bonn/Cologne, Germany

Graphic abstract



Keywords TAVR · TAVI · Risk prediction · Mortality · EuroSCORE · GAVS

Abbreviations

ES II	EuroSCORE II
EuroSCORE	European System for Cardiac Operative Risk Evaluation
GAVS-II	German Aortic Valve Score (2nd update)
LogES I	Logistic EuroSCORE I
OBSERVANT	Observational study of appropriateness, efficacy and effectiveness of AVR-TAVR procedures for the treatment of severe symptomatic aortic stenosis
STS (PROM)	Society of Thoracic Surgeons Predicted Risk of Mortality
TAVR	Transcatheter aortic valve replacement

Introduction

Transcatheter aortic valve replacement (TAVR) is the treatment option of choice in inoperable and high-risk patients with severe aortic valve stenosis [1, 2] and recently showed favorable outcomes also in intermediate-risk [3, 4] and low-risk patients [5, 6]. Patient and procedural characteristics determine risk for adverse clinical outcomes. Cardiovascular societies recommend clinical decision-making for TAVR

within the Heart Team [1, 2], assisted by the use of statistical risk scoring models for individual patient risk stratification. Classical surgical risk models such as logistic EuroSCORE I (LogES I) [7, 8], its update EuroSCORE II (ES II) [9], or the Society of Thoracic Surgeons Predicted Risk of Mortality model (STS PROM) [10] are routinely used to gauge operative risk; however, they are known to overestimate mortality risk in TAVR [11–13]. Several TAVR-specific risk assessment models have been developed from national registry databases to optimize the predictive performance in the context of TAVR [14–16]. However, the optimal approach to risk prediction in TAVR still remains unknown, as their performance has been poor in validation studies [12, 17, 18].

We here aimed to evaluate the 30-day mortality prediction performance of six surgery- and TAVR-specific risk models (Table 1) in our German Rhine Transregio Aortic Diseases cohort of patients undergoing TAVR.

Methods

Study population and data collection

This study was performed as an all-comer analysis of patients treated with either TF or TA TAVR between 2008

Table 1 Risk model characteristics

Risk model	Year of publication	Recruitment period	Recruitment region	Cohort size: development/validation (<i>n</i>)	Procedure characteristics	Discrimination performance (<i>c</i> -indices in validation cohort)
LogES I [7, 8]	1999/2003	1995	Europe	13,302/1479	Cardiac surgery	0.76
ES II [9]	2012	2010	Worldwide, predominantly Europe	16,828/5553	Cardiac surgery	0.81
STS PROM [10]	2009	2002–2006	US	65,855/43,904	Cardiac valve surgery	0.80
FRANCE-2 [15]	2014	2010–2011	France + Monaco	2552/1281	TAVR (TF 73.4%, TA 17.8%, other 8.8%)	0.59
OBSERVANT [14]	2014	2010–2012	Italy	1256/622	TAVR	0.71
GAVS-II [16]	2017	2011–2012	Germany	9027/9027	55% surgical AVR, 45% TAVR	0.74

Characteristics of the analyzed risk models for the prediction of 30-day mortality risk in TAVR

and 2018 at the Heart Centers in Düsseldorf and Bonn, Germany, within the Rhine Transregio research consortium on aortic diseases. All patients were referred for TAVR procedures by local heart teams according to contemporary clinical practice and provided written informed consent for inclusion in prospective registries, with collection of clinical, procedural and follow-up data. The study was performed in accordance with the Declaration of Helsinki. The institutional Ethics Committee of the Heinrich-Heine University approved the study protocol (4080).

Clinical outcomes were systematically assessed using the Valve Academic Research Consortium consensus statement [19]. Primary clinical outcome of interest for risk model performance evaluation was 30-day mortality, secondary in-hospital outcomes are additionally reported.

Statistics

Statistical and graphical data analysis was performed using Excel (Microsoft, USA), SPSS (version 23.0, SPSS Inc., Chicago, IL, USA), MedCalc 18.10 (MedCalc Software, Belgium) and GraphPad Prism (version 7.0, Graphpad Software, San Diego, CA, USA). Continuous data are described as means \pm standard deviation (SD), ordinal/categorical data as counts and % of total, and receiver-operating-characteristic (ROC) curve analysis is summarized as *c*-indices (area-under-the-curve) with 95% confidence intervals (CI). Continuous variables were evaluated for normal distribution using the Kolmogorov–Smirnov test and compared with either Student's *t* tests (for normally distributed samples) or Mann–Whitney-*U* tests; categorical data were compared with Chi-squared tests

or Fisher's exact tests (for expected values < 5); ROC curves were compared using a non-parametric approach according to DeLong et al. [20]. All statistical tests were two tailed and an α probability of $p < 0.05$ was considered statistically significant.

Six risk models for the prediction of 30-day mortality (Table 1: LogES I [7, 8], ES II [9], STS PROM [10], FRANCE-2 [15], OBSERVANT [14] and GAVS-II [16]) were calculated for individual patients. LogES I, ES II and STS were routinely available in the databases and used for heart team decision-making. FRANCE-2, OBSERVANT and GAVS-II were calculated as follows: (1) relevant variables (for details, see Supplementary material) were weighed according to model definitions; (2) weighed variables were transformed into 30-day mortality probability predictions according to model definitions. Missing values necessary for risk model calculation were imputed to the mean (continuous) or median (categorical) of the whole population. Risk model discrimination accuracy was evaluated using ROC analysis and the *c*-index [area-under-the-curve (AUC)] as a cumulative measure; risk model calibration accuracy/goodness-of-fit was evaluated by stratification of patients into risk quintiles and comparison of observed vs. expected events within risk strata; additionally, it was formally tested by calculation of a logistic regression model for 30-day mortality, with the score value as independent variable and the Hosmer–Lemeshow goodness-of-fit test [21]. Prespecified analyses of patients stratified by TF vs. TA approach, time period of the procedure and prosthetic valve device types were additionally performed to account for learning curves and technical developments.

Results

Patient population

A total of 2,946 patients underwent TAVR between 2008 and 2018 at the German Heart Centers of Düsseldorf ($n = 1605$) and Bonn ($n = 1341$) and were included in this analysis, with 2,625 TF TAVR (89%) and 321 TA TAVR (11%) procedures, respectively. Mean age was 80.9 ± 6.1 years (Suppl. Table 1), patients suffered from a high-cardiovascular-risk profile, including comorbidities of cardiovascular disease (67% coronary artery disease), previous cardiac interventional or surgical procedures (39% previous PCI, 15% previous CABG) and diabetes mellitus (30%). Characteristics differed significantly between patients treated with TF and TA TAVR procedures (Suppl. Table 1).

Procedural characteristics across the TAVR era

TAVR developed from a rare procedure with high mortality (2008/2009: $n = 61$ with $\sim 10\%$ mortality) to a high-volume routine procedure (2018: $n = 538$ with 1.5% mortality) with low adverse events and shortened procedure duration (Table 2), which reflects a learning curve as well as continuous technical development. Patient age and estimated surgical mortality risk (LogES I) only declined marginally. TF TAVR became the primary access of choice (95.5% of procedures in 2018) over TA TAVR. The self-expandable Medtronic valves in different generations (early: CoreValve; newer generations: Evolut R and Evolut R Pro) were preferred to the balloon-expandable Edwards valves (early:

Sapien XT; newer generation: Sapien 3) in TF TAVR; the latter were the primary choice in TA TAVR (Suppl. Table 1).

Clinical outcomes

The primary outcome of 30-day mortality occurred in 3.7% overall, more likely in TA vs. TF TAVR patients (7.5% vs. 3.2%; $p < 0.0001$). Secondary clinical outcomes according to VARC-2 criteria are additionally reported in Suppl. Table 2.

Risk model performance for the prediction of 30-day mortality

Risk model characteristics are provided in Table 1 and in the Supplementary material.

Model discrimination

Risk model discrimination performance indices are reported in Table 3A, Fig. 1 and Table 4: ROC analyses showed mediocre performance of all risk models, without significant differences between them. Numerically, STS PROM (c -index 0.67, 95% CI 0.62–0.72) performed best, followed by FRANCE-2 (c -index 0.66, 95% CI 0.60–0.71) and ES II (c -index 0.65, 95% CI 0.60–0.70), OBSERVANT performed worst (c -index 0.60, 95% CI 0.55–0.66). All risk models performed worse than in their original validation cohorts (Table 1), except for FRANCE-2 (c -index 0.66 vs. 0.59 original). STS PROM showed the most consistent performance across TF and TA TAVR subgroups (both c -indices 0.66), while OBSERVANT performed numerically worst in both.

