Prediction models for mortality after transcatheteric aortic valve implantation (TAVI)

Al-Farra, H.

Publication date
2023

Document Version
Final published version

Prediction models for mortality after transcatheteric aortic valve implantation (TAVI)

Hatem Al-Farra

Prediction models for mortality after transcatheteric aortic valve implantation (TAVI)

Hatem Al-Farra

Prediction models for mortality after transcatheteric aortic valve implantation (TAVI)

PhD thesis, University of Amsterdam, Amsterdam, the Netherlands

ISBN: 978-94-6469-344-7

Cover and layout: Hatem Al-Farra

Printed by: ProefschriftMaken

Prediction models for mortality after transcatheteric aortic valve implantation (TAVI)

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op 04 juli 2023, te 16:00 uur

door

Hatem Al-Farra

geboren te Homs

Promotiecommissie

Promotores:	prof. dr. A. Abu-Hanna	AMC - UvA
	prof. mr. dr. B.A.J.M. de Mol	AMC - UvA
Copromotores:	dr. ir. A.C.J. Ravelli	AMC - UvA
	prof. dr. J.P.S. Henriques	AMC - UvA
Overige leden:	prof. dr. H.A. Marquering	AMC - UvA
	prof. dr. F.W. Asselbergs	Universiteit Utrecht
	prof. dr. M.C. Schut	Vrije Universiteit Amsterdam
	dr. M.M. Vis	AMC - UvA
	dr. G. Cina	AMC - UvA

Faculteit der Geneeskunde

Financial support by the Dutch Heart Foundation for the publication of this thesis is gratefully acknowledged

Imagination is essential and it comes first,
for without imagination, we are aimless.

C. Northcote Parkinson

Contents

List of tables

Chapter 2

Chapter 3

- E-component Table 7: Sensitivity analysis by simulating the values of the absent variable predictor (acute-pulmonary-oedema) and calculated the performance measures of the updated-model in each of 1000 bootstrap samples
- E-component Table 8: TRIPOD Checklist: Prediction Model Validation

Chapter 4

- Table 1. Early (30-day) mortality rates after TAVI in the total population (n=9144), and the univariate analysis of the variable predictors
- Table 2. The predictor variables of the NHR-TAVI model, with the pooled regression coefficients, standard errors of the regression coefficients, the odds ratios and p-values
- Table 3. Results of temporal-validation and cross-validation for the updated-models ACC-TAVI (updated with model-revision) and FRANCE-2 (updated with model-intercept-update).
- E-supplement table 1. Basic characteristics of the total (N=9144) TAVI-population, stratified by the status of the early 30-day mortality after TAVI (Survivors = 8776, vs. non-Survivors = 368)
- E-supplementary table 2. Missing values and percentages of the variable predictors before imputation of the NHR TAVI dataset of 9144 patients
- E-supplementary table 3. Results of the cross-validated predictive performance of model revision ACC-TAVI on TAVI-NHR cohort 2013-2018
- E-supplementary table 4. Results of the external validation and the performance of the updated version of the Model revision IRRMA on the TAVI-NHR cohort 2013-2018
- E-supplement table 5. Observed early-mortality rates for the continuous variables divided into 3, or 4 subgroups.
- E-supplementary table 6. Performance measures of the internal validation of the prediction strategy in 10-fold cross-validation, for including predictors appearing in at least 3 or 4 of the 5 imputation datasets
- E-supplementary table 7A. Predictor variables for the TAVI logistic regression model predicting the 30-day early-mortality, variable selection based on LASSO
- E-supplementary table 7B. Internal validation of the prediction strategy based on LASSO
- E-supplementary table 8A. Predictor variables and their pooled coefficients, odds ratios, standard errors (SE), and p-value for the model after including the variable frailty-status in the five imputed datasets
- E-supplementary table 8B. Predictor variables and their pooled coefficients, odds ratios, standard errors (SE), and p-value for the model after including the variable frailty-status in the five imputed datasets from the dataset that include the patients with complete frailty-status score (n = 1295)
- E-supplementary table 8C. Results of the sensitivity analysis in which we included the variable frailty-status. The performance measures of the internal validation of this prediction strategy in 10-fold cross-validation of these two models
- E-supplementary table 9. Research Reporting Guideline checklist TRIPOD: Prediction model development and validation

Chapter 5

- Table 1. The incidences and trends of major adverse cardiac events (MACE) and mortality after TAVI procedures and the trends over the years (2013-2018) for the 9144 TAVI-patients from the Netherlands Heart Registration registry (NHR)
- Table 2. The incidence and totals of the major adverse cardiac events (MACE) in different age groups of the 9144 TAVI patients. The patients are sub-grouped according to the following age groups <75, 75-80, and >80 year

- Table 3. Incidences of TAVI-related major adverse cardiac events (MACE) in the 9144 patients divided into three operative risk groups (33.3 % tertiles) EuroSCORE-II estimated
- E-supplementary table 1. Baseline characteristics of the TAVI study population stratified by the outcome of Permanent pacemaker implantation within the first 30 days after the TAVI procedure. Baseline characteristics stratified by mortality status can be reviewed in our pervious published study in (31)
- E-supplementary table 2a. Incidences (%) of the MACE and mortality in each calendar-year as calculated in this study, and as reported in the NHR annual report of 2021
- E-supplementary table 2b. Numbers of patients in each year and each age group, numbers and percentages of mortality that occurred on 30-days or one year after TAVI in each calendar year from 2013 up to 2018. Note this data is about the patients with known mortality outcomes (N= 9144)
- E-supplementary table 2c. Total number of mortality that occurred in each sequential year following the TAVI procedure, in the whole cohort, and each age group.
- E-supplementary table 2d. Cumulative numbers of mortality the occurred in the years following the TAVI procedure
- E-supplementary table 3. Incidence and totals of the TAVI-related major adverse cardiac events (MACE) in three age groups of 33.3% tertiles of the age of the 9144 TAVI patients
- E-supplementary table 4. Sub-analysis for incidences of TAVI-related major adverse cardiac events in different age groups over the years. Note that the number of TAVI procedures increased over time from 20% in 2013-2014 to 48% in 2017-2018. Also, note that 60% of the patients were 80 years and older. The percentage of 80+ patients slightly declined over time from 62% to 58%

Chapter 6

- Table 1. Details on the steps undertaken during the development of the SIM-plan
- Table 2. Details of the main end products from each step
- Table 3. The business goals areas of the Heart Center SIM-plan (2018-2022)
- Table 4. The IT goals areas of the Heart Center SIM-plan (2018-2022)
- Table 5. The top priority business goals for the Heart Center (2018-2022)
- Table 6. The strategic project portfolio of the Heart Center (2018-2022), (6 out of the 25 projects)

List of figures

Chapter 1

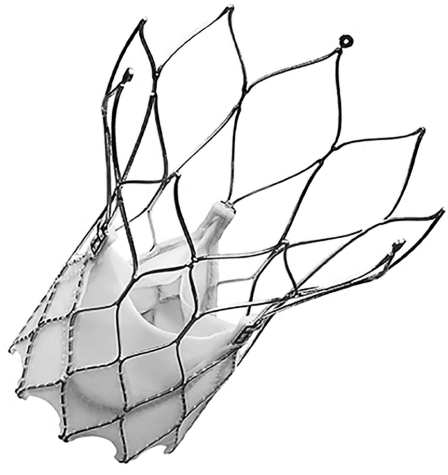
Chapter 2

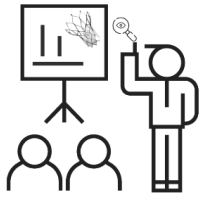
Chapter 3

Chapter 4

Chapter 5

-





Chapter 1:

General Introduction

General introduction

Cardiovascular diseases (CVD) are a global leading cause of death. Each year about 17.9 million people die due to CVD. This represents, according to the World Health Organization (WHO), 32% of all worldwide deaths. Valvular heart diseases form an important category of CVD. Aortic valve stenosis (AS) is the most common valvular disease in Europe and the United States of America. Generally, AS results from the aging process, which affects the aortic valve. About 12.4% of elderly people (>75 years) have AS (1). AS is considered the most common pathology that obstructs the blood outflow from the left heart ventricle. The degree of stenosis varies, and it results in asymptomatic to very severe conditions. Some cases of AS can be treated medically with careful observation of the patient. In more severe aortic stenosis, surgical aortic valve replacement (SAVR) may be required. However, not all patients would be considered surgically fit candidates for this procedure. And up till recently, these patients had no option for treatment and remained highly symptomatic, and had a very poor prognosis.

Transcatheter aortic valve implantation (TAVI) has been introduced to treat severe symptomatic AS (2). In 2002, TAVI was introduced as an alternative treatment option for patients with severe AS who were deemed SAVR inoperable due to high operative-risk (2-4). The first TAVI in the Netherlands was in 2005 (5). Several studies reported satisfactory safety results for the TAVI procedure (6, 7). Over the years, TAVI became the standard treatment choice in that high operative-risk population (8, 9).

Since 2017, treatment of AS with TAVI has been extended to patients with moderate and moderate-low operative-risk (10). Despite lower perioperative risks, care for patients undergoing TAVI requires proper perioperative-risk assessment for the post-operative major adverse cardiac events (MACE), such as mortality, the need for permanent pacemaker implantation (PPI), major severe bleeding (MVB), and stroke.

Prediction models should enable heart teams to identify patients at higher risk for developing procedure-related adverse outcomes. These models assist clinicians by objectively assessing the risk of outcomes, like mortality or MACE, after TAVI, and optimally informing patients. They are intended to eliminate some of the undesired subjectivity inherent in clinical decision-making and provide personalized care. Therefore, prediction models can aid the joint management with the patient and can help improve the provided patients' care.

Today, various mortality prediction models (MPM) are used to predict mortality outcomes for TAVI-patients. Some of them were specifically developed in a TAVI population (further referred to as TAVI-specific MPMs) (11-13). While other models were developed in a cardiac surgery population (referred to as surgical MPMs) (14, 15). Moreover, there are few- prediction models that estimate the risk of individual TAVI-related complications (MACEs), such as predicting the occurrence of MVB or PPI after TAVI.

Practice variations, differences in perioperative care, and national consensus guidelines driven by reimbursement differences amongst countries may affect the selection of patients. Therefore, it is necessary to investigate to which extent models can be generalized to patient groups, or whether specific validation based on national registries is mandatory. There are various prediction models with the outcome mortality (MPMs), which are not validated for external populations and certainly not for the Dutch population. For using such MPMs, it is essential to study their merits in external populations. Nevertheless, it is not known beforehand whether these national validation efforts bear additional value to the existing knowledge.

In this thesis, we investigate the predictive performance of such models in a large Dutch TAVI population. However, for providing accurate predictions of TAVI outcomes, more information may be needed than currently captured at the hospital level. That's why a study of the information environment as a whole in a heart center merits investigation. Specifically, aside from the various concrete needs to support the diagnosis, prognosis, and treatment plans, a heart center, as a whole, needs to strategically manage its health information systems. The introduction of several health information systems was aimed to reduce the workload and enhance patient care.

However, these various systems generate a huge amount of data, accompanied by complexities to deal with. In addition, communication between various health information systems (from one to another hospital) is poor. This requires that specialized as well as general heart centers are always ready for market changes, and armed with a flexible strategic information management plan (SIM-plan) that would help to manage and efficiently operate the information systems. Few studies have addressed the experience of developing such SIM-plans. Therefore, we share and elaborate on our experience in developing a SIM-plan in a heart center including means for regional, national, and international communication.

Below, we first present preliminaries on TAVI care, and its post-operative complications; followed by a description of relevant prediction models, especially their validation and performance measures; and SIM planning. We then state the research questions and outline the structure of this thesis.

Transcatheter aortic valve implantation

Transcatheter aortic valve implantation (TAVI) was introduced as an alternative treatment option for inoperable high operative-risk patients.

In 1989, TAVI was experimentally introduced. The Danish cardiologist Henning Rud Andersen developed a foldable heart valve prosthesis called the stent-valve (16). In 2002, the first human TAVI case was performed. TAVI is a minimally invasive percutaneous procedure to replace the stenotic, thickened, and calcified aortic valve, through a catheter.

The material used in the replacement valves is made of biological tissues (pig or cow). During the procedure an X-ray -or echocardiogram- guided catheter will be inserted to reach the heart via the transfemoral artery, the aorta directly, or through the heart apex from the chest wall (figure 1). The in-place valve implantation occurs after either expanding the tip of the catheter with a balloon or using another sort of self-expandable valve.

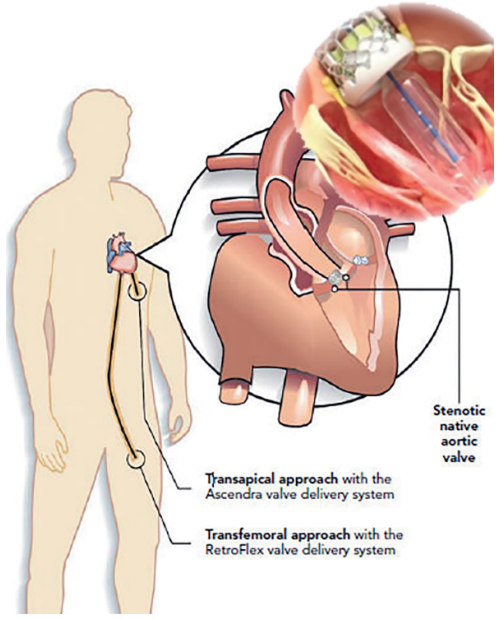


Figure 1 Illustration for the TAVI procedure

Over the years, TAVI techniques and devices have been intensely developed and improved. Due to continuous scientific and industrial advancements, different generations of implants (figure 2) and procedural tools have been introduced to alleviate procedure-related MACEs. Each type (brand) of TAVI-implants has its features, advantages, and disadvantages (17).

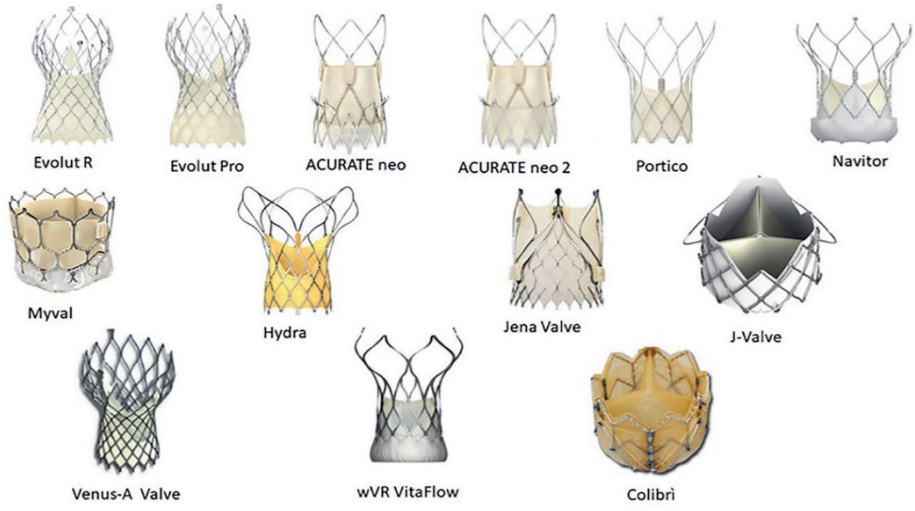


Figure 2 Different generations of the implants used for TAVI

On the other hand, the experience of the operators has improved over the last years. In other words, TAVI has become a common cardiac procedure, and the learning curve is likely to have flattened. All these factors have led to better outcomes and a lower rate of procedure-related complications. After 15 years after its introduction, the number of performed procedures was about 300,000 globally, in more than 1200 centers using either balloon-expandable or self-expanding prostheses (18).

Durko et. al. has estimated that the annual number of TAVI candidates in Europe and Northern-America is around 115,000 and 58,000, respectively. These figures would increase, to 177,000 and 90,000 respectively, if the evidence demonstrates that more low-risk patients would be safely operated (19). The increase of TAVI in The Netherlands will be described in detail in chapter 5.

Prediction models for TAVI

TAVI-patient selection is usually carried out by heart teams. Heart teams consist of cardiothoracic surgeons, radiologists, imaging experts, cardiologists, and interventional cardiologists. Patient selection is based on a careful assessment of the aortic valve and the patient's condition. If a patient with severe AS is deemed SAVR inoperable (due to high operative-risk) by cardiothoracic surgeons, the patient might be considered a TAVI candidate. In many countries, heart teams apply their adjusted national treatment guidelines, although they are still based on guidelines such as the European Society of Cardiology guidelines for the management of valvular heart disease (which are used in Europe and many other countries too).

In general, such guidelines recommend that a patient can be considered as TAVI-candidate, if life expectancy is more than one year, TAVI would improve his/her quality of life, and the patient has no absolute contraindication

for TAVI (10). Heart teams, therefore, play a crucial role in TAVI-patients' selection after identification of the relative and absolute contraindications for the procedure.

As mentioned earlier, there are different operative-risk groups of AS patients; low, intermediate, high, and extreme surgical risk. Within the same risk group, considerable patient variation may exist based on demographics, comorbidity, and severity of calcification. This fact might complicate clinical judgment and decision-making. In clinical practice and next to the decisive clinical judgments, heart teams use various decision support tools based on prediction models for the perioperative mortality risk estimation in patients with severe AS. These models are called mortality prediction models (MPMs). Accordingly, heart teams can decide if a patient will be a candidate for SAVR or TAVI. If the post-operative risk is estimated to be high, the patient will be considered as TAVI patient.

The role of prediction models as a basis for decision support is more pronounced when the clinical estimation of the risk resides in a grey area. Usually, such cases require much discussion between the different disciplines of the heart teams to decide on the appropriate choice of treatment.

The mortality prediction models used for TAVI patients are either general cardiac surgical models or specific for TAVI-patients. Examples of the most popular surgical models are the European System for Cardiac Operative Risk Evaluation (EuroSCORE-I, 2003; and the newer version EuroSCORE-II, 2012) (14, 20), and the Society of Thoracic Surgeons Predicted Risk of Mortality (STS-PRoM, 2009) model (21). The more recent TAVI-specific early-mortality MPMs include OBSERVANT (13), FRANCE-2 (11), German-AV (22), and American College of Cardiology TAVI (ACC-TAVI) (12). All these models have been developed in different countries using data from local TAVI populations.

Prediction models: development

Predictive modeling is a statistical approach that uses historical data to anticipate future outcomes (23). Over the last decennia, predictive modeling has dramatically evolved from conventional statistic-based modeling to machine learning and data mining-based techniques. Based on the clinical data of a patient a probabilistic prediction model assesses the probability of how likely the patient is to exhibit a specific outcome within a defined time, for example, early (30-day) or 1-year mortality or a major adverse event. Therefore, a prediction model can provide patient risk stratification that helps tailor medical decision-making, which could improve the outcomes and the quality of care.

To build a prediction model, data and a development strategy are needed (24). In the ideal world, data required to build a new prediction model would be prospectively collected based on well-defined criteria, where all the required variables are completely collected to avoid missing data and to standardize the meaning of the variables. However, collecting prospective data is hard and expensive to do, and therefore retrospective datasets are usually used for this purpose even though they often inherit several limitations. These include missing variables altogether, missing values of variables, different definitions of the variables, mislabelling, and lack of auditing. In terms of the model development strategy, there are several methods used to develop (train) prediction models such as regression models (linear, logistic, or Cox), decision trees, neural networks, and many others (24). Most of the existing MPMs -used for TAVI patients are logistic regression-based models and they predict postoperative mortality risk.

During model development, it is often important to select a subset of input variables from the dataset, which contributes to the prediction of the target outcome (dependent variable). This is meant to alleviate the risk of overfitting the data. Deciding on the final set of independent variables is called variable (or feature) selection. This step will reduce the dataset's dimensionality and can, consequently, improve the performance of the model. Features may be pre-selected by clinical experts or based on literature reviews. However, this might enlist too many variables, or miss useful variables that the clinicians overlooked. There are many automatic techniques used for variable selection such as step-wise forward and backward selection, least absolute shrinkage and selection operator (LASSO) regression, and tree-based approaches (24, 25).

Prediction models: validation

Before using a prediction model, it is essential to evaluate it by measuring how accurate the model is in predicting the target outcome (26-28). Evaluation of the predictive performance of a model on data not used during its training is called model validation. A classical used method is to split the dataset randomly into a training (development) set, and a test (validation) set (24). In this approach, developers use their data to internally validate their model and report on the performance of the test set. The final model will still be trained on the whole dataset. However, this method has some drawbacks, specifically if the dataset is small. The result will be too much dependent on the random split. Another commonly used method for internal validation is to perform cross-validation in which the dataset is first partitioned into K (usually 5 or 10) mutually exclusive sets and then trained on all but one dataset, which will function for that fold as a test set. This is repeated K times in which the test set alternates among the K datasets (24, 27). Bootstrapping is another approach for performing internal validation in which bootstrap samples, often hundreds of them, of the whole dataset are taken and the model developmental strategy is repeated on all of them. The difference in the performance of the original model and the mean performance on the bootstrap samples provides an unbiased measure of performance.

Internal validation relies on random resampling of the given dataset. This does not test the model in a new population with potentially different characteristics. The interest may indeed lie in how the model performs in such external datasets. Such external validation can concern the performance of the model from the same setting but prospectively over time. This is called temporal validation in which a model is developed using a dataset of patients over a period of time, and tested on the cohort of a later time period. This method best imitates what the model would face in real-life practice, and helps understand whether and how much a model suffers from changes over time due to population drift (e.g. changes in the patient population over time and changes in the technology and treatment plans). Another form of external validation is evaluating the model's performance in a different setting, for example, a different country. This is important because potential users might need to measure the predictive performance of a model on their population before adopting that model for use (24). For instance, in the TAVI literature, some studies showed that using cardiac surgery MPMs for TAVI patients is inadequate (29). Furthermