ESC European Society of Cardiology

European Heart Journal - Digital Health (2023) **4**, 225–235
https://doi.org/10.1093/ehjdh/ztad021

ORIGINAL ARTICLE

Development and validation of explainable machine learning models for risk of mortality in transcatheter aortic valve implantation: TAVI risk machine scores

Andreas Leha [1,2], Cynthia Huber [1], Tim Friede [1,2], Timm Bauer[3], Andreas Beckmann[4,5], Raffi Bekeredjian [6], Sabine Bleiziffer [7], Eva Herrmann [8,9], Helge Möllmann[10], Thomas Walther [11], Friedhelm Beyersdorf [12,13], Christian Hamm[14,15], Arnaud Künzi [16], Stephan Windecker [17], Stefan Stortecky [17], Ingo Kutschka [18], Gerd Hasenfuß [2,19], Stephan Ensminger [20,22], Christian Frerker[21,22], and Tim Seidler [2,19,*]

[1]Department of Medical Statistics, University Medical Center Göttingen, Humboldtallee 32, 37073 Göttingen, Germany; [2]DZHK (German Center for Cardiovascular Research), Partner Site Göttingen, Robert-Koch str. 40, 37075 Göttingen, Germany; [3]Department of Cardiology, Sana Klinikum Offenbach, Starkenburgring 66, 63069 Offenbach am Main, Germany; [4]German Society for Thoracic and Cardiovascular Surgery, Langenbeck-Virchow-Haus, Luisenstraße 58/59, 10117 Berlin, Germany; [5]Department for cardiac and pediatric cardiac surgery, Heart Center Duisburg, EVKLN, Gerrickstr. 21, 47137 Duisburg, Germany; [6]Department of Cardiology, Robert-Bosch-Krankenhaus, Auerbachstraße 110, 70376 Stuttgart, Germany; [7]Clinic for Thoracic and Cardiovascular Surgery, Heart and Diabetes Center Northrhine-Westphalia, Georgstr 11, 32545 Bad Oeynhausen, Germany; [8]Goethe University Frankfurt, Department of Medicine, Institute of Biostatistics and Mathematical Modelling, Theodor-Stern-Kai 7, 60590 Frankfurt Main, Germany; [9]DZHK (German Centre for Cardiovascular Research), Partner Site Rhine/Main, Theodor-Stern-Kai 7, 60590 Frankfurt Main, Germany; [10]Department of Cardiology, St.-Johannes-Hospital Dortmund, Johannesstrasse 9-17, 44137 Dortmund, Germany; [11]Department of Cardiothoracic Surgery, University Hospital Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt, Germany; [12]Medical Faculty of the Albert-Ludwigs-University Freiburg, University Hospital Freiburg, Hugstetterstr. 55, 79106 Freiburg, Germany; [13]Department of Cardiovascular Surgery, Heart Centre Freiburg University, Freiburg, Germany; [14]Department of Cardiology and Angiology, University Hospital Gießen, Klinikstr. 33, 35392 Gießen, Germany; [15]Department of Cardiology, Kerckhoff Heart and Thorax Center, Benekestraße 2-8, D-61231 Bad Nauheim, Germany; [16]CTU Bern, University of Bern, Mittelstrasse 43, 3012 Bern, Switzerland; [17]Department of Cardiology, Inselspital, Bern University Hospital, University of Bern, 3010 Bern, Switzerland; [18]Clinic for Cardiothoracic and Vascular Surgery/Heart Center, University Medical Center Göttingen, Robert-Koch Str. 40, 37075 Göttingen, Germany; [19]Clinic for Cardiology and Pulmonology, Heart Center, University Medical Center Göttingen, Robert-Koch Str. 40, 37075 Göttingen, Germany; [20]Department of Cardiac and Thoracic Vascular Surgery, University Heart Center Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany; [21]Department of Cardiology, University Heart Center Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany; and [22]DZHK (German Centre for Cardiovascular Research), partner site Hamburg/Kiel/Lübeck, Lübeck, Germany

Received 27 October 2022; revised 28 February 2023; accepted 16 March 2023; online publish-ahead-of-print 17 March 2023

Aims Identification of high-risk patients and individualized decision support based on objective criteria for rapid discharge after transcatheter aortic valve implantation (TAVI) are key requirements in the context of contemporary TAVI treatment. This study aimed to predict 30-day mortality following TAVI based on machine learning (ML) using data from the German Aortic Valve Registry.

Methods and results Mortality risk was determined using a random forest ML model that was condensed in the newly developed TAVI Risk Machine (TRIM) scores, designed to represent clinically meaningful risk modelling before (TRIMpre) and in particular after (TRIMpost) TAVI. Algorithm was trained and cross-validated on data of 22 283 patients (729 died within 30 days post-TAVI) and generalisation was examined on data of 5864 patients (146 died). TRIMpost demonstrated significantly better performance than traditional scores [*C*-statistics value, 0.79; 95% confidence interval (CI)] [0.74; 0.83] compared to Society of Thoracic Surgeons (STS) with *C*-statistics value 0.69; 95%-CI [0.65; 0.74]). An abridged (aTRIMpost) score comprising 25 features (calculated using a web interface) exhibited significantly higher performance than traditional scores (*C*-statistics value, 0.74; 95%-CI [0.70; 0.78]). Validation on external data of 6693 patients (205 died within 30 days post-TAVI) of the Swiss TAVI Registry confirmed significantly better performance for the TRIMpost (*C*-statistics value 0.75, 95%-CI [0.72; 0.79]) compared to STS (*C*-statistics value 0.67, CI [0.63; 0.70]).

* Corresponding author. Tel: +49 (0) 551/39-63907, Fax: +49(0)551/39-63906, Email: tim.seidler@med.uni-goettingen.de

Conclusion TRIM scores demonstrate good performance for risk estimation before and after TAVI. Together with clinical judgement, they may support standardised and objective decision-making before and after TAVI.

Structured Graphical Abstract

Key Question

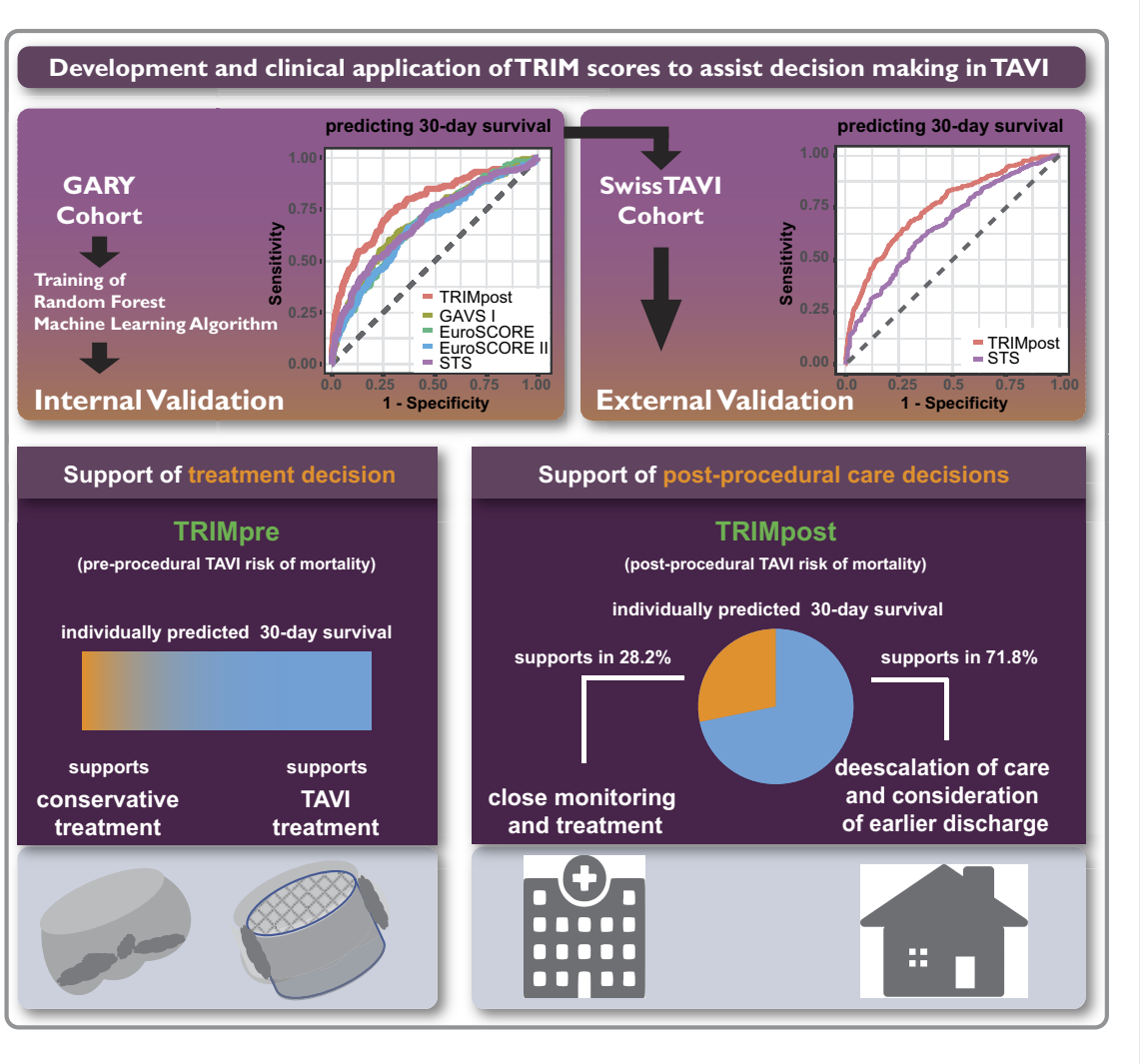
Given that current decision support tools for transcatheter aortic valve implantation (TAVI) are limited to relatively simple models and that guidelines still recommend scores that were not developed specifically for TAVI, the question arises whether machine learning (ML) can be used to create improved and interpretable risk models.

Key Finding

In this diagnostic study of the nationwide cohort of patients undergoing TAVI in Germany, we observed that risk assessment both before and especially after TAVI can be significantly improved using ML-based risk models. The prognostic value of our TRIM scores was confirmed in the Swiss TAVI Registry cohort and outperformed established risk scores. TRIM scores can be interrogated for the influential factors as well as the profile of this influence.

Take Home Message

Our results argue for the use of TRIM scores in routine clinical care, while improving the interpretability of risk assessment through an app that visually explains the variables that influence an individual's risk assessment.



Keywords TAVI • machine learning • risk score • random forest • decision support

Introduction

Aortic valve stenosis can be treated with either surgical aortic valve replacement (SAVR) or transcatheter aortic valve implantation (TAVI). Statistical models for predicting the procedural risk based on preprocedural data have been widely recommended.[1] Scores are usually helpful when the number of variables and their complex interplay place high demands on clinical judgment skills and an objective appreciation of the variables is desired. Given the recent option for TAVI treatment across the entire spectrum of surgical risk,[2,3] general surgical risk scores remain important to objectively support the identification of patients not suited for SAVR. However, while there are TAVI specific scores,[4–10] none of the common scores were specifically developed for a TAVI/SAVR procedure or population. In addition, post-TAVI risk models are lacking, although diverse variables (some of which are intraprocedural) need to be integrated for postinterventional care decision-making. Thus, a post-TAVI risk model could support decisions between post-TAVI treatment approaches, such as fast but safe discharge or closer surveillance. Machine learning (ML) as an approach to build predictive models is especially capable of modelling non-linear and complex interdependencies between a possibly large set of predictors and has, thus, the potential to build particularly strong models in feature rich settings. Prediction of patient risk has been attempted using ML methods for TAVI[4–10] and other cardiac treatments,[11–15] but several challenges have precluded the translation of trained models from bench to bedside.[16] These challenges include the lack of prospective evaluations or even randomized controlled trials for ML models, a large body of models published only as preprints without peer-review, the failure to follow reporting standards such as transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD), failure to capture or assess dataset shift due to model fitting on historic data, difficulties to generalise to new populations, and hard to interpret black-box models. While this manuscript presents a retrospective analysis, we try to address many of the other mentioned challenges.

In this diagnostic study, we provide a ML-based risk scoring for post-TAVI risk (TRIMpost) to support objective decision-making between treatment approaches. Furthermore, a pre-TAVI (TRIMpre) scoring model that outperformed classical models is provided. Both models highlight the individuals risk context with the variables that influence it, leaving the physician not only with a score, but with the variables that impacted on the score (explainable ML). As ML performance relies on a large dataset, our modelling used the German Aortic Valve Registry (GARY), one of the world's largest aortic valve replacement registries combining data from over 90 centers.[17] The performance of the ML models was evaluated on a separated test set from GARY and validated on external data from the Swiss TAVI Registry (SwissTAVI).[18] The ML models were interrogated for interpretation. An interactive easy-to-use web app that calculates the score and provides insight for decision-making is also described. We suggest TRIM scores as the basis for advanced decision support in the course of TAVI treatment.

Methods

This study is reported in adherence to the TRIPOD guidelines[19] (see Supplementary material online, *Figure S1*). Full details on the methods are provided in the Supplement.

Study population

The analysed cohort comprised all completed TAVI interventions through transfemoral access with annotated survival status at 30 days post-TAVI recorded in GARY between 2011 and 2017.[17] GARY (NCT01165827) was approved by local ethics committees of all participating centres and patients gave informed consent to the analysis of the data. External validation was performed by using patient data from the SwissTAVI between 2014 and 2018.[18] The SwissTAVI (NCT01368250) is a national prospective cohort study, which is mandated by the Swiss Federal Office of Public Health to include all consecutive patients undergoing TAVI at approved sites in Switzerland since 2011.

Outcomes

To support clinical decision-making towards early discharge, mortality 30 days post-TAVI was used as study endpoint.

Existing scores evaluated in this study

Several established scores served as a baseline for comparison in this study: the logistic EuroSCORE,[20] EuroSCORE II,[21] German aortic valve score (GAVS),[22] and Society of Thoracic Surgeons (STS) Valve score.[23] However, none of these scores is TAVI specific. The logistic EuroSCORE and EuroSCORE II are surgical risk scores; the GAVS and STS Valve score are specific for valve procedures. We also evaluated the TAVI-specific TVT in both versions (with NYHA (TVT(NYHA))[24] and with health status and gait speed (TVT(GaitSpeed)),[25]) but many variables used in these scores were missing, so that we refrained from a formal comparison.

Potential predictors

Only those ($k = 97$) variables in the GARY registry that are assessed before or immediately after TAVI ($k = 58$) were considered.

Training and validation

The training and validation pipeline is visualised in Supplementary material online, *Figure S2*.

To avoid information leakage, great care was taken to conduct all training and tuning of the ML method using only the training set. The test set was used only to estimate the generalization performance in the last step after the model tuning. Due to effects of data shift between historic and future cohorts and progress in medical care in general and TAVI treatment in particular, we expect differences in TAVI cohorts as well as in our outcome (30-day mortality) over time. To assess the robustness of our model towards the evolution of TAVI, the training test split was done time-based: the training was done on cases from 2011 to 2016 ($n = 22\,283$) the validation on cases from 2017 ($n = 5864$ for (a)TRIMpost and 5926 for (a) TRIMpre).

ML-technique: random forest

Among several ML methods (model based recursive partitioning, deep learning, and random forest), the random forest yielded the best performing models. Parameters of the random forest (RF)[26] were tuned on the training data via grid search using 5-fold cross validation.

Improving class balance in the training data is beneficial to the performance of RF.[27] Therefore, patients with events within 30 days were selected with increased probability during the training.

When using the trained model to predict new data, this new data might have missing values where the training data did not. In order to allow the trained model to be applicable to new data with such missing values, the training data was extended by duplicating for each variable of interest randomly (stratified by the variable of interest) chosen 10% of the data while setting the variable of interest to a new missing category (categorical variable) or missing (numerical variable) as suggested in Chollet and Allaire.[28]

The character of missing values was examined by testing missing value indicators for independence from the outcome and by Jamshidian and Jalal's non-parametric missing completely at random (MCAR) test[29] on the numeric variables with missing values. Missing values in categorical variables were set to a new 'missing' category, missing values in the numeric variables were imputed by median. As a sensitivity analysis multiple imputation by chained equations was performed to check for influence on classification performance or variable importance.

The RF models were trained twice: using all available features (TRIM scores) and using selected features (abridged TRIM scores).

For feature selection, an initial random forest classifier was fit to the training data, variable importance was calculated for all features, and the most important features were selected.

Model interpretation

Variable importance[30] in the RF models was measured as the total decrease in node impurities from splitting on the variable, averaged over all trees, where the node impurity was measured by the Gini index.

Breakdown plots[31] were generated to interpret risk scores for individual patient predictions, highlighting the shift in expected prediction induced by a specific value of a variable if conditioned on values of earlier variables. The order of the variables was chosen considering potential interaction (iBreakdown-Plots) as implemented in the R package *iBreakDown*.[32]

Partial dependency plots[33] are aggregations of Ceteris Paribus profiles which are generated for each sample by varying the value of only one feature, while keeping the remaining features fixed. The partial dependecy plots were generated using the R package *DALEX*.[34]

Statistical analysis

Performance of risk scores on the test set was evaluated using receiver operating characteristic curves, and area under the curve (AUC, also called *C*-statistic) with 95% confidence intervals (CIs). AUC comparisons were conducted using bootstrap tests.

Independent predictive information in TRIM scores was tested via multivariable logistic regression models fit to the test data using the TRIM score, STS, and age as predictors. To test for incremental information model, improvement due to the inclusion of the TRIM scores was tested via likelihood-ratio tests and the net reclassification improvement was assessed.

Calibration curves were generated comparing the estimated and the observed proportion of events in the deciles of the risk scores. The logits of the predicted scores were regressed onto the observed events via logistic regression and intercept and slope were calculated with corresponding 95% confidence intervals.

A newly implemented interactive app (see Supplementary material online, *Figure S3*) allows the input of the variables necessary for the aTRIMpost risk score.

The significance level was set to 5% for all statistical tests. All analyses were performed using R (version 3.6.1).

Results

Study population

The analysed cohort comprised 28 147 completed TAVI interventions with annotated 30-day mortality recorded in GARY between 2011 and 2017. Of these cases, 875 (3.11%) patients died within 30 days post-TAVI. Further, 6693 interventions (205 (3.06%) with event for 30-day mortality) recorded in SwissTAVI were used for external validation (*Figure 1*).

The mean age of patients was 81 ± 6.1 years, and 13 185 (46.8%) were men. The mean STS score was 5.9 ± 4.6. (Additional baseline population characteristics: Supplementary material online, *Tables S1–S3*).

Missing values

Of the 155 variables used in TRIMpost 104 do not have a missing entry. Further 34 have less than 5% missing values, further 6 have less than 20% missing values, and only 9 have more than 30% missing values. The character of the missing values cannot be assumed to be MCAR. When testing for independence between the missing indicator and the outcome, 13 of the 51 variables with missing values show significant dependencies. This is also supported by the Jamshidian and Jalal's nonparametric MCAR test on the 25 numeric variables with missing values (Anderson–Darling test, $P < 0.001$). Therefore, a more involved imputation by multiple imputation by chained equations was employed and the model predictions averaged. Neither the classification performance nor the variable importance was affected much, so that we went with the simpler imputation for the remainder of the manuscript (see Supplementary material online, *Figure S4*).

Feature selection, performance, and comparison to previous models

We established RF models that condensed several patient features that are available prior to (TRIMpre) or (in addition) immediately after TAVI (TRIMpost) into risk scores. While risk prediction prior to TAVI (as in TRIMpre) is useful to guide treatment decision between TAVI and open surgery, risk prediction immediately after TAVI (as in TRIMpost) is useful to guide postprocedural decisions on intensification or de-escalation of monitoring, therapy, discharge, and follow-up frequency.

TRIMpre takes advantage of 97 features (see Supplementary material online, *Table S4*), achieving an AUC of 0.74 (95%-CI [0.70; 0.78]) on the GARY test set which was significantly superior to any of the examined traditional scores ($P = 0.019$ in comparison to STS score calculated on the same data, see Supplementary material online, *Figure S5*). Using the cutoff corresponding to the Youden index sensitivity and specificity are 0.587 and 0.781 compared to 0.514 and 0.792 for STS (*Table 1*). An abridged version of TRIMpre (aTRIMpre), which is easily manageable manually, exhibited not only overlapping features, but also equal performance to that of established scores, suggesting that the superior performance of TRIMpre depends on several informative features. Likewise, for TRIMpost, RF performance continued to improve with more features and reached its maximum with the entire 155 feature-set (see Supplementary material online, *Figure S6* and *Tables S4 and S5*). A steep increase in prediction performance was observed until 25 features. Therefore, the setting with 25 features was chosen for the aTRIMpost score. Both TRIMpost (AUC = 0.78; 95%-CI [0.74; 0.82]) and aTRIMpost (AUC = 0.74; 95%-CI [0.70; 0.78]) performed significantly better than STS ($P < 0.001$ and $P = 0.034$, respectively) and any of the established scores on the GARY test set (*Figure 2A–D*). Sensitivity and Specificity at the cutoff corresponding to the Youden index amount to 0.726 and 0.725 (TRIMpost) and 0.562 and 0.788 (aTRIMpost, *Table 1*). Despite some variations, those scores performed comparably on our data (AUCs between 0.678 and 0.698). Since only 28 out of 40 features (TVT(NYHA)) and 27 out of 42 (TVT(GaitSpeed)) are part of GARY, lower performance values for TVT(NYHA) (AUC = 0.623; 95%; CI [0.569–0.676]) and TVT(GaitSpeed) (AUC = 0.595; 95%-CI [0.542–0.647]) were observed. True tables for the TRIM scores and STS binarised at the Youden index can be found in Supplementary material online, *Table S6*.

SwissTAVI also differs markedly from GARY: We were able to extract or infer 78 (49.7%) of the 157 variables used by TRIMpost in SwissTAVI. Similarly, for aTRIMpost, TRIMpre, and aTRIMpre, only 63.0%, 49.5%, and 66.7% of the variables could be inferred in SwissTAVI (see Supplementary material online, Tables *S4* and *S5*). Despite the many missing variables aTRIMpre (AUC = 0.66; 95%-CI [0.62; 0.70]) and aTRIMpost (AUC = 0.67; 95%-CI [0.63; 0.71]) performed comparably to STS (AUC = 0.67; 95%-CI [0.63; 0.70]), TRIMpre (AUC = 0.69; 95%-CI [0.65; 0.72]) performed numerically better than STS and TRIMpost (AUC = 0.75; 95%-CI [0.72; 0.79]) significantly outperformed STS ($P < 0.001$; *Figure 2E* and see Supplementary material online, *Figure S7*).

We observed overestimation of the event risks (*Figure 2F–G*, see Supplementary material online, *Figure S8*) in the TRIM and, to a lower extent, STS score. The selection of patients with events with increased probability during training, which we implemented to address the class imbalance, leads to an overestimation of the prevalence of events. When we omitted the up-weighting of the minority and counteracted the expected time-effect by disregarding the first 100 TAVI

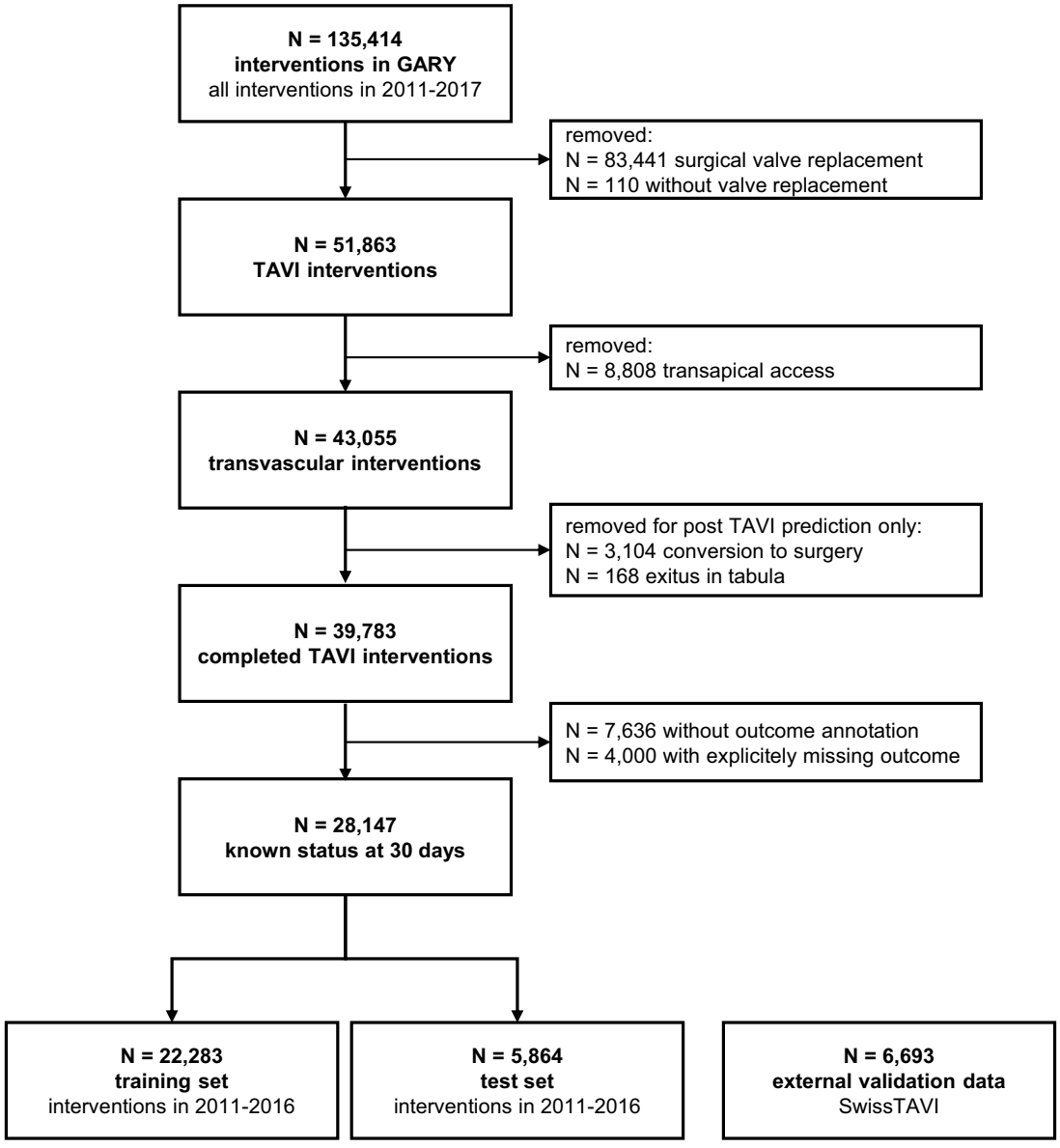


Figure 1 Description of the patient cohort for training, test, and external validation. Starting with all interventions in GARY from 2011 to 2017, the dataset was split into training (2011–2016) and test sets (2017). The TRIM scores were developed on the training set, the performance was assessed on the test set. Data from SwissTAVI was available for external validation.

interventions per hospital, a more favourable calibration was achieved (see Supplementary material online, *Figure S9*) at the cost of numerically lower classification performance (AUC = 0.78; 95%-CI [0.74; 0.82]). Thus, we stay with the proposed TRIM scores.

When factored into multi-variable logistic regression models including age and STS, the TRIM scores proved to be independent predictors of 30-day mortality on the GARY test set (TRIMpre: OR = 1.13; 95%-CI = [1.10; 1.16]; $P < 0.001$; TRIMpost: OR = 1.17; 95%-CI = [1.14; 1.20]; $P < 0.001$) and provided incremental prognostic significance (TRIMpre: change in model $\chi^2 = 67$; $P < 0.001$; net reclassification improvement (NRI) = 0.54; $P < 0.001$; TRIMpost: change in model $\chi^2 = 138$; $P < 0.001$; NRI = 0.70; $P < 0.001$), which showed that the ML algorithm did not simply relearn the STS score (*Figure 2H and I*).

Variable importance and partial dependency profiles

In TRIMpost, the duration of the intervention and fluoroscopy emerged as the most influential variables followed by serum creatinine level (*Figure 3A*). Among the 20 most important variables were also basic characteristics including weight, height, and age and expected echo parameters (peak AV gradient [Pmax], mean AV gradient [Pmean], LVEF, and pulmonary artery pressure [PAP]). Other contributing parameters were TAVI related, including aortic annular diameter, minimal iliac artery diameter, and heights of the right and left coronary arteries. Variable importance for the TRIMpre is provided in the Supplementary material online, *Figure S10*.

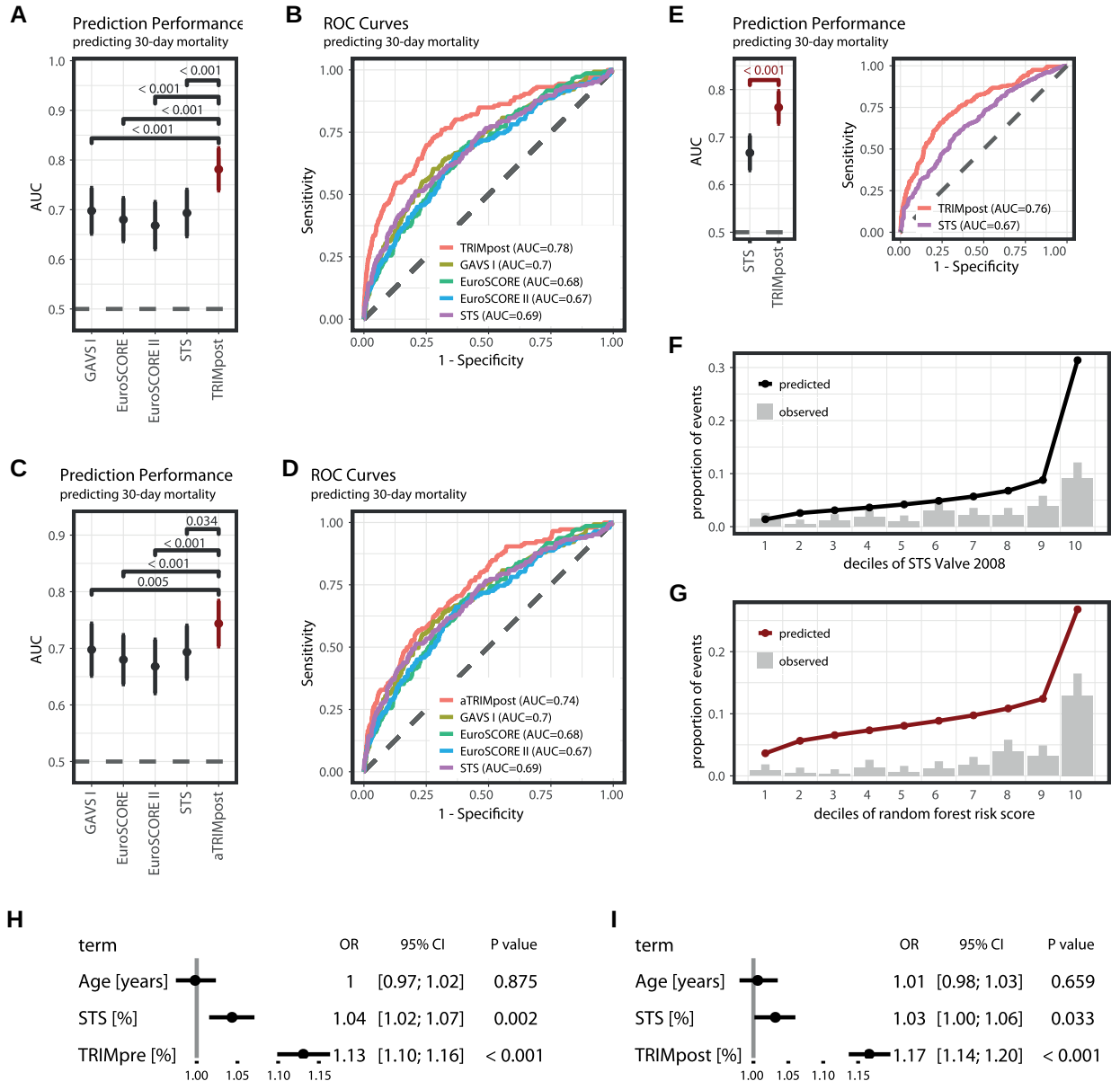


Figure 2 Performance and characteristics of TRIM scores. The ability to correctly predict 30-day mortality directly post-TAVI was assessed via ROC analyses. (*B*) Compares TRIMpost and several established scores. The dashed line on the diagonal shows the performance of an uninformative prediction. (*A*) Shows the area under the ROC curves as a measure of predictive performance with 95%-CI and pairwise comparisons against the performance of existing scores. (*D*) and (*C*) depict analogous results for aTRIMpost. (*E*) depicts analogous results for the external validation data (SwissTAVI). (*F*)-(G) Calibration Plots: The proportion of events as observed (bars) with 95%-CI and predicted (lines) for STS (*F*) and TRIMpost score (*G*) when evaluated for the test data ($n = 5889$). The derived TRIM Scores offer predictive information exceeding the STS alone, as they are strongly predictive for short term survival independent of the STS: Panels H (TRIMpre) and I (TRIMpost) show model coefficients (as odds-ratios) with 95%-CIs and associated *P* values against the null hypothesis of no association from multi-variable logistic regression models for 30-day mortality on the GARY test cohort ($n = 5889$). CI, confidence interval; OR, odds ratio.

Breakdown plots can be displayed for any individual and are part of the online tool for the aTRIMpost score for enhanced interpretability and transparency. For example, we display the contributions of individual variables to the outcome of the most extreme test set patient (the individual with the highest predicted TRIM score) in *Figure 3C*. Even in this very obvious high-risk case, no single variable was clearly dominant.

Although variables integrating intraprocedural complications and possibly operator skills (e.g. fluoroscopy duration and contrast use) expectedly reflected increased risk, most profiles of established variables

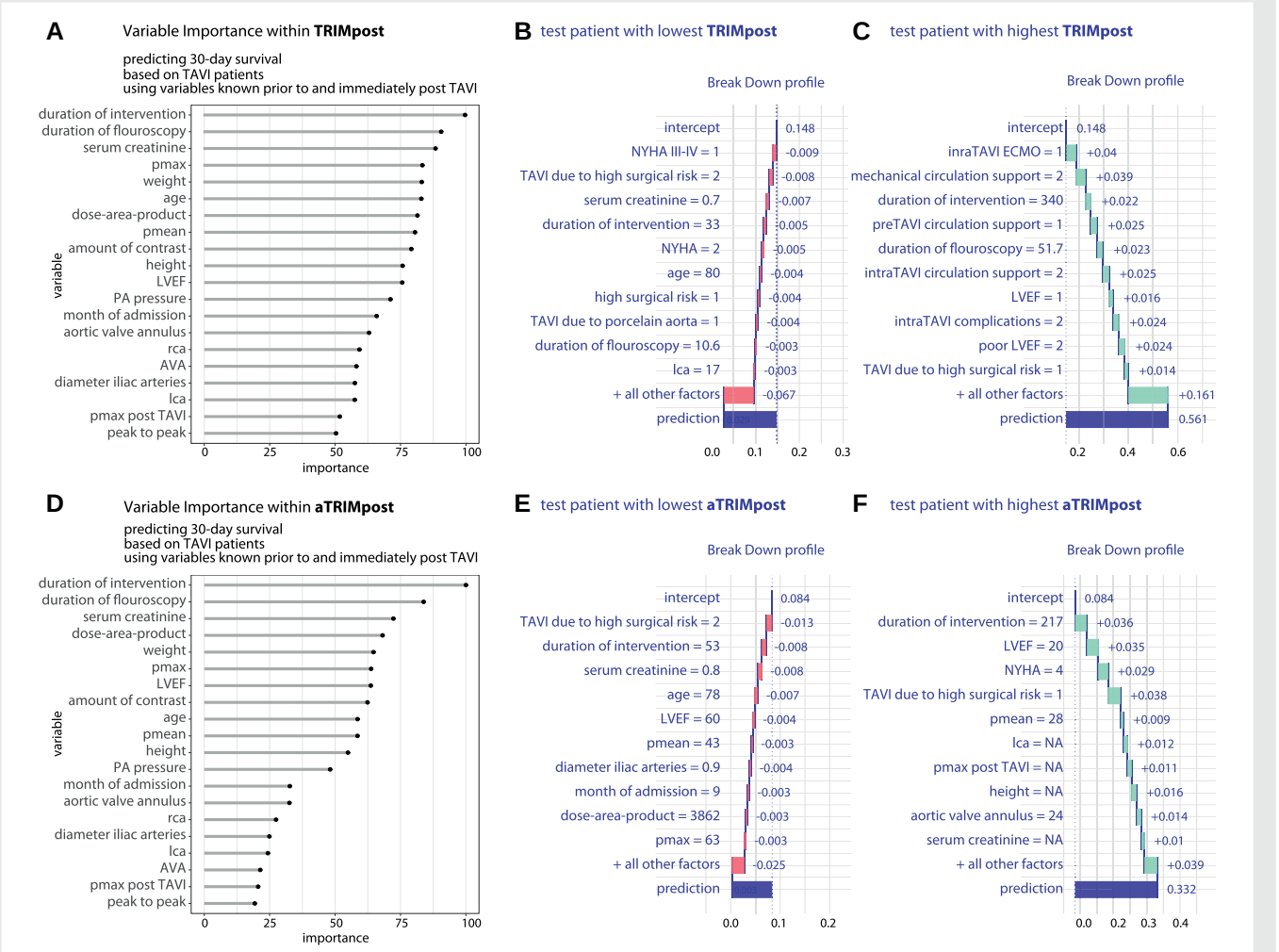


Figure 3 Variable importance. (*A*) Shows the 20 most important variables in TRIMpost score to predict 30-day mortality post-TAVI ordered by their contribution (importance) in discriminating the survivors from non-survivors. These values quantify how prominently the variables are used in the trained TRIMpost model. (*B*) and (*C*) exemplify interrogation of the trained model for the variable contribution in the prediction for two (most extreme) specific patients from the test set: Data of the patient with the highest (*B*) and the lowest (*C*) predicted survival probability are shown. These plots are meant to be interpreted from top to bottom: each row shows the change in the predicted TRIMpost score induced by the indicated variable. The variables are ordered by the size of this induced change in prediction and the top 10 influential variables are shown. Note that these contribution values are specific to the cases shown and since they are ordered by size these will induce an order that is typically different from the order of the variable importance values in (*A*). (*E*), (*F*), and (*G*) depict analogous results for aTRIMpost, i.e. where the most important 25 features are used in the training of the ML risk score. These plots are displayed individually for each patient in the score calculator app (see Supplementary material online, *Figure S3*).

were non-linear and with complex patterns of influence which are not easily graspable by a clinician or detected or modelled conventionally, thus supporting computational decision-support (*Figure 4*).

Potential application: early discharge recommendation

When accepting a risk of 1% for mortality within 30 days using a TRIMpost score cut-off <10.5, 72.2% of the cases in our test cohort would be candidates for early discharge. According to STS score, only 1812 (31%) were candidates for early discharge at the same risk level (*Table 1*). As the retrospective data does not tell how many of the surviving patients did survive only due to life-saving in-hospital treatment, we analysed the fractions of patients with complications as a proxy. Mediastinitis, any cerebral event, rupture of a vessel, dissection, bleeding, hematoma, and ischemia were regarded as serious complication. In the GARY test data, 506 (8.6%) of the 5864 patients suffered a serious complication. We found that among the 4236 patients predicted as dischargeable by TRIMpost at the 1% risk level were 309 (7.3%) with a serious complication. This fraction is significantly lower than in the patients not predicted as dischargeable (Fisher's exact test, $P < 0.001$). Among the 1812 patients marked as dischargeable by STS at the same 1% risk level there were 141 (7.8%) with a serious complication which is not significantly lower than in the patients not marked as dischargeable (Fisher's exact test, $P = 0.13$).

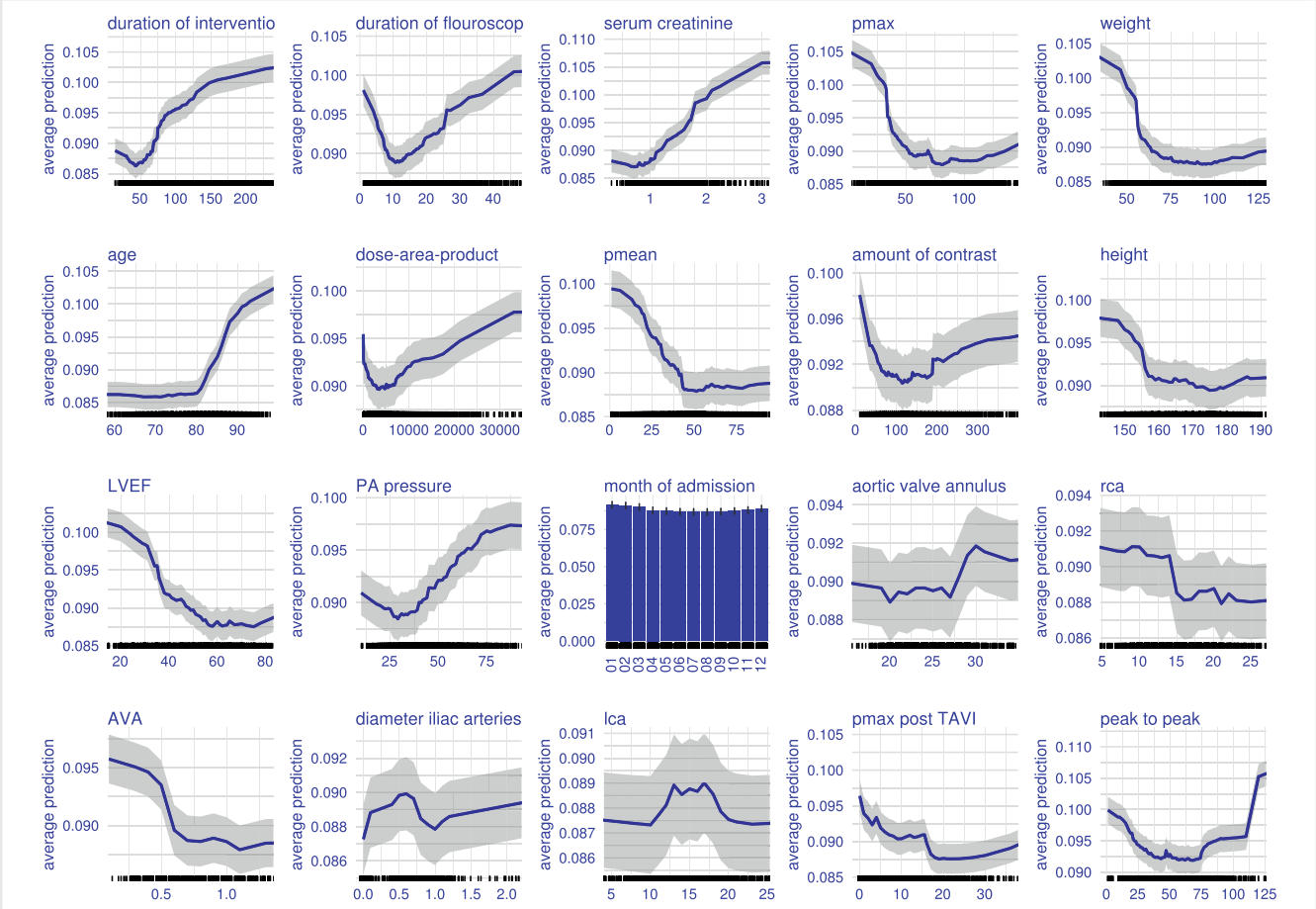


Figure 4 Profiles of influence. For each of the variables in *Figure 3* (*A*) the partial dependency profile is displayed. Each shows the influence of the variable on the predicted TRIMpost as an average across the full dataset. In individual patients, the value of variable of interest is changed across the range of the observed values while all other variables are kept fixed. The plots show the resulting prediction profiles as mean with 95%-CI across patients. CI, confidence interval.

Discussion

TRIM scores improve prediction of short-term mortality

Outdated risk models and a lack of clearly superior alternatives raise a need for updated and clinically meaningful risk scores in TAVI.[35] Using ML we, thus, developed and validated a risk score from one of the world's largest TAVI registries based on the prediction of 30-day mortality. Our ML-based TRIMpre score outperformed widely established risk scores and showed better performance on our test data than recently published ML models trained on less data reported from their respective test data sets[5,8] and TRIMpost is an entirely new tool for objective postinterventional risk estimation after TAVI. The superior accuracy of the TRIMpost models argues for post-procedural adjustment of preoperative risk assessment to better accommodate risk after TAVI has been performed.

Integration of ML-based decision-making has been anticipated to occur across all phases of patient care and yield valuable predictors for post-procedural care.[11] In contrast to the popularity of models assessing risk before TAVI including guideline recommendations, objective awareness of postoperative risk is much less developed; although important postprocedural decisions (on intensification or de-escalation of monitoring, therapy, discharge, and follow-up frequency) require consideration of the short-term mortality risk. Indeed, our data suggest that the preoperative risk does not simply propagate, but depends considerably on several intraprocedural events, justifying the use of a specific post-TAVI score. Moreover, TRIMpost can be useful for research applications such as covariate adjustment in randomised controlled trials considering post-TAVI care. Finally, examining the difference between TRIMpre and TRIMpost risk is suitable for measuring or benchmarking the quality of the interventional procedure, and the difference between TRIMpost and observed mortality is suitable for examining the quality of post-interventional care.

Potential clinical implications: earlier discharge

The patients and health care system would benefit from a TAVI-based risk score (together with clinical judgement) if patients can be safely discharged early after the intervention. While we do not suggest to solely rely on a score for obvious limitations, there are also limitations of clinical judgement in clinical practice that can be meaningfully addressed with a score. For example, a same day discharge decision based on clinical judgement alone might be re-evaluated in light of a high TRIMpost, or a low TRIM score might provide objective re-assurance. In this context, accepting a risk of 1% for mortality within 30 days TRIMpost

would identify more than twice as many candidates for early discharge than the STS. We compared the task-specific 'post' score (TRIMpost) to the STS score mainly to highlight the added value of a formalised post-interventional adjustment of pre-interventional risk estimation. Reducing the accepted mortality risk even lower to nearly 0%, TRIMpost selects 5 times as many patients as the STS score (*Table 2*).

In a prospective setting, the number of candidates for early discharge will likely be lower, as 7.3% of the candidates suffered from a serious complication and might have been survivors mainly due to life-saving in-hospital treatment. But interestingly, TRIMpost seems to pick up on serious complications, too, as there were significantly less patients with serious complications among the candidates for early discharge than among the remaining patients ($P < 0.001$). This was not the case for STS-derived candidates for early discharge ($P = 0.13$).

Of course, TRIMpost should only be applied in combination with additional parameters to decide on the safety of early discharge, warranting future updates and prospective validation.

Digitalization will enable the use of highly complex scores

Although physicians are capable of risk estimation, ML-based methods are less subjective, model non-linear relationships and complex interdependencies and significantly benefit from feature rich data. With developments in digital hospital information infrastructure, automated rather than manual data entry into scoring models will become routine, removing any limit on the number of features entered, and certainly the full TRIMpost score (considering 155 variables) requires automated data entry from a hospital information system. Hence, we anticipate that feature-rich models will soon be automatically calculated routinely to update physicians about risks along the entire stay of the patient. ML-based methods are especially strong in such feature rich situations where they can pick up on complex interdependencies between variables. In situations with limited numbers of variables (such as the aTRIM scores) this advantage is lost to some extent.

In case manual data entry or a more transparent calculation is desired, we have however provided abridged versions (aTRIM). The aTRIMpost score was trained on 25 data driven selected features and implemented in an interactive web app that is easily usable in clinical practice (for example, the STS-AVR requires 34 features to be entered, which we consider to be the upper limit for a manually operated scoring system). Interpretability is a crucial component in the translation of an ML method to clinical practice.[36] For each individual, our web app graphically displays the risk context and interpretation support rather than only the calculated score (see Supplementary material online, *Figure S3*).

Table 1 Performance measures on binary predictions. Predictions on the GARY test data were binarized according to the Youden index into 'highrisk' and 'lowrisk' predictions

Model	Sensitivity	Specificity	PPV	NPV
STS	0.514	0.792	0.059	0.985
TRIMpost	0.726	0.725	0.063	0.990
aTRIMpost	0.562	0.788	0.063	0.986
TRIMpre	0.587	0.781	0.074	0.984
aTRIMpre	0.593	0.733	0.062	0.984

For each of the TRIM scores as well as for STS this table shows the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of the resulting classification

Meaningful performance evaluation

As with classical statistical risk scoring, careful data pre-processing, selection of potential predictors, and handling of missing values are crucial components of successful ML training. We deliberately excluded patients with a conversion from TAVI to SAVR from the post-TAVI prediction models as predicting high risks for such patients is trivial and does hardly require decision support tools. Additionally, we limited the variables to those that are known directly post-TAVI for decision support that is applicable in truly fast discharge settings. In contrast, Hernandez-Suarez et al. included severe in-hospital complications in their risk score, which adds several patients to the dataset that are easily classified as high-risk even without ML.[4] The addition of such cases tremendously improves overall classification performance, but without increasing accuracy in a modern scenario, where the possibility of discharge is most often decided on the day of the procedure or the first post-procedural day. Moreover, we validated TRIM not against a subset of all data, but only against the most recent year available, whereas training was performed across a set of earlier years. Again, this reduced the apparent overall classification performance, but addresses more

Table 2 Number and percentage of candidates for early discharge in the test data depending on chosen risk model. Of all patients with a calculated risk score below the given threshold an estimated 1% (top part) or 0% (bottom part) died within 30 days. FN, number of false negatives

accepted mortality risk [%]	Risk score	N	Number of candidates for early discharge	[%]	FN	Threshold
1	GAVS I	5732	1301	22.7	13	2.78
	Euro I	5800	1669	28.8	16	9.77
	Euro II	5745	69	1.2	0	1.12
	STS	5849	1812	31	18	3.39
	TRIMpost	5864	4236	72.2	42	0.10
0	GAVS I	5732	21	0.4	0	0.58
	Euro I	5800	66	1.1	0	3.65
	Euro II	5745	69	1.2	0	1.12
	STS	5849	26	0.4	0	1.15
	TRIMpost	5864	130	2.2	0	0.04

adequately that TAVI is a moving field. Despite this stringent (and more meaningful) approach, we found *C*-statistics for TRIMpost and aTRIMpost of 0.78 and 0.74, respectively. This level of performance matches many guideline recommended clinical scores (including the Heart Failure Survival Score,[37] the Seattle Heart Failure Model,[38] the EuroSCORE II,[21] and the MIDA mortality risk score.[39])

Interpretable ML allows valuable insights

In addition to predictive analytics, our analysis yielded valuable insights into the determinants of postprocedural risk. Most variables are either known as strong predictors or intuitively plausible. High aorto-ventricular pressure gradients correlate inversely with the risk and may recapitulate LV function and the potential procedural gain in cardiac output and left ventricular unloading. In contrast to previously established scores, some variables were less expectedly informative in aTRIMpost: LCA and RCA heights may reflect the risk of prosthesis related coronary occlusion, but (due to non-linear relation and a rare event rate) could also reflect additional more occult aspects of aortic root anatomy. The informative variable 'month of TAVI', peaking from January to March, could simply indicate that influenza and pneumococcus vaccination deserve more attention to reduce mortality in this elderly cohort.

External validation and robustness of the TRIM scores

We could validate the TRIM scores externally on data of the SwissTAVI. The differences in the recorded data were remarkably large between SwissTAVI and GARY. Despite our best efforts to extract or reconstruct the variables used in the TRIM scores, we were able to match 50–66% (depending on the TRIM score) of the variables only. The large fraction of missing values exists although the variables collected in GARY can be considered quite standard and routinely collected variables and is due to different foci given in the two national registries. Given that up to 50% of the training variables were missing from the SwissTAVI validation data, the TRIM scores proved highly robust: all TRIM scores performed comparably to the STS while TRIMpost still significantly outperformed the STS also in this external validation.

Limitations

This study has some limitations. Despite consideration of a wide range of features, to date all models rely on patient record data and are therefore incomplete. However, TRIM may advance objectivity in the process of risk estimation and provide valuable guidance: According to a comment in a recent systematic review, 'even models with poor *C*-statistics (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'.[35] The same authors attribute models with *C*-statistics levels >0.8 (which is very close to that of TRIMpost) with 'strong support to guide medical decision-making and can reliably indicate whether a patient will experience an event'.

We had to exclude cases due to missing endpoint annotations. We did not observe an obvious bias in the excluded cohort. Patients in the lower NYHA classes exhibited the endpoint slightly more often, but interestingly the distribution of the risk groups did not differ much between patients with and without endpoint annotation (see Supplementary material online, *Table S7*).

There is a time gap between the analyzed cohort and future cases to be predicted. This is similar to other risk scores developed on registry data, e.g. the MIDA score published in 2018 was adopted by the 2021 ESC guidelines for valvular heart disease[1] and was derived on patient data from 1980 to 2005.[39] However, the TRIM score is more up to date than the more established scores and we anticipate a uniform overestimation—as visible in the calibration—of patient risk as the main consequence. This can be mitigated by simple rescaling of the resulting risk scores, and we suggest updating TRIM every 5 years to take advantage of the continuously growing GARY cohort similar to updates in other scores, e.g. FRANCE-2 and ACC-TAVI.[40] Although TAVI has become a very standardised procedure, ideally, additional data from different countries or regions of the world and imaging data should be included in the future to further improve individualized decision support.

Conclusions

In conclusion, the mortality risk before and after TAVI was calculated and condensed in newly developed TRIM scores based on the large GARY patient cohort and ML. TRIM scores advance preprocedural and postprocedural risk assessment, integrating relevant information from the course of TAVI treatment with strong predictive performance.

Supplementary material

Supplementary material is available at *European Heart Journal – Digital Health*.

Funding

This analysis was supported by the German Center for Cardiovasclar Research (DZHK). The responsible body of the GARY registry is a non-profit organization named Deutsches Aortenklappenregister gGmbH founded by the Deutsche Gesellschaft für Kardiologie (DGK) and Deutsche Gesellschaft für Thorax- Herz- und Gefäßchirurgie (DGTHG). The responsible societies and the BQS Institute are by the virtue of their constitutions independent organizations in both—legal and scientific—views. The registry receives financial support in the format of unrestricted grants by medical device companies (Edwards Lifesciences, JenaValve Technology, Medtronic, Sorin, St Jude Medical, Symetis SA), which however have neither access to data nor influence on their publication.

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

Conflict of interest: T Friede reports personal fees for statistical consultancies (includingdata monitoring committees) from Novartis, Bayer, Janssen, Roche, Galapagos, Penumbra, Parexel, Vifor, BiosenseWebster, CSLBehring, Fresenius Kabi, Coherex Medical, LivaNova, Minoryx, IQVIA, Enanta, Relaxera, Immunic; all outside the submitted work. CHamm reports membership in the International Strategic Advisory Board of Medtronic Inc. H Möllmann reports Speaker honoraria, advisory board and proctor fees Abbott, Boston Scientific, Edwards Lifesciences, Medtronic. S Stortecky reports research grants to theinstitution from Edwards Lifesciences, Medtronic, Abbott and Boston Scientific, and personal fees from Boston Scientific, Teleflex, and BTG. S Windecker reports research, travel or educational grants to the institution without personal remuneration from Abbott, Abiomed, Amgen, Astra Zeneca, Bayer, Braun, Biotronik, Boehringer Ingelheim, Boston Scientific, Bristol Myers Squibb, Cardinal Health,CardioValve, Cordis Medical, Corflow Therapeutics, CSL Behring, Daiichi Sankyo, Edwards Lifesciences, Farapulse Inc. Fumedica, Guerbet, Idorsia, Inari Medical, InfraRedx, Janssen-Cilag, Johnson & Johnson, Medalliance, Medicure, Medtronic,

Merck Sharp &Dohm, Miracor Medical, Novartis, Novo Nordisk, Organon, OrPha Suisse, Pharming Tech. Pfizer, Polares, Regeneron, Sanofi-Aventis, Servier, Sinomed, Terumo, Vifor, V-Wave. He served as advisory board member and/or member of the steering/executive group oftrials funded by Abbott, Abiomed, Amgen, Astra Zeneca, Bayer, Boston Scientific, Biotronik, Bristol Myers Squibb, Edwards Lifesciences,MedAlliance, Medtronic, Novartis, Polares, Recardio, Sinomed, Terumo, and V-Wave with payments to the institution but no personalpayments. He is also member of the steering/executive committee group of several investigator-initiated trials. G Hasenfuß reportshonorarium for consultating and/or for lectures from AstraZeneca, Bayer, Boehringer, Impulse Dynamics, Novartis, Pfizer, Servier, Viforand Consultation without honoraria from Corvia. T Seidler reports research or educational grants to the institution by Edwards LifeSciences and honoraria for lectures or advisory board consultations or travel grants from Abbott Vascular, AstraZeneca, BoehringerIngelheim, Bristol Myers Squibb, Corvia, Cytokinetics, Edwards Life Sciences, Medtronic, Myocardia, Novartis, Pfizer, Teleflex. S. Bleiziffer reports Speaker fees from Boston Scentific, Abbott, Edwards Lifesciences and Medtronic. All other authors declare no conflicts of interest exist.

Data availability

The data for this analysis was provided by the GARY and SwissTAVI consortia. The data is closed source, but individual requests may be addressed to the respective registries and will be handled according to their guidelines.

References

1. Vahanian A, Beyersdorf F, Praz F, Milojevic M, Baldus S, Bauersachs J, *et al.* 2021 ESC/EACTS guidelines for the management of valvular heart disease. *Eur Heart J*. 2021;**43**:ehab395.
2. Popma JJ, Deeb GM, Yakubov SJ, Mumtaz M, Gada H, O'Hair D, *et al.* Transcatheter aortic-valve replacement with a self-expanding valve in low-risk patients. *N Engl J Med* 2019;**380**:1706–1715.
3. Mack MJ, Leon MB, Thourani VH, Makkar R, Kodali SK, Russo M, *et al.* Transcatheter aortic-valve replacement with a balloon-expandable valve in low-risk patients. *N Engl J Med* 2019;**380**:1695–1705.
4. Hernandez-Suarez DF, Kim Y, Villablanca P, Gupta T, Wiley J, Nieves-Rodriguez BG, *et al.* Machine learning prediction models for in-hospital mortality after transcatheter aortic valve replacement. *JACC Cardiovasc Interv* 2019;**12**:1328–1338.
5. Mamprin M, Lopes RR, Zelis JM, Tonino PAL, van Mourik MS, Vis MM, *et al.* Machine learning for predicting mortality in transcatheter aortic valve implantation: an intercenter cross validation study. *J Cardiovasc Dev Dis* 2021;**8**:65.
6. Penso M, Pepi M, Fusini L, Muratori M, Cefalù C, Mantegazza V, *et al.* Predicting long-term mortality in TAVI patients using machine learning techniques. *J Cardiovasc Dev Dis* 2021;**8**:44.
7. Agasthi P, Ashraf H, Pujari SH, Girardo ME, Tseng A, Mookadam F, *et al.* Artificial intelligence trumps TAVI2-SCORE and CoreValve score in predicting 1-year mortality post-transcatheter aortic valve replacement. *Cardiovasc Revasc Med* 2021;**24**:33–41.
8. Lopes RR, van Mourik MS, Schaft EV, Ramos LA, Baan J, Vendrik J, *et al.* Value of machine learning in predicting TAVI outcomes. *Neth Heart J* 2019;**27**:443–450.
9. Abdul Ghffar Y, Osman M, Shrestha S, Shaukat F, Kagiyama N, Alkhouli M, *et al.* Usefulness of semisupervised machine-learning-based phenogrouping to improve risk assessment for patients undergoing transcatheter aortic valve implantation. *Am J Cardiol* 2020;**136**:122–130.
10. Debonnaire P, Fusini L, Wolterbeek R, Kamperidis V, van Rosendael P, van der Kley F, *et al.* Value of the "TAVI2-SCORe" versus surgical risk scores for prediction of one year mortality in 511 patients who underwent transcatheter aortic valve implantation. *Am J Cardiol* 2015;**115**:234–242.
11. Sardar P, Abbott JD, Kundu A, Aronow HD, Granada JF, Giri J. Impact of artificial intelligence on interventional cardiology. *JACC Cardiovasc Interv* 2019;**12**:1293–1303.
12. Commandeur F, Slomka PJ, Goeller M, Chen X, Cadet S, Razipour A, *et al.* Machine learning to predict the long-term risk of myocardial infarction and cardiac death based on clinical risk, coronary calcium, and epicardial adipose tissue: a prospective study. *Cardiovasc Res* 2019;**116**:2216–2225.
13. Pieszko K, Hiczkiewicz J, Budzianowski P, Rzeźniczak J, Budzianowski J, Blaszczyński J, *et al.* Machine-Learned models using hematological inflammation markers in the prediction of short-term acute coronary syndrome outcomes. *J Transl Med* 2018;**16**:334.
14. Nanayakkara S, Fogarty S, Tremeer M, Ross K, Richards B, Bergmeir C, *et al.* Characterising risk of in-hospital mortality following cardiac arrest using machine learning: a retrospective international registry study. *PLoS Med* 2018;**15**:e1002709.
15. Leha A, Hellenkamp K, Unsöld B, Mushemi-Blake S, Shah AM, Hasenfuß G, *et al.* A machine learning approach for the prediction of pulmonary hypertension. *PLoS One* 2019;**14**:e0224453.
16. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019;**17**:195.
17. Hamm CW, Mollmann H, Holzhey D, Beckmann A, Veit C, Figulla HR, *et al.* for the GARY-Executive Board. The German aortic valve registry (GARY): in-hospital outcome. *Eur Heart J*. 2014;**35**:1588–1598.
18. Stortecky S, Franzone A, Heg D, Tueller D, Noble S, Pilgrim T, *et al.* Temporal trends in adoption and outcomes of transcatheter aortic valve implantation: a SwissTAVI registry analysis. *Eur Heart J Qual Care Clin Outcomes* 2019;**5**:242–251.
19. Collins GS, Reitsma JB, Altman DG, Moons KGM; TRIPOD Group. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. The TRIPOD group. *Circulation*. 2015;**131**:211–219.
20. Roques F, Michel P, Goldstone AR, Nashef SA. The logistic EuroSCORE. *Eur Heart J* 2003;**24**:882.
21. Nashef SAM, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, *et al.* EuroSCORE II. *Eur J Cardiothorac Surg* 2012;**41**:734–745.
22. Kotting J, Schiller W, Beckmann A, Schafer E, Dobler K, Hamm C, *et al.* German Aortic valve score: a new scoring system for prediction of mortality related to aortic valve procedures in adults. *Eur J Cardiothorac Surg* 2013;**43**:971–977.
23. O'Brien SM, Shahian DM, Filardo G, Ferraris VA, Haan CK, Rich JB, *et al.* The society of thoracic surgeons 2008 cardiac surgery risk models: part 2-isolated valve surgery. *Ann Thorac Surg* 2009;**88**:S23–S42.
24. O'Brien SM, Cohen DJ, Rumsfeld JS, Brennan JM, Shahian DM, Dai D, *et al.* Variation in hospital risk–adjusted mortality rates following transcatheter aortic valve replacement in the United States: A report from the society of thoracic surgeons/American college of cardiology transcatheter valve therapy registry. *Circ Cardiovasc Qual Outcomes* 2016;**9**:560–565.
25. Arnold SV, O'Brien SM, Vemulapalli S, Cohen DJ, Stebbins A, Brennan JM, *et al.* Inclusion of functional Status measures in the risk adjustment of 30-day mortality after transcatheter aortic valve replacement. *JACC Cardiovasc Interv* 2018;**11**:581–589.
26. Breiman L. Random forests. *Mach Learn* 2001;**45**:5–32.
27. O'Brien R, Ishwaran H. A random forests quantile classifier for class imbalanced data. *Pattern Recognit* 2019;**90**:232–249.
28. Chollet F, Allaire JJ. *Deep learning with R. Shelter island*. NY: Manning Publications Co; 2018.
29. Jamshidian M, Jalal S. Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. *Psychometrika* 2010;**75**:649–674.
30. Louppe G, Wehenkel L, Sutera A, Geurts P. Understanding variable importances in forests of randomized trees. In: Burges CJC, Bottou L, Welling M, Ghahramani Z, Weinberger KQ, eds. *Advances in neural information processing systems 26*. Red Hook, NY: Curran Associates, Inc.; 2013. p431–439.
31. Staniak M, Biecek P. Explanations of model predictions with live and breakDown packages. *R J* 2019;**10**:395.
32. Gosiewska A, Biecek P. Do Not Trust Additive Explanations. ArXiv190311420 Cs Stat [Internet]. 2020 May [cited 2021 Nov 26]; Available from: http://arxiv.org/abs/1903.11420
33. Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann Stat* 2001;**29**:1189–1232.
34. Biecek P. DALEX: explainers for Complex predictive models in R. *J Mach Learn Res* 2018;**19**:1–5.
35. Siddiqi TJ, Usman MS, Khan MS, Khan MAA, Riaz H, Khan SU, *et al.* Systematic review and meta-analysis of current risk models in predicting short-term mortality after transcatheter aortic valve replacement. *EuroIntervention* 2020;**15**:1497–1505.
36. Checcucci E, Autorino R, Cacciamani GE, Amparore D, De Cillis S, Piana A, *et al.* Uro-technology and SoMe working group of the young academic urologists working party of the European association of urology. Artificial intelligence and neural networks in urology: current clinical applications. *Minerva Urol E Nefrol Ital J Urol Nephrol* 2020;**72**:49–57.
37. Aaronson KD, Schwartz JS, Chen TM, Wong KL, Goin JE, Mancini DM. Development and prospective validation of a clinical Index to predict survival in ambulatory patients referred for cardiac transplant evaluation. *Circulation* 1997;**95**:2660–2667.
38. Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, *et al.* The Seattle heart failure model: prediction of survival in heart failure. *Circulation* 2006;**113**:1424–1433.
39. Grigioni F, Clavel MA, Vanoverschelde JL, Tribouilloy C, Pizarro R, Huebner M, *et al.* The MIDA mortality risk score: development and external validation of a prognostic model for early and late death in degenerative mitral regurgitation. *Eur Heart J* 2018;**39**:1281–1291.
40. Al-Farra H, de Mol B, Ravelli ACJ, ter Burg WJPP, Houterman S, Henriques JPS, *et al.* Update and, internal and temporal-validation of the FRANCE-2 and ACC-TAVI early-mortality prediction models for transcatheter aortic valve implantation (TAVI) using data from The Netherlands heart registration (NHR). *IJC Heart Vasc* 2021;**32**:100716.