Table 2 Patient and procedural characteristics across the study period

Year	Age, years	LogES I	Patient count, n (%)	New gen. device, n (%)	TF/TA TAVR, n/n (% TF)	Procedure duration, min	Contrast volume, ml	30-Day mortality, n (%)
2008	84.5 ± 4.6	31.7 ± 17.6	17 (0.6)	0 (0)	17/0 (100)	59.1 ± 24.0	186.3 ± 45.8	1 (5.9)
2009	79.1 ± 6.7	27.5 ± 18.1	44 (1.5)	0 (0)	43/1 (97.7)	75.9 ± 29.1	217.1 ± 76.8	5 (11.4)
2010	82.5 ± 5.7	26.4 ± 14.5	78 (2.6)	0 (0)	60/18 (76.9)	82.6 ± 46.1	192.1 ± 80.9	5 (6.4)
2011	80.9 ± 6.3	23.2 ± 14.1	188 (6.4)	0 (0)	156/32 (83.0)	85.3 ± 39.0	166.9 ± 75.1	11 (5.9)
2012	81.1 ± 6.1	22.6 ± 14.2	118 (4.0)	0 (0)	88/30 (74.6)	88.6 ± 30.7	145.3 ± 88.6	9 (7.6)
2013	81.5 ± 6.1	22.5 ± 16.2	212 (7.2)	15 (7.1)	187/25 (88.2)	76.6 ± 27.0	128.7 ± 27.0	9 (4.2)
2014	80.7 ± 6.0	21.0 ± 15.0	297 (10.1)	125 (42.1)	252/45 (84.8)	78.2 ± 31.4	143.7 ± 60.2	13 (4.4)
2015	80.7 ± 5.9	26.0 ± 17.9	361 (12.3)	318 (88.1)	323/38 (89.5)	96.9 ± 43.9	138.7 ± 49.8	14 (3.9)
2016	80.4 ± 6.0	24.6 ± 15.8	466 (15.8)	461 (98.9)	406/60 (87.1)	87.2 ± 35.2	121.7 ± 39.3	12 (2.6)
2017	81.0 ± 6.3	23.7 ± 16.1	627 (21.3)	627 (100)	579/48 (92.3)	71.5 ± 32.9	119.9 ± 46.5	21 (3.4)
2018	80.7 ± 6.1	21.0 ± 15.0	538 (18.3)	538 (100)	514/24 (95.5)	64.9 ± 27.4	106.1 ± 46.1	8 (1.5)
Overall	80.9 ± 6.1	23.4 ± 15.9	2946 (100)	2084 (70.7)	2625/321 (89.1)	78.7 ± 35.5	130.3 ± 56.0	108 (3.7)

Patient and procedural characteristics of TAVR from 2008 to 2018 and overall means/counts. Continuous variables are displayed as mean with standard deviation, categorical variables as n (%), unless specified differently

TF transfemoral, TA transapical, TAVR transcatheter aortic valve replacement, LogES I Logistic EuroScore I

Table 3 Risk model performance for the prediction of 30-day mortality risk in TAVR

Risk model	(A) Model discrimination: ROC analysis			(B) Model calibration: prediction of 30-day mortality risk		
	All patients (<i>n</i> = 2946)	TF TAVR (<i>n</i> = 2625)	TA TAVR (<i>n</i> = 321)	All patients (<i>n</i> = 2946)	TF TAVR (<i>n</i> = 2625)	TA TAVR (<i>n</i> = 321)
LogES I [7, 8]	0.63 (0.58–0.68)	0.61 (0.55–0.67)	0.67 (0.56–0.78)	23.4 ± 15.9 (1.1–95.0)	22.8 ± 15.7* (1.1–95.0)	28.0 ± 16.9* (2.5–88.4)
ES II [9]	0.65 (0.60–0.70)	0.64 (0.58–0.70)	0.67 (0.55–0.79)	7.5 ± 7.4 (0.5–78.1)	7.3 ± 7.2* (0.5–78.1)	9.3 ± 8.8* (1.0–56.1)
STS PROM [10]	0.67 (0.62–0.72)	0.66 (0.60–0.72)	0.66 (0.57–0.76)	6.2 ± 5.7 (0.5–69.3)	6.1 ± 5.6* (0.6–69.3)	7.5 ± 6.7* (0.6–58.1)
FRANCE-2 [15]	0.66 (0.60–0.71)	0.63 (0.57–0.69)	0.67 (0.55–0.79)	9.2 ± 6.8 (4.0–62.0)	8.3 ± 5.6* (4.0–55.0)	16.3 ± 10.7* (6.8–62.0)
OBSERVANT [14]	0.60 (0.55–0.66)	0.60 (0.54–0.67)	0.58 (0.47–0.68)	5.8 ± 5.0 (1.8–58.7)	5.7 ± 5.0 (1.8–58.7)	6.2 ± 4.9 (1.8–33.2)
GAVS-II [16]	0.63 (0.58–0.69)	0.62 (0.56–0.69)	0.67 (0.56–0.78)	9.4 ± 9.6 (1.0–89.8)	9.2 ± 9.1 (0.9–89.8)	10.9 ± 12.4 (1.1–80.8)

Summary of risk model performance indices for the prediction of 30-day mortality risk in TAVR. (A) Model discrimination, reported as *c*-indices (area-under-the-ROC-curve, Fig. 1) with 95% confidence intervals (in brackets); (B) model calibration reported as predicted mean 30-day mortality risk (in %) ± standard deviations (in %) and risk range (in brackets, in %)

TF transfemoral, TA transapical, TAVR transcatheter aortic valve replacement

*Statistically significant differences between TF and TA TAVR groups

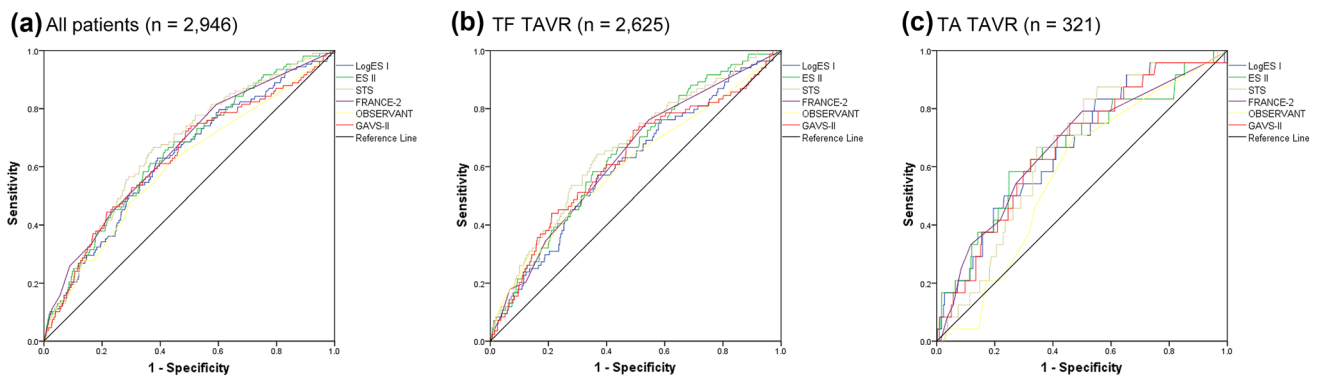


Fig. 1 Risk model discrimination for the prediction of 30-day mortality. Model discrimination (ROC curves) of the six risk models for the prediction of 30-day mortality, for all studied patients (a), patients with TF TAVR only (b) and patients with TA TAVR only (c)

The stratified analyses for time period and for new vs. old generation devices (Table 4) showed no significant differences and also no visible performance trend in either risk model over the years.

Model calibration

Mean predicted mortality risk (Table 3B) exceeded observed mortality in all models, ranging from $5.8 \pm 5.0\%$ (OBSERVANT) to $23.4 \pm 15.9\%$ (LogES I). Predictions were significantly higher in the TA TAVR subgroup for LogES I, ES II, STS PROM and FRANCE-2, while the OBSERVANT and GAVS-II models showed no differences between subgroups.

Graphical analysis of calibration is displayed in Fig. 2. Overestimation of mortality risk was especially pronounced in high-risk patients, while the lower risk quintiles were more adequately calibrated (e.g. STS PROM, ES II, OBSERVANT models). The classical LogES I surgical model grossly overestimated mortality in all risk strata.

Results of formal statistical testing for goodness-of-fit are displayed in Suppl. Table 3: STS PROM and GAVS-II showed significant Hosmer–Lemeshow *p* values (=inadequate calibration) of the respective logistic regression models for the whole population, GAVS-II also across TF and TA TAVR subgroups. OBSERVANT could not be calculated due to lack of events in some risk groups.

Table 4 Risk model discrimination stratified by time period and device type

Risk model	2008–2010 (n = 139)	2011–2013 (n = 518)	2014–2016 (n = 1124)	2017–2018 (n = 1165)	Old-generation TAVR device (n = 862)	New-generation TAVR device (n = 2084)	Overall (n = 2946)
LogES I [7, 8]	0.67 (0.50–0.85)	0.59 (0.48–0.70)	0.68 (0.59–0.77)	0.57 (0.47–0.68)	0.63 (0.55–0.71)	0.63 (0.56–0.71)	0.63 (0.58–0.68)
ES II [9]	0.65 (0.46–0.83)	0.57 (0.48–0.67)	0.70 (0.61–0.78)	0.65 (0.55–0.74)	0.63 (0.56–0.71)	0.66 (0.59–0.73)	0.65 (0.60–0.70)
STS PROM [10]	0.53 (0.32–0.74)	0.64 (0.56–0.73)	0.67 (0.58–0.76)	0.64 (0.54–0.74)	0.63 (0.56–0.70)	0.65 (0.57–0.72)	0.67 (0.62–0.72)
FRANCE-2 [15]	0.63 (0.49–0.76)	0.63 (0.51–0.74)	0.66 (0.57–0.75)	0.66 (0.57–0.76)	0.61 (0.53–0.69)	0.68 (0.61–0.76)	0.66 (0.60–0.71)
OBSERVANT [14]	0.61 (0.41–0.82)	0.53 (0.41–0.65)	0.68 (0.60–0.76)	0.57 (0.47–0.68)	0.57 (0.49–0.66)	0.63 (0.56–0.70)	0.60 (0.55–0.66)
GAVS-II [16]	0.58 (0.37–0.79)	0.54 (0.42–0.66)	0.68 (0.60–0.77)	0.66 (0.57–0.76)	0.59 (0.50–0.68)	0.67 (0.60–0.74)	0.63 (0.58–0.69)

Risk model discrimination performance across the study period from 2008 to 2018, stratified by time period and device type, in comparison to the overall mean, displayed as c-indices (area-under-the-ROC-curve) with 95% confidence intervals (in brackets)

TAVR transcatheter aortic valve replacement

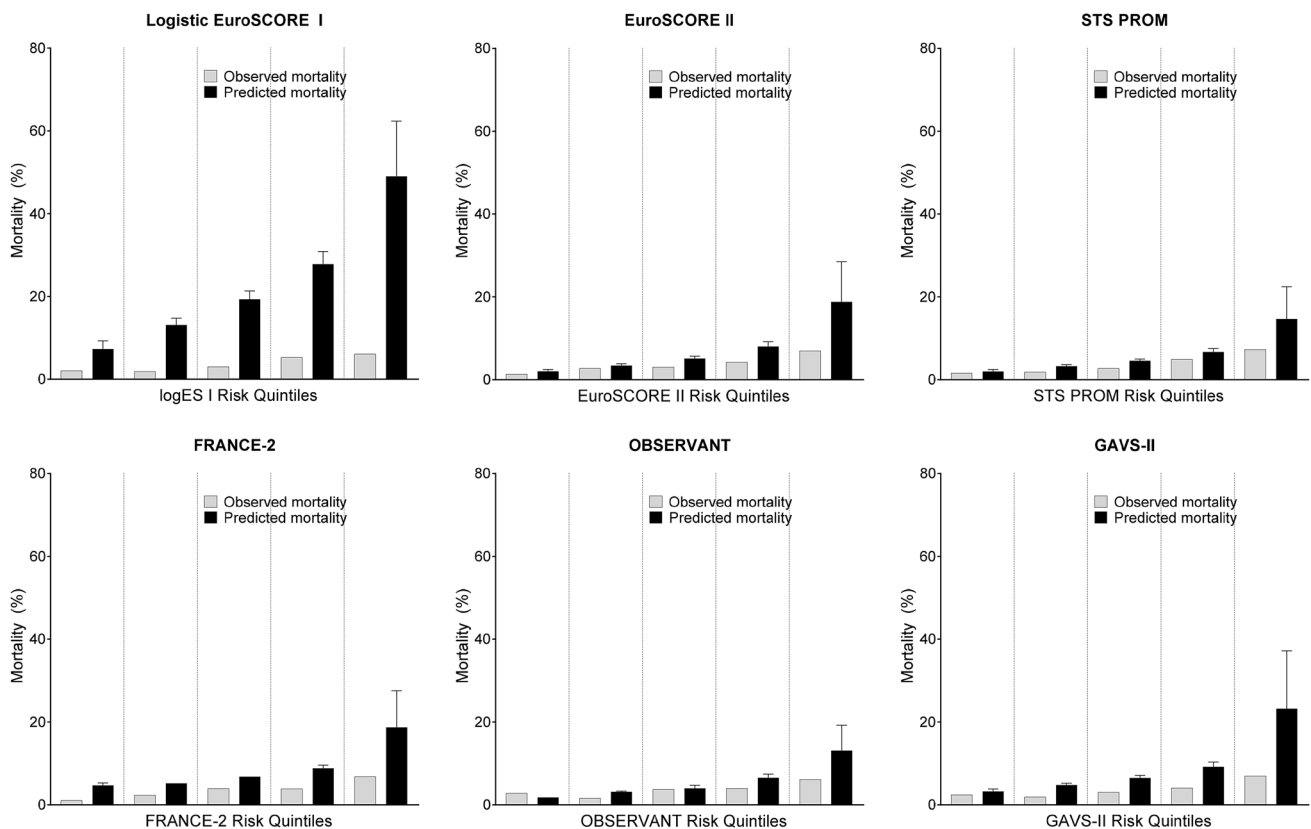


Fig. 2 Risk model calibration for the prediction of 30-day mortality. Graphical model calibration comparison of the six risk models, stratified into risk quintiles for observed vs. predicted 30-day mortality

Discussion

The main results of this comparative external validation study of six risk models for the prediction of 30-day mortality in patients undergoing TAVR at two German

high-volume Heart Centers from 2008 to 2018 are (1) surgical and TAVR-specific risk models showed similarly mediocre discrimination performance (*c*-indices 0.60–0.67); (2) all models overestimated 30-day mortality risk and were poorly calibrated, especially in high-risk

patients; (3) no significant influence of time of procedure or device type on the results could be found.

Patient risk stratification for coronary procedures is well established—ranging from the elective setting to acute coronary syndromes [22–24]—but the risk of TAVR procedures is thus far considerably less predictable: Patients suffer from a multitude of interdependent comorbid conditions, especially *frailty* as an essential risk factor is difficult to classify [25, 26]. The expansion of TAVR from highest risk/inoperable patients into intermediate- and lower risk groups introduces bias. On the procedural side, there are clear associations of center volume [27], operator experience [28], and choice of approach (TA vs. TF) [29] with clinical outcomes.

Classical surgical risk models (LogES I, STS PROM, ES II) have well-known limitations in TAVR [13]. LogES I is not recommended for use in TAVR anymore (1), we included the model for historical comparisons. National TAVR-registries have provided platforms for the development of TAVR-specific risk models [14–16], which were mostly developed in the “early years” of TAVR (Table 1). Resulting limitations in their performance have already been seen when applied outside of their original populations [30, 31], external validation studies with a similar design to our study were performed in the United Kingdom [12] and in the Netherlands [18]—and also retrieved disappointing results.

This analysis from the German Transregio Aortic Diseases cohort elucidates risk prediction in the whole era of TAVR, which has developed from an experimental procedure in 2008 to a routine and high-volume alternative to surgical aortic valve replacement in 2018. With steadily increasing operator experience and new-generation devices, procedure counts have dramatically increased and complication rates declined over the study period. The TF approach has become the routine access route.

Surgical risk models performed similar to expectations [13]: Discrimination analysis essentially showed similar performance of STS PROM [10], LogES I [7, 8] and ES II [9] models. None could even remotely match discrimination performance in their original surgical patient cohorts. Calibration analysis underlined the general overestimation of 30-day mortality risk—most pronounced in the LogES I model. Surgical models are thus confirmed to have severe limitations to judge TAVR risk.

However, dedicated TAVR-specific models also disappointed: FRANCE-2 (overall *c*-index 0.66, 95% CI 0.60–0.71) was better than in the original validation cohort (*c*-index 0.59 [15]) and in the UK (*c*-index 0.62 [12]) and Netherlands (*c*-index 0.63 [18]) external validation studies, but not superior to surgical models and slightly worse than in an Israeli external validation study (*c*-index 0.71 [30]). The model considerably overestimated risk (Fig. 2) in our patients. The GAVS-II model [16] derived from the German Aortic Valve Registry was expected to be most adapted

to German TAVR conditions, but it performed similar to the other models in the overall cohort (*c*-index 0.63) and in TF and TA subgroups and could not meet the performance in the development cohort (*c*-index 0.74), similar to the Dutch external validation study [18]. Numerically, its discrimination improved over time and in new vs. old generation devices (Table 4). The model also overestimated mortality (expected: 9.4%; Fig. 2). Reasons for the lower-than-expected performance may lie in the GAVS-II model being developed in 55% surgical/45% TAVR patients from 2011–2012 (Table 1), which significantly differed in age (74 vs. 81 years) and comorbidities from our population [16]. The OBSERVANT model [14] discriminated numerically worst in the overall cohort and in TF/TA subgroups, and could not match performance in the original population (*c*-index 0.71), which confirms findings in other external validation studies [12, 18, 30]. However—of all analyzed models—its mean mortality prediction (5.8%) came closest to observed events (3.7%).

Taken together, this analysis clarifies that risk prediction in TAVR is still an unsolved issue: all tested models are severely limited in their performance, and dedicated TAVR-specific models are not superior to decade-old surgical scores. Using any of these models for risk stratification is better than a coin-flip—but with ample room for improvement. A recently published analysis from the Netherlands [18] found similar results to the UK study [12] and to our results, which supports the validity of these findings across Europe. With TAVR evermore becoming a routine procedure and 30-day mortality rates reaching all-time-lows in high-volume Heart Centers, a re-calibration of existing risk models or a new development from growing registries is thus mandatory to improve accuracy [32, 33]. High-quality TAVR databases should enable us to produce risk models with performance comparable to surgical procedures or acute coronary syndrome. Incorporation of functional status/frailty assessment as important predictive factors may additionally improve model accuracy in the TAVR setting [34, 35].

Study limitations

All conclusions from our study are limited to risk model performance in German—or at best European—TAVR patients, findings might be different in other patient populations and procedural conditions. Bias may originate from retrospective calculation of FRANCE-2, OBSERVANT and GAVS-II risk models. While data quality in our prospective databases was high at all times, there is unavoidable risk for bias by changes in clinical practice and adverse event rates from 2008 to 2018: We accounted for this with sub-analyses of patients stratified by time of procedure and device generations (Table 4); however, their statistical power is limited.

Conclusions

Three surgical as well as three TAVR-specific risk scoring models showed mediocre performance in prediction of 30-day mortality risk for TAVR in the German Rhine Transregio Aortic Diseases cohort. Development of new or updated risk models is necessary to improve risk stratification.

Acknowledgements We thank Jenni Scharlau for substantial contribution in data acquisition.

Funding Open Access funding provided by Projekt DEAL. This work was supported by the Forschungskommission of the Medical Faculty of the Heinrich-Heine-University Düsseldorf (No. 2018-32 to GW) for a Clinician Scientist Track and by the Deutsche Forschungsgemeinschaft for the collaborative research center Transregio 259: Aortic diseases. There was no financial or other support from the industry.

Code availability Not applicable.

Compliance with ethical standards

Conflict of interest Georg Wolff, Christian Jung, Kathrin Klein, Amin Polzin, Ralf Westenfeld, Tobias Zeus, and Verena Veulemans have received consulting fees, travel expenses or study honoraria from Medtronic and/or Edwards Lifesciences. Jan-Malte Sinning and Georg Nickenig have received speaker honoraria and research grants from Abbott, Abiomed, Medtronic, Boston Scientific, and Edwards Lifesciences. Dr. Grube is a proctor for Medtronic and Boston Scientific. All other authors have nothing to disclose with regard to this project.

Ethics approval This study was ethics approved (see “Methods”) and performed in accordance with the Declaration of Helsinki.

Consent to participate All patients provided written informed consent to participate in TAVR registries at both sites.

Consent for publication All authors read and consented to the final manuscript version for publication.

Availability of data and material All authors had access to the data used in this study. Data can be made available upon request.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Baumgartner H, Falk V, Bax JJ, De Bonis M, Hamm C, Holm PJ et al (2017) 2017 ESC/EACTS guidelines for the management of valvular heart disease. *Eur Heart J* 38(36):2739–2791
- Nishimura RA, Otto CM, Bonow RO, Carabello BA, Erwin JP 3rd, Fleisher LA et al (2017) 2017 AHA/ACC focused update of the 2014 AHA/ACC guideline for the management of patients with valvular heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol* 70(2):252–289
- Leon MB, Smith CR, Mack MJ, Makkar RR, Svensson LG, Kodali SK et al (2016) Transcatheter or surgical aortic-valve replacement in intermediate-risk patients. *N Engl J Med* 374(17):1609–1620
- Reardon MJ, Van Mieghem NM, Popma JJ, Kleiman NS, Sondergaard L, Mumtaz M et al (2017) Surgical or transcatheter aortic-valve replacement in intermediate-risk patients. *N Engl J Med* 376(14):1321–1331
- Mack MJ, Leon MB, Thourani VH, Makkar R, Kodali SK, Russo M et al (2019) Transcatheter aortic-valve replacement with a balloon-expandable valve in low-risk patients. *N Engl J Med* 380(18):1695–1705
- Popma JJ, Deeb GM, Yakubov SJ, Mumtaz M, Gada H, O’Hair D et al (2019) Transcatheter aortic-valve replacement with a self-expanding valve in low-risk patients. *N Engl J Med* 380(18):1706–1715
- Roques F, Michel P, Goldstone AR, Nashef SA (2003) The logistic EuroSCORE. *Eur Heart J* 24(9):881–882
- Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R (1999) European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg* 16(1):9–13
- Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR et al (2012) EuroSCORE II. *Eur J Cardiothorac Surg* 41(4):734–44 (discussion 44–5)
- O’Brien SM, Shahian DM, Filardo G, Ferraris VA, Haan CK, Rich JB et al (2009) The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2—isolated valve surgery. *Ann Thorac Surg* 88(1 Suppl):S23–42
- Piazza N, Wenaweser P, van Gameren M, Pilgrim T, Tzikas A, Otten A et al (2010) 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* 159(2):323–329
- Martin GP, Sperrin M, Ludman PF, de Belder MA, Gale CP, Toff WD et al (2017) Inadequacy of existing clinical prediction models for predicting mortality after transcatheter aortic valve implantation. *Am Heart J* 184:97–105
- Wang TKM, Wang MTM, Gamble GD, Webster M, Ruygrok PN (2017) Performance of contemporary surgical risk scores for transcatheter aortic valve implantation: a meta-analysis. *Int J Cardiol* 236:350–355
- Capodanno D, Barbanti M, Tamburino C, D’Errigo P, Ranucci M, Santoro G et al (2014) A simple risk tool (the OBSERVANT score) for prediction of 30-day mortality after transcatheter aortic valve replacement. *Am J Cardiol* 113(11):1851–1858
- Iung B, Laouenan C, Himbert D, Eltchaninoff H, Chevreul K, Donzeau-Gouge P et al (2014) Predictive factors of early mortality after transcatheter aortic valve implantation: individual risk assessment using a simple score. *Heart (British Cardiac Society)* 100(13):1016–1023
- Schiller W, Barnewold L, Kazmaier T, Beckmann A, Masseli F, Welz A et al (2017) The German aortic valve score II. *Eur J Cardiothorac Surg* 52(5):881–887

