

## Clinical Investigation

### Inadequacy of Existing Clinical Prediction Models for Predicting Mortality after Transcatheter Aortic Valve Implantation

Glen P. Martin, MSc <sup>1</sup>; Matthew Sperrin, PhD<sup>1</sup>; Peter F. Ludman, MA, MD, FRCP, FESC<sup>2</sup>;  
Mark A. de Belder, MA, MD, FRCP<sup>3</sup>; Chris P. Gale, PhD, FRCP, FESC<sup>4</sup>; William D. Toff,  
MD, FRCP, FESC<sup>5,6</sup>; Neil E. Moat, MBBS, MS<sup>7</sup>; Uday Trivedi, MBBS<sup>8</sup>; Iain Buchan, MD,  
FFPH<sup>1</sup>; Mamas A. Mamas, MA, DPhil, FRCP<sup>1,9</sup>

1: Health e-Research Centre, University of Manchester, Manchester, United Kingdom

2: Queen Elizabeth Hospital, Birmingham, United Kingdom

3: James Cook University Hospital, Middlesbrough, United Kingdom

4: MRC Bioinformatics Unit, Leeds Institute of Cardiovascular and Metabolic Medicine,  
University of Leeds

5: Department of Cardiovascular Sciences, University of Leicester, Clinical Sciences Wing,  
Glenfield General Hospital, Leicester, United Kingdom

6: NIHR Leicester Cardiovascular Biomedical Research Unit, Leicester, United Kingdom

7: Royal Brompton and Harefield National Health Service (NHS) Foundation Trust, London,  
United Kingdom

8: Sussex Cardiac Centre, Brighton and Sussex University Hospitals, Brighton,  
United Kingdom

9: Keele Cardiovascular Research Group, Keele University, Stoke-on-Trent, UK, Royal  
Stoke Hospital, University Hospitals North Midlands, Stoke-on-Trent, United Kingdom

**Abbreviated Title:** *CPMs predicting mortality after TAVI*

**Conflict of Interest:** No conflicts of interest to declare

**Corresponding Author:**

Mamas A. Mamas

Professor of Cardiology / Honorary Consultant Cardiologist

Keele Cardiovascular Research Group,

University of Keele

Stoke-on-Trent, United Kingdom

Email: mamasmamas1@yahoo.co.uk

## **Abstract and Keywords**

### **Background**

The performance of emerging Transcatheter Aortic Valve Implantation (TAVI) clinical prediction models (CPMs) in national TAVI cohorts distinct from those where they have been derived is unknown. This study aimed to investigate the performance of the German Aortic Valve, FRANCE-2, OBSERVANT and American College of Cardiology (ACC) TAVI CPMs compared with the performance of historic cardiac CPMs such as the EuroSCORE and STS-PROM, in a large national TAVI registry.

### **Methods**

The calibration and discrimination of each CPM were analysed in 6676 patients from the UK TAVI registry, as a whole cohort and across several subgroups. Strata included gender, diabetes status, access route and valve type. Furthermore, the amount of agreement in risk classification between each of the considered CPMs was analysed at an individual patient level.

### **Results**

The observed 30-day mortality rate was 5.4%. In the whole cohort, the majority of CPMs over-estimated the risk of 30-day mortality, although the mean ACC score (5.2%) approximately matched the observed mortality rate. The areas under ROC curve were between 0.57 for OBSERVANT and 0.64 for ACC. Risk classification agreement was low across all models, with Fleiss's kappa values between 0.17 and 0.50.

### **Conclusions**

Although the FRANCE-2 and ACC models outperformed all other CPMs, the performance of current TAVI-CPMs was low when applied to an independent cohort of TAVI patients.

Hence, TAVI specific CPMs need to be derived outside populations previously used for model derivation, either by adapting existing CPMs or developing new risk scores in large national registries.

**Keywords:**

Aortic Stenosis – Transcatheter Aortic Valve Implantation – Mortality Risk Prediction –  
Clinical Prediction Models

## **Introduction**

Despite surgical aortic valve replacement (SAVR) being the definitive treatment strategy for severe symptomatic Aortic Stenosis (AS), a significant proportion of patients are not offered surgery due to co-morbidities or frailty that contribute to high surgical risks and adverse outcomes in such patient groups (1). Transcatheter aortic valve implantation (TAVI) has emerged as an efficacious but less invasive treatment option in high and intermediate operative risk patients (2–5). As such, treatment allocation between medical management, SAVR and TAVI depends on multiple factors, but key is the assessment of the patient's procedural risk. Clinical prediction models (CPMs), which quantify the risks associated with the proposed treatment strategy at an individual patient level, can aid heart-teams in this clinical decision-making process and are vital for audit purposes between TAVI centres.

Cardiac surgery CPMs for short-term mortality prediction, such as the European System for Cardiac Operative Risk Evaluation Score (EuroSCORE) (6,7) and the Society of Thoracic Surgeons Predicted Risk of Mortality (STS) model (8), have been used to identify high-risk patients in randomised trials of TAVI (2,3). However, these surgical CPMs perform poorly in predicting risk after both SAVR and TAVI, as exemplified in the Placement of Aortic Transcatheter Valves (PARTNER) cohort A trial where there was large disagreement between the observed and STS-expected 30-day mortality (3). Moreover, several cohort studies have shown the inaccuracy of the surgical CPMs in predicting mortality after TAVI (9–11).

Consequently, TAVI specific CPMs are beginning to emerge from large cohorts of TAVI patients (12–15). In particular, the German Aortic Valve Score (German AV) was developed using patients who underwent either surgical replacement or TAVI (13), while TAVI-specific CPMs have been derived in the France TAVI registry (FRANCE-2 model) (14), the Italian TAVI registry (OBSERVANT model) (12) and the Society of Thoracic

Surgeons/ American College of Cardiology Transcatheter Valve Therapy registry (ACC model) (15). However, the performance of the aforementioned TAVI-CPMs in large cohorts of patients outside of their derivation cohorts is unknown. Hence, it is unclear if they can be reliably used in other national settings.