17. Martin GP, Sperrin M, Mamas MA (2018) Pre-procedural risk models for patients undergoing transcatheter aortic valve implantation. *J Thoracic Dis* 10(Suppl 30):S3560–S3567
18. Al-Farra H, Abu-Hanna A, de Mol BAJM, ter Burg WJ, Houterman S, Henriques JPS et al (2020) External validation of existing prediction models of 30-day mortality after transcatheter aortic valve implantation (TAVI) in the Netherlands Heart Registration. *Int J Cardiol*
19. Kappetein AP, Head SJ, Genereux P, Piazza N, van Mieghem NM, Blackstone EH et al (2013) Updated standardized endpoint definitions for transcatheter aortic valve implantation: the Valve Academic Research Consortium-2 consensus document. *J Thoracic Cardiovasc Surg* 145(1):6–23
20. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44(3):837–845
21. Hosmer DWJ, Lemeshow SXS (2013) Applied logistic regression, 3rd edn. Wiley, Hoboken
22. Brennan JM, Curtis JP, Dai D, Fitzgerald S, Khandelwal AK, Spertus JA et al (2013) Enhanced mortality risk prediction with a focus on high-risk percutaneous coronary intervention: results from 1,208,137 procedures in the NCDR (National Cardiovascular Data Registry). *JACC Cardiovasc Interv* 6(8):790–799
23. Fox KA, Dabbous OH, Goldberg RJ, Pieper KS, Eagle KA, Van de Werf F et al (2006) Prediction of risk of death and myocardial infarction in the six months after presentation with acute coronary syndrome: prospective multinational observational study (GRACE). *BMJ* 333(7578):1091
24. Wolff G, Lin Y, Quade J, Bader S, Kosejian L, Brockmeyer M et al (2019) Validation of National Cardiovascular Data Registry risk models for mortality, bleeding and acute kidney injury in interventional cardiology at a German Heart Center. *Clin Res Cardiol*
25. Flaatten H, De Lange DW, Morandi A, Andersen FH, Artigas A, Bertolini G et al (2017) The impact of frailty on ICU and 30-day mortality and the level of care in very elderly patients (\geq 80 years). *Intensive Care Med* 43(12):1820–1828
26. Shimura T, Yamamoto M, Kano S, Kagase A, Kodama A, Koyama Y et al (2017) Impact of the clinical frailty scale on outcomes after transcatheter aortic valve replacement. *Circulation* 135(21):2013–2024
27. Wassef AWA, Rodes-Cabau J, Liu Y, Webb JG, Barbanti M, Munoz-Garcia AJ et al (2018) The learning curve and annual procedure volume standards for optimum outcomes of transcatheter aortic valve replacement: findings from an International Registry. *JACC Cardiovasc Interv* 11(17):1669–1679
28. Salemi A, Sedrakyan A, Mao J, Elmously A, Wijeyesundera H, Tam DY et al (2019) Individual operator experience and outcomes in transcatheter aortic valve replacement. *JACC Cardiovasc Interv* 12(1):90–97
29. Kumar N, Khara R, Fonarow GC, Bhatt DL (2018) Comparison of outcomes of transfemoral versus transapical approach for transcatheter aortic valve implantation. *Am J Cardiol* 122(9):1520–1526
30. Halkin A, Steinvil A, Witberg G, Barsheshet A, Barkagan M, Assali A et al (2016) 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* 215:227–231
31. Pilgrim T, Franzone A, Stortecky S, Nietlispach F, Haynes AG, Tueller D et al (2017) Predicting mortality after transcatheter aortic valve replacement: external validation of the transcatheter valve therapy registry model. *Circ Cardiovasc Intervent* 10(11)
32. Hickey GL, Grant SW, Murphy GJ, Bhabra M, Pagano D, McAllister K et al (2013) Dynamic trends in cardiac surgery: why the logistic EuroSCORE is no longer suitable for contemporary cardiac surgery and implications for future risk models. *Eur J Cardiothorac Surg* 43(6):1146–1152
33. Matheny ME, Ohno-Machado L, Resnic FS (2005) Discrimination and calibration of mortality risk prediction models in interventional cardiology. *J Biomed Inform* 38(5):367–375
34. Martin GP, Sperrin M, Ludman PF, deBelder MA, Gunning M, Townend J et al (2018) Do frailty measures improve prediction of mortality and morbidity following transcatheter aortic valve implantation? An analysis of the UK TAVI registry. *BMJ Open* 8(6):e022543
35. van der Wulp K, van Wely MH, Schoon Y, Vart P, Olde Rikkert MGM, Morshuis WJ et al Geriatric assessment in the prediction of delirium and long-term survival after transcatheter aortic valve implantation. *J Thorac Cardiovasc Surg*