Therefore, the aim of this study was to investigate the performance and agreement of the German AV, FRANCE-2, OBSERVANT and ACC TAVI-CPMs for predicting 30-day mortality outside their development cohorts, to examine if the performance was sufficient for them to be used for this purpose. The study compared the TAVI-CPM performance against surgical CPMs, namely the Logistic EuroSCORE (LES), EuroSCORE II (ESII) and STS score.

## **Methods**

### **UK TAVI Registry**

Prospectively collected data on every TAVI procedure in the United Kingdom from January 2007 to December 2014 were obtained through the UK TAVI registry (16). By the end of 2014, 34 UK centres were performing TAVI procedures with multi-disciplinary teams of cardiologists, surgeons and other health-care professionals at each centre deciding on patients' suitability for TAVI (16). The web-based registry comprises 95 variables detailing patient demographics, risk factors for intervention, procedural details and adverse outcomes up to the time of hospital discharge. All-cause mortality tracking was obtained from the Office for National Statistics (ONS) providing the life-status of English and Welsh patients (two countries of the UK). Mortality tracking was unavailable for patients in Northern Ireland and Scotland and as such, these patients were removed from the analysis.

## Statistical Analysis

Multiple imputation was used for missing values, with ten datasets imputed (17). Missing life-status was not imputed and this analysis excluded any patient who had such a missing endpoint. To avoid underestimation of covariate-outcome associations, 30-day mortality indication was used in the imputation models for missing covariates (18). [Further details of the imputation procedure are given in the supplementary material.](#)

The risk of 30-day mortality implied by each CPM was retrospectively calculated for each patient based on the published regression coefficients (6–8,12–15). This analysis used clinical reasoning to make assumptions regarding translation between variable definitions in the published CPMs and those in the UK TAVI dataset. Any CPM risk-prediction variable that was not recorded in the UK TAVI registry was assumed risk factor absent for all patients. The full translation between each CPM and the TAVI registry variables is given in the Supplementary Material (**Supplementary Table 1 – Supplementary Table 7**) along with the statistical code used to calculate the scores.

The performance of each CPM was assessed in terms of calibration and discrimination. Calibration is the agreement between the expected and observed event rates across the full risk range; discrimination is the ability of the CPM to distinguish between those who will experience an event and those who will not. Discrimination of the risk models was analysed using the area under the receiver operating characteristic (ROC) curve, with values between 0.5 and 1 where higher values indicate better discrimination. To examine the calibration of each CPM, a logistic regression model was fitted with the event indicator as the outcome and the linear predictor from the CPM as the only covariate (19). Perfect calibration would occur when the corresponding intercept and slope are zero and one respectively, with the intercept estimated assuming a slope of unity. Furthermore, the Brier Score was used as a measure of overall performance, with values between 0 (perfect prediction) and 1 (worst

prediction) (20); ~~the McFadden's pseudo-R<sup>2</sup> was calculated to give an indication of the explained variation in the data.~~ CPM performance was analysed in the whole cohort and within several subgroups. The following subgroups were considered: age ( $\leq$  or  $>$  75), sex, diabetes status, access route (Transfemoral vs. non-Transfemoral), valve type (SAPIEN<sup>TM</sup> vs. CoreValve<sup>TM</sup>), previous coronary artery bypass graft (CABG) status, LV function (LVEF  $<$  50% or LVEF  $\geq$  50%) and procedure urgency (elective vs. non-elective).

Patient-level risk agreement between CPMs was analysed in the surgical models and the TAVI models separately to facilitate fair comparisons. It was decided, a priori, to derive  $\epsilon$  cut-off values for each CPM ~~were used to that~~ defined three risk levels (low-, medium- and high-risk), with approximately equal patient numbers ~~of patients~~ in each. The proportions of patients for whom risk classification agreed between multiple CPMs was then calculated.

Additionally, Fleiss's Kappa was calculated in the surgical and TAVI models (21). A sensitivity analysis was conducted in which the risk stratifications were re-defined to give a population ratio of 1:3:1 for low-, medium- and high-risk, respectively.

R version 3.3.1 (22) was used for all statistical analyses. Multiple imputation of the dataset was completed using the mice package (23), graphical plots were made using the ggplot2 package (24) and the package pROC was used for constructing ROC curves (25).

The Health e-Research Centre, funded by the Medical Research Council [MR/K006665/1] and the North Staffordshire Heart Committee supported this work. The authors are solely responsible for the design and conduct of this study, all study analyses, the drafting and editing of the paper and its final contents.

## Results

The UK TAVI registry included all 7431 patients who underwent a TAVI procedure between January 2007 and December 2014. All patients from Northern Ireland (n=400) and

the majority of Scottish patients (n=193) were excluded from the analysis due to absence of ONS mortality tracking. Out of the remaining 6838 patients, a further 162 were removed due to missing life status, leaving 6676 patients studied in this analysis. The observed survival rates were 94.6%, 83.3% and 64.4% at 30-days, one-year and three-year follow-up, respectively~~Median follow up time was 1.9 years (inter-quartile range 0.95–3.3 years).~~ **Table 1** presents summary statistics for baseline characteristics of the patients in the UK TAVI registry.

### Performance Analysis

From January 2007 to December 2014, there were 360 deaths within 30-days of the TAVI procedure (5.4%). The expected 30-day mortalities in the whole cohort were 21.9%, 8.1%, 5.1%, 7.4%, 9.2, 7.1% and 5.2% from the LES, ESII, STS, German AV, FRANCE-2, OBSERVANT and ACC CPMs, respectively (**Table 2**). The ACC score and STS model were the closest to the observed mortality in terms of absolute and relative differences, while the LES overestimated risk by a factor of four (**Table 2**). After a decrease from 2007 to 2008, the observed 30-day mortality per year remained approximately constant, with further decreases in 2013 and 2014 (**Figure 1**). In contrast, the profile of the majority of CPMs remained approximately constant throughout (**Figure 1**). The inflated observed 30-day mortality in the first two years likely reflects the UK learning curve and advances in TAVI technology, while the CPMs do not account for such factors. The observed and expected 30-day mortality rates over each subgroup are given in **Supplementary Table 8**.

**Table 3** shows the performance of each CPM in the whole cohort. While the calibration intercepts of the ACC and STS models were significantly close to zero (i.e. the observed and expected mortalities agreed), the 95% confidence intervals for the calibration slopes did not span one, indicating model miscalibration. Poor discrimination was observed, with area under the ROC curves between 0.57 and 0.64 for the whole cohort; the FRANCE-2

TAVI score and the ACC TAVI score had the highest AUC values of 0.62 and 0.64, respectively. Overall performance, as measure by the Brier score, was similar for the majority of models with values of 0.05; a Brier score of 0.09 for the LES was the highest (worst) amongst the models. Quantitatively similar results were obtained from a sensitivity analysis that excluded patients who underwent TAVI in 2007 or 2008 (n=337) where the observed mortality was elevated over that in subsequent years (Supplementary Table 9).

The performances of all the CPMs in each subgroup are given in the supplementary material (**Supplementary Table 10**). The expected mortality from the ACC TAVI model was significantly close to the observed mortality across all strata, but satisfactory calibration (calibration intercept and slope close to zero and one, respectively) was only observed for this CPM in female and diabetic subgroups. All other models were miscalibrated across strata. The area under the ROC curve was below 0.7 for all CPMs across the subgroups, with the majority close to 0.6; the ACC and FRANCE-2 CPMs had the highest discrimination across subgroups.

### Agreement Analysis

~~Cut-off values were selected for all CPMs that gave approximately equal numbers of patients in low-, medium- and high-risk categories.~~ The chosen cut-off values that gave approximately equal numbers of patients in low-, medium- and high-risk categories- are given in **Table 4**. Based on these cut-off values, the proportions of patients ~~in-~~classified in each risk level who were similarly classified across the other CPMs were calculated (**Figure 2** for the surgical based CPMs and **Figure 3** for the TAVI based CPMs). A low level of agreement at an individual patient level was observed; for example, only 31.8% of the 1951 patients grouped as high-risk group by FRANCE-2 >10% were also grouped as high-risk by the OBSERVANT and ACC models (**Figure 3**). Quantifying agreement between the CPMs using Fleiss's Kappa ( $\kappa$ ), highlighted that agreement between all the surgical scores was moderate

( $\kappa = 0.37$ ), while that between all the TAVI models was poor ( $\kappa = 0.20$ ). The pairwise Fleiss's Kappa values are given in **Table 4**, which shows that there was moderate agreement between the FRANCE-2 and ACC TAVI models ( $\kappa = 0.33$ ). Risk stratifications were re-defined to give a population ratio of approximately 1:3:1 for low-, medium- and high-risk. Here, the results indicated marginally improved levels of agreement, but these were still moderate. Specifically, the Fleiss's Kappa across the surgical scores was 0.40 and that between the TAVI models was 0.20, with pairwise Fleiss's Kappa values given in Supplementary Table 11.

## Discussion

Clinical prediction models form the cornerstone of risk stratification for patients undergoing invasive procedures, helping to guide both treatment allocation and the consent process. However, their performance needs to be tested in large datasets independent to those in which the models were developed before they can be used in external populations (26,27). Our analysis of the UK TAVI registry has systematically demonstrated that outside their development cohorts, the German AV, FRANCE-2, OBSERVANT and ACC TAVI CPMs are miscalibrated and have low discrimination at predicting 30-day mortality. These results support previous work in this area (28). In the current study, the FRANCE-2 and ACC models had the highest discrimination out of all those considered, with the discrimination of the FRANCE-2 model these comparing favourably to the the internal validation results results reported reported when this these models was were derived (14,15). Additionally, although the ACC model was miscalibrated, the expected mortality was significantly close to the observed mortality across all subgroups considered in this analysis. However, of note is that the ACC model was predominately developed to predict in-hospital mortality, which potentially contributes to the agreement between the observed and expected event rates for this model.

The performance of any CPM is expected to drop when they are applied in populations external to the development set since patient mix and procedure techniques can vary between populations (26,27,29). Consequently, the findings of the current study are, perhaps, unsurprising given that the TAVI-CPMs achieved only moderate performance in their respective development datasets (12,14,15). Current TAVI cohorts predominantly represent a particularly high-risk and homogenous group of patients, potentially contributing to the lack of a highly predictive TAVI-CPM. Future TAVI-CPMs need to be developed by utilising the contemporary large registries that are emerging, which will inevitably require greater harmonisation between variable and outcome definitions amongst national datasets.

Moreover, many of the co-morbidities used in the development of CPMs are cardiovascular risk factors, with important non-cardiovascular co-morbidities not considered (30). In particular, frailty is not reflected in many of the CPMs, despite being particularly prevalent in elderly patients with AS and previous work suggesting frailty to be associated with poor TAVI outcomes (31,32). A CPM that aims to predict long-term mortality following TAVI found that the inclusion of frailty in their model significantly increased the discrimination (33). Similarly, a previously published CPM that aims to predict mortality and/ or a decline in quality of life following TAVI included an indication of 6-min walk test distance (34).

The present study indicated that the 30-day mortality was elevated in 2007 and 2008 over that in subsequent years, but the sensitivity analysis that excluded 2007/08 procedures indicated similar results to the main analysis. Previous studies have shown a learning curve associated with TAVI, but centre/ operator volume and outcome relationships remain debated (35–37). Nevertheless, measures of operator volume or experience are not used in CPMs since accounting for such variables would be inappropriate, particularly when the purpose of a CPM might be to benchmark an individual operators / centres performance. Similarly, the

addition of operator volume/ experience in a CPM would make it almost impossible for a physician to convey the predicted risk to a patient.

### **Comparison with Performance of the Surgical CPMs**

The current study confirms previous work in showing that the performance of the LES, ESII and STS models are poor at predicting 30-day mortality post TAVI (9–11). Despite being poor, the STS model outperformed the other surgical models, with the STS expected 30-day mortality rate not significantly different from the observed 30-day mortality rate. This finding has been previously observed (9,11) and is perhaps attributable to the fact the STS score has a specific model for isolated valve surgery (8). Of note, previous TAVI registries have reported mean STS values higher than that found in this study, perhaps due to the assumptions made in our study regarding the calculation of the STS model. For example, the FRANCE TAVI registry reported STS values of around 18%, while the Italian CoreValve registry reported values of 11% (38,39).

Nonetheless, comparing the surgical CPMs to the TAVI-CPMs highlights that the latter performed better than the former when internally validated (12,14) and the current study shows that the FRANCE-2 and ACC models outperformed the surgical scores. Surgical CPMs are limited in their use in transcatheter procedures because they were derived from surgical populations. Not only are the procedural risks of TAVI different from those in SAVR, but there is a lack of grading between the severities of co-morbidities in the surgical CPMs. For example, chronic obstructive pulmonary disease (COPD) is a risk factor in LES, but there is no further distinction between the severity of COPD or even other severe lung disease. Since the heart-team considers such severities when deciding between SAVR and TAVI, grading of co-morbidities should be included in future TAVI-CPMs.

## Patient-level Agreement Analysis

This study highlighted that the classification of patient risk varies between multiple CPMs, even when comparing surgical and TAVI based CPMs separately. A Pearson correlation coefficient of 0.56 has previously been reported between the LES and STS score (10), with similar correlation between these scores reported in other studies (11). Such an analysis does not necessarily indicate the level of agreement between two risk models, since the correlation is only assessing the linear relationship between them (40). Although the current study found higher agreement between the surgical models than between the TAVI models, this was driven by the ESII being an updated version of the LES. The lack of agreement between the scores further highlights previous published recommendations that risk assessment should be based on heart-team discussion in combination with multiple CPMs (4).

## Limitations

A limitation of the current work is that assumptions were required when linking the definitions of model variables with the TAVI dataset, as described in the Supplementary Material. For example, the lowest LV function category in the ESII model is LVEF<20% whereas that in the UK TAVI dataset is LVEF<30%, with this analysis assuming these definitions to be equivalent. Such assumptions are an artefact of different recording practices between national registries. Accordingly, some of the surgical CPMs could not be calculated exactly as they were published, which could induce bias into the calculated predicted risks.

This study used Surrogate variables ~~were used~~ to mitigate this wherever possible and all assumptions were made to reflect the TAVI procedure as accurately as possible. As noted above, the calculated STS score in this study is lower than previously reported values from other TAVI registries. Lack of variables including Mitral valve, hypertension and severity of pulmonary disease ~~variables~~ could have contributed to this, although but our findings

compare favourably to previous work. Similarly, the assumption of risk factor absent for variables that were included in CPMs but not recorded in the UK TAVI registry (e.g. mitral valve replacement or infective endocarditis) may induce bias, but any such bias is likely to be negligible given the variables where this assumption was needed.

### **Implications for Future Work**

Based on this work, the development of further TAVI-CPMs is recommended in populations of interest. Although there is an indication of feasibility of TAVI in intermediate risk patients (5), TAVI-CPMs are still required, especially for procedure audit purposes and risk stratification analyses. Rather than developing new scores from scratch, model updating techniques could be applied to the current TAVI-CPMs to adapt them to new national cohorts (41). For instance, re-fitting the current models to the population of interest and/or the addition of new risk factors, such as frailty, could improve prediction (31,42). Further work in this area is recommended. Secondly, developing TAVI models that predict both short- and long-term outcomes would be particularly valuable, especially if they included a measure of futility.

### **Conclusions**

The FRANCE-2 and ACC TAVI models had the highest performance across all CPMs considered. However, all the CPMs had low calibration and discrimination, reducing their suitability for risk stratification outside their development cohorts. Future iterations of existing TAVI models may benefit from including non-cardiovascular co-morbidities such as frailty. ~~, rendering them unsuitable for use for risk stratification outside of their development populations.~~ Hence, ~~†~~the derivation of new TAVI-CPMs in contemporary large registries is recommended, but it remains to be determined if this is best achieved by updating/ revising

existing TAVI scores, by developing new CPMs in specific cohorts, or a combination of the two.

### **Acknowledgements**

We would like to acknowledge the National Institute for Cardiovascular Outcomes Research (NICOR) for provided the UK TAVI registry extract for this study.

## References

1. Lung B, Cachier A, Baron G, Messika-Zeitoun D, Delahaye F, Tornos P, et al. Decision-making in elderly patients with severe aortic stenosis: Why are so many denied surgery? *Eur Heart J*. 2005;26:2714–20.
2. Leon M, Smith C, Mack M, Miller DC, Moses JW, Svensson LG, et al. Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. *N Engl J Med*. 2010;363(17):1597–607.
3. Smith C, Leon M, Mack M, Miller DC, Moses JW, Svensson LG, et al. Transcatheter versus surgical aortic-valve replacement in high-risk patients. *N Engl J Med*. 2011;364(23):2187–98.
4. Vahanian A, Alfieri O, Andreotti F, Antunes MJ, Baron-Esquivias G, Baumgartner H, et al. Guidelines on the management of valvular heart disease (version 2012): The Joint Task Force on the Management of Valvular Heart Disease of the European Society of Cardiology (ESC) and the European Association for Cardio-Thoracic Surgery (EACTS). *Eur Heart J*. 2012;33(19):2451–96.
5. Leon MB, Smith CR, Mack MJ, Makkar RR, Svensson LG, Kodali SK, et al. Transcatheter or Surgical Aortic-Valve Replacement in Intermediate-Risk Patients. *N Engl J Med*. 2016;374(17):1609–20.
6. Nashef SAM, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, et al. EuroSCORE II. *Eur J Cardio-Thoracic Surg*. 2012;41(4):734–45.
7. Roques F. The logistic EuroSCORE. *Eur Heart J*. 2003;24(9):882.
8. 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(1):S23–42.
9. Durand E, Borz B, Godin M, Tron C, Litzler P-Y, Bessou J-P, et al. Performance analysis of EuroSCORE II compared to the original logistic EuroSCORE and STS scores for predicting 30-day mortality after transcatheter aortic valve replacement. *Am J Cardiol.* 2013;111(6):891–7.
  10. Piazza N, Wenaweser P, van Gameren M, Pilgrim T, Tsikas A, Otten A, et al. Relationship between the logistic EuroSCORE and the Society of Thoracic Surgeons Predicted Risk of Mortality score in patients implanted with the CoreValve ReValving system--a Bern-Rotterdam Study. *Am Heart J.* 2010;159(2):323–9.
  11. Ben-Dor I, Gaglia MA, Barbash IM, Maluenda G, Hauville C, Gonzalez MA, et al. Comparison between Society of Thoracic Surgeons score and logistic EuroSCORE for predicting mortality in patients referred for transcatheter aortic valve implantation. *Cardiovasc Revasc Med.* 2011;12(6):345–9.
  12. Capodanno D, Barbanti M, Tamburino C, D'Errigo P, Ranucci M, Santoro G, et al. A simple risk tool (the OBSERVANT score) for prediction of 30-day mortality after transcatheter aortic valve replacement. *Am J Cardiol.* 2014;113(11):1851–8.
  13. 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 Cardio-Thoracic Surg.* 2013;43(5):971–7.
  14. Iung B, Laouéan C, Himbert D, Eltchaninoff H, Chevreul K, Donzeau-Gouge P, et al. Predictive factors of early mortality after transcatheter aortic valve implantation: individual risk assessment using a simple score. *Heart.* 2014;100(13):1016–23.
  15. Edwards FH, Cohen DJ, O'Brien SM, Peterson ED, Mack MJ, Shahian DM, et al. Development and Validation of a Risk Prediction Model for In-Hospital Mortality

- After Transcatheter Aortic Valve Replacement. *JAMA Cardiol.* 2016;1(1):46.
16. Ludman PF. The UK transcatheter aortic valve implantation registry; one of the suite of registries hosted by the National Institute for Cardiovascular Outcomes Research (NICOR). *Heart.* 2012;98(24):1787–9.
  17. Rubin DB. *Multiple Imputation for Nonresponse in Surveys.* John Wiley & Sons; 1987.
  18. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;338(1):b2393–b2393.
  19. Cox D. Two further applications of a model for binary regression. *Biometrika.* 1958;45(3):562–5.
  20. Brier GW. Verification of Forecasts Expressed in Terms of Probability. *Mon Weather Rev.* 1950;78(1):1–3.
  21. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull.* 1971;76(5):378–82.
  22. R Core Team R. *R: A Language and Environment for Statistical Computing.* Team RDC, editor. R Foundation for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2016. Available from: <http://www.r-project.org/>
  23. Buuren S Van, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw.* 2011;45(3):1–67.
  24. Wickham H. *ggplot2: elegant graphics for data analysis.* Springer New York; 2009.
  25. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC*

- Bioinformatics. 2011;12:77.
26. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605–b605.
  27. Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ*. 2016;353:i3140.
  28. Halkin A, Steinvil A, Witberg G, Barsheshet A, Barkagan M, Assali A, et al. 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 A. *Int J Cardiol.*; 2016;215:227–31.
  29. Steyerberg EW, Harrell FE, Borsboom GJJ., Eijkemans MJ., Vergouwe Y, Habbema JDF. Internal validation of predictive models. *J Clin Epidemiol*. 2001;54(8):774–81.
  30. Mamas MA, Fath-Ordoubadi F, Danzi GB, Spaepen E, Kwok CS, Buchan I, et al. Prevalence and Impact of Co-morbidity Burden as Defined by the Charlson Co-morbidity Index on 30-Day and 1- and 5-Year Outcomes After Coronary Stent Implantation (from the Nobori-2 Study). *Am J Cardiol*. 2015;116(3):364–71.
  31. Puls M, Sobisiak B, Bleckmann A, Jacobshagen C, Danner BC, Hünlich M, et al. Impact of frailty on short and long term morbidity and mortality after transcatheter aortic valve implantation: risk assessment by Katz Index of activities of daily living. *EuroIntervention*. 2014;10(5):609–19.
  32. Schoenenberger AW, Stortecky S, Neumann S, Moser A, Jüni P, Carrel T, et al. Predictors of functional decline in elderly patients undergoing transcatheter aortic valve implantation (TAVI). *Eur Heart J*. 2013;34(9):684–92.

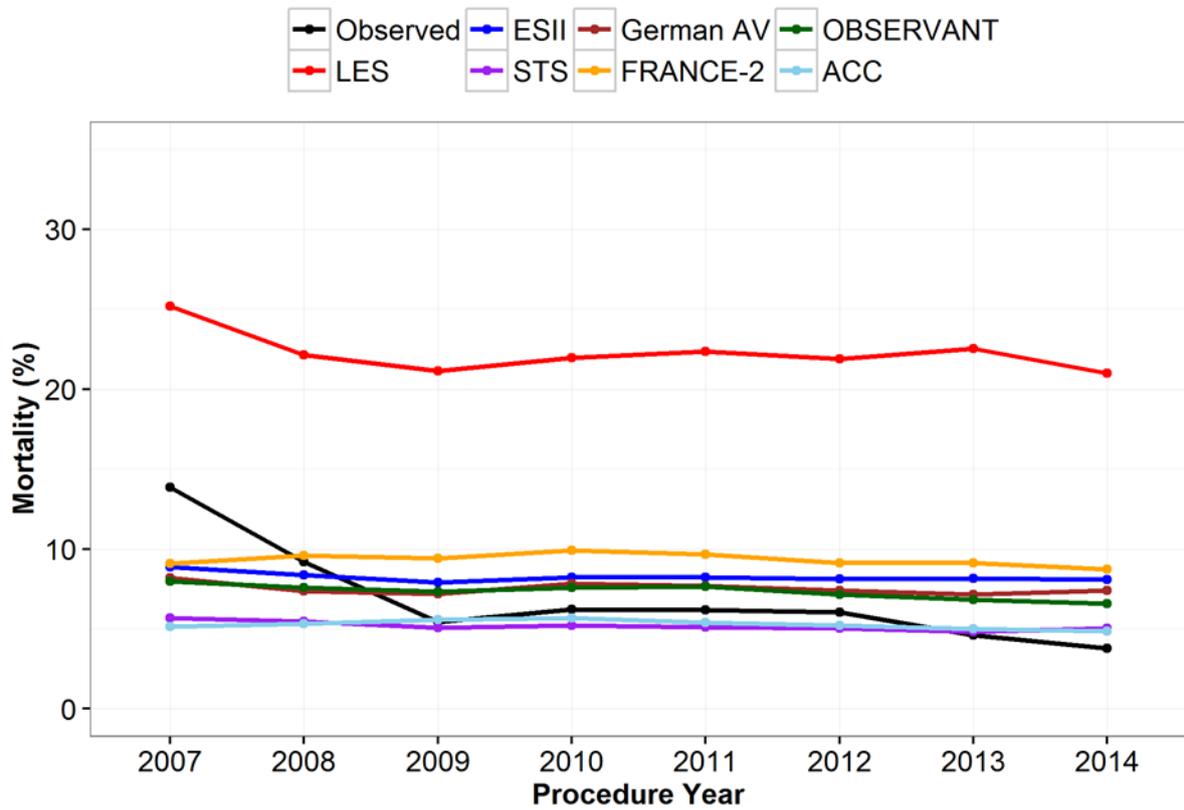
33. Seiffert M, Sinning J-M, Meyer A, Wilde S, Conradi L, Vasa-Nicotera M, et al.  
Development of a risk score for outcome after transcatheter aortic valve implantation.  
*Clin Res Cardiol.* 2014;103(8):631–40.
34. Arnold S V., Reynolds MR, Lei Y, Magnuson EA, Kirtane AJ, Kodali SK, et al.  
Predictors of poor outcomes after transcatheter aortic valve replacement results from the PARTNER (Placement of Aortic Transcatheter Valve) trial. *Circulation.* 2014;129(25):2682–90.
35. Ludman PF, Moat N, de Belder MA, Blackman DJ, Duncan A, Banya W, et al.  
Transcatheter Aortic Valve Implantation in the UK: Temporal Trends, Predictors of Outcome and 6 Year Follow Up: A Report from the UK TAVI Registry 2007 to 2012. *Circulation.* 2015;131(13):1181–90.
36. Gurvitch R, Tay EL, Wijesinghe N, Ye J, Nietlispach F, Wood DA, et al.  
Transcatheter aortic valve implantation: Lessons from the learning curve of the first 270 high-risk patients. *Catheter Cardiovasc Interv.* 2011;78(7):977–84.
37. Kempfert J, Rastan A, Holzhey D, Linke A, Schuler G, van Linden A, et al.  
Transapical Aortic Valve Implantation: Analysis of Risk Factors and Learning Experience in 299 Patients. *Circulation.* 2011;124(suppl 1):S124–9.
38. Eltchaninoff H, Prat A, Gilard M, Leguerrier A, Blanchard D, Fournial G, et al.  
Transcatheter aortic valve implantation: early results of the FRANCE (FRench Aortic National CoreValve and Edwards) registry. *Eur Heart J.* 2011;32(2):191–7.
39. Ussia GP, Barbanti M, Petronio AS, Tarantini G, Ettori F, Colombo A, et al.  
Transcatheter aortic valve implantation: 3-year outcomes of self-expanding CoreValve prosthesis. *Eur Heart J.* 2012;33(8):969–76.
40. Altman D, Bland J. Measurement in medicine: the analysis of method comparison

studies. *Stat.* 1983;32(3):307–17.

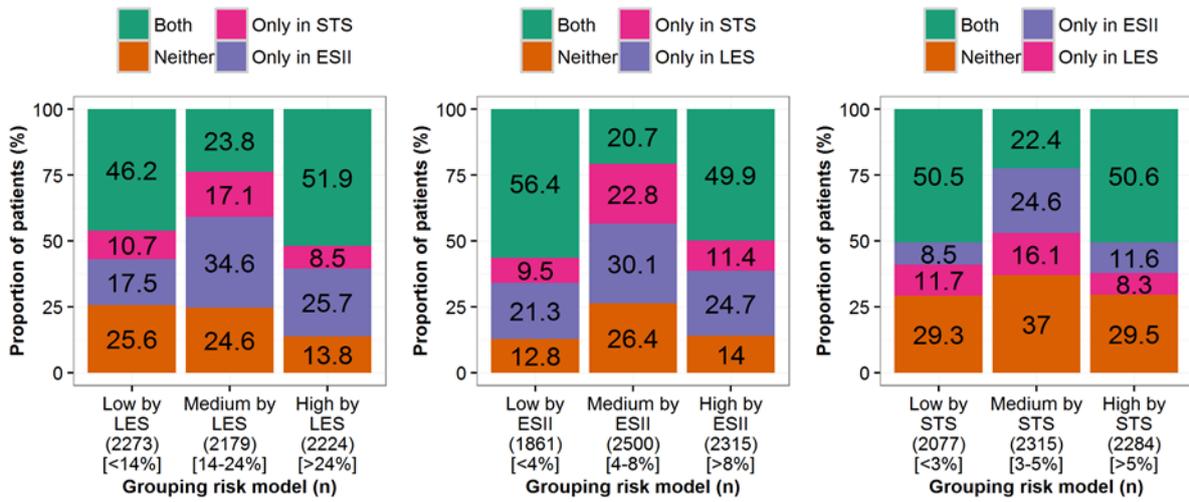
41. Steyerberg EW, Borsboom GJJM, van Houwelingen HC, Eijkemans MJC, Habbema JDF. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med.* 2004;23(16):2567–86.
42. Puri R, Iung B, Cohen DJ, Rodés-Cabau J. TAVI or No TAVI: identifying patients unlikely to benefit from transcatheter aortic valve implantation. *Eur Heart J.* 2016;ehv756.

## Figure Legends

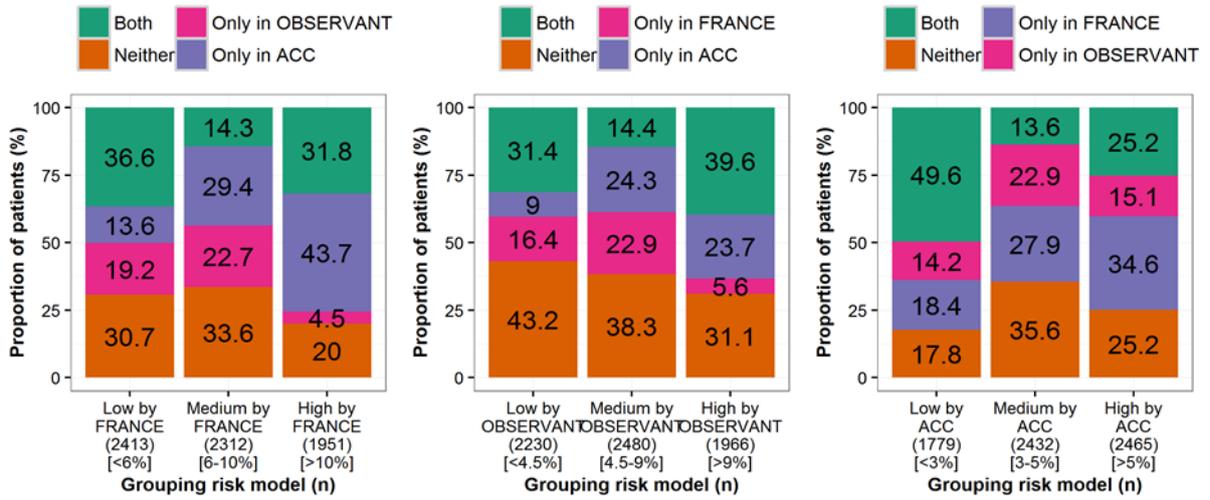
**Figure 1.** Temporal changes in observed and expected mortality over each of the CPMs.



**Figure 2.** The proportion of patients that agree in risk allocation over the surgical based CPMs. Each bar represents a risk stratification by one of the surgical CPMs, with the segments of that bar showing the proportion of patients that were also grouped in that risk strata by none, one or both of the other surgical CPMs.



**Figure 3.** The proportion of patients that agree in risk allocation over the TAVI based CPMs. Each bar represents a risk stratification by one of the TAVI-CPMs, with the segments of that bar showing the proportion of patients that were also grouped in that risk strata by none, one or both of the other TAVI-CPMs



## Tables

**Table 1. Summary statistics, before multiple imputations of the missing data, of baseline and procedural characteristics in the UK TAVI dataset.**

| Variable  | Summary (% of <i>n</i> = 6676) | Missing (% of <i>n</i> = 6676) |
|---|--------------------------------|--------------------------------|
| Age, mean [range]                                     | 81.3 [29-101]                  | 0 (0%)                         |
| Females, n (%)  | 3085 (46.2%)                   | 22 (0.3%)                      |
| Weight (Kg), mean [range]                             | 74.0 [32.0-190.0]              | 131 (2.0%)                     |
| Height (m), mean [range]                              | 1.6 [1.1-2.4]                  | 159 (2.4%)                     |
| NYHA  |                                | 42 (0.6%)                      |
| Class I, n (%)  | 185 (2.8%)                     |                                |
| Class II, n (%)                                       | 1116 (16.7%)                   |                                |
| Class III, n (%)                                      | 4186 (62.7%)                   |                                |
| Class IV, n (%)                                       | 1147 (17.2%)                   |                                |
| Creatinine, $\mu\text{mol/L}$ , mean [range]          | 114.3 [29.0-1044.0]            | 73 (1.1%)                      |
| Creatinine greater than 200 $\mu\text{mol/L}$ , n (%) | 379 (5.7%)                     | 73 (1.1%)                      |
| LVEF  |                                | 59 (0.88%)                     |
| $\geq 50\%$ , n (%)                                   | 4074 (61.0%)                   |                                |
| 30 – 49%, n (%)                                       | 1929 (28.9%)                   |                                |
| $< 30\%$ , n (%)                                      | 614 (9.2%)                     |                                |

|                                  |              |            |
|----------------------------------|--------------|------------|
| Extracardiac Arteriopathy, n (%) | 1572 (23.5%) | 88 (1.3%)  |
| Diabetes                         |              | 35 (0.52%) |
| Dietary control, n (%)           | 290 (4.3%)   |            |
| Oral medicine, n (%)             | 884 (13.2%)  |            |
| Insulin, n (%)                   | 363 (5.4%)   |            |
| Dialysis, n (%)                  | 127 (1.9%)   | 66 (0.99%) |
| MI                               |              | 33 (0.49%) |
| within 90 days of TAVI, n (%)    | 153 (2.3%)   |            |
| within 30 days of TAVI, n (%)    | 65 (0.97%)   |            |
| within 24 hours of TAVI, n (%)   | 6 (0.09%)    |            |
| Procedure Urgency                |              | 7 (0.10%)  |
| Elective, n (%)                  | 5853 (87.7%) |            |
| Urgent, n (%)                    | 772 (11.6%)  |            |
| Emergency, n (%)                 | 35 (0.52%)   |            |
| Salvage, n (%)                   | 9 (0.13%)    |            |
| Valve Type                       |              | 31 (0.46%) |
| Edwards SAPIEN Valve, n (%)      | 3684 (55.2%) |            |

|                                 |              |             |
|---------------------------------|--------------|-------------|
| Medtronic CoreValve, n          | 2735 (41.0%) |             |
| (%)                             |              |             |
| Access Route                    |              | 13 (0.19%)  |
| TF Access, n (%)                | 4965 (74.4%) |             |
| Transapical Access, n (%)       | 1064 (15.9%) |             |
| Chronic Lung Disease, n (%)     | 1879 (28.1%) | 94 (1.4%)   |
| Cerebrovascular Disease, n      | 1139 (17.1%) | 35 (0.52%)  |
| (%)                             |              |             |
| Previous cardiac surgery, n (%) | 2087 (31.3%) | 35 (0.52%)  |
| Critical preoperative state, n  | 110 (1.6%)   | 81 (1.2%)   |
| (%)                             |              |             |
| PA systolic > 60mmHg            | 785 (11.8)   | 1860 (27.9) |
| LMS > 50% or Triple vessel      | 887 (13.3%)  | 74 (1.1%)   |
| disease, n (%)                  |              |             |

---

*LMS: Left Main Stem Disease, LVEF: Left Ventricular Ejection Fraction, MI: Myocardial Infarction,*

*NYHA: New York Heart Association Functional Classification, PA: Pulmonary Artery, TF:*

*Transfemoral Access Route*

**Table 2. Absolute and relative differences of the expected to observed 30-day mortalities.**

| <b>Risk Model</b> | <b>Expected 30-day Mortality (%)</b> | <b>Absolute Difference to observed mortality*</b> | <b>Relative Difference to observed mortality † (%)</b> |
|-------------------|--------------------------------------|---|--|
| LES               | 21.9                                 | 16.5  | 405.6  |
| ESII              | 8.1                                  | 2.7   | 150.0  |
| STS               | 5.1                                  | 0.3   | 94.4   |
| German AV         | 7.4                                  | 2.0   | 137.0  |
| FRANCE-2          | 9.2                                  | 3.8   | 170.4  |
| OBSERVANT         | 7.1                                  | 1.7   | 131.5  |
| ACC TAVI          | 5.2                                  | 0.2   | 96.3   |

*\*: calculated as the absolute value of expected minus observed. †: calculated as (expected/observed)×100.*

**Table 3. Calibration, discrimination and Brier score for 30-day mortality in the whole cohort.**

| <b>Risk Model</b> | <b>Calibration Intercept<br/>(95% CI) *</b> | <b>Calibration<br/>Slope (95% CI)</b> | <b>AUC (95% CI)</b> | <b>Brier<br/>Score</b> |
|-------------------|---|---------------------------------------|---------------------|------------------------|
| LES               | -1.75 (-1.86, -1.64)                        | 0.35 (0.23, 0.48)                     | 0.57 (0.54, 0.61)   | 0.093                  |
| ESII              | -0.47 (-0.59, -0.36)                        | 0.40 (0.28, 0.53)                     | 0.59 (0.55, 0.62)   | 0.054                  |
| STS               | <b>0.07 (-0.04, 0.18)</b>                   | 0.56 (0.42, 0.71)                     | 0.60 (0.57, 0.63)   | 0.051                  |
| German AV         | -0.36 (-0.47, -0.25)                        | 0.44 (0.32, 0.57)                     | 0.59 (0.56, 0.62)   | 0.053                  |
| FRANCE-2          | -0.60 (-0.71, -0.49)                        | 0.69 (0.53, 0.86)                     | 0.62 (0.59, 0.65)   | 0.053                  |
| OBSERVANT         | -0.31 (-0.42, -0.20)                        | 0.39 (0.25, 0.53)                     | 0.57 (0.54, 0.60)   | 0.052                  |
| ACC TAVI          | <b>0.04 (-0.07, 0.15)</b>                   | 0.67 (0.52, 0.82)                     | 0.64 (0.60, 0.67)   | 0.051                  |

*\*The reported calibration intercept is that estimated assuming a slope of one; satisfactory calibration would occur if the 95% confidence intervals for the calibration intercept and slope span zero and one respectively.*

**Table 4. Cut-off values and the pairwise kappa values for the surgical and TAVI based CPMs.**

| CPM            | Low    | High  | Fleiss's kappa † |          |           |           |
|----------------|--------|-------|------------------|----------|-----------|-----------|
|                | Risk*  | Risk* |                  |          |           |           |
| Surgical Based |        |       | LES              | ESII     | STS       | German AV |
| LES            | ≤ 14%  | >24%  | n/a              | 0.50     | 0.29      | 0.34      |
| ESII           | ≤ 4%   | >8%   | 0.50             | n/a      | 0.34      | 0.27      |
| STS            | ≤ 3%   | >5%   | 0.29             | 0.34     | n/a       | 0.47      |
| German AV‡     | ≤ 4%   | >8%   | 0.34             | 0.27     | 0.47      | n/a       |
| TAVI Based     |        |       | German AV        | FRANCE-2 | OBSERVANT | ACC       |
| German AV‡     | ≤ 4%   | >8%   | n/a              | 0.17     | 0.13      | 0.26      |
| FRANCE-2       | ≤ 6%   | >10%  | 0.17             | n/a      | 0.14      | 0.33      |
| OBSERVANT      | ≤ 4.5% | >9%   | 0.13             | 0.14     | n/a       | 0.18      |
| ACC            | ≤ 3%   | >5%   | 0.26             | 0.33     | 0.18      | n/a       |

\*: All cut-off values were chosen to give approximately equal numbers of patients in low-, medium- and high-risk categories. Patients with predicted risks between the low- and high-risk cut-off values were classified as medium risk. †: Values give the pairwise agreement between the two indicated CPMs. ‡: The German AV model was derived in a cohort with both surgical and TAVI patients, thus is considered in both groups of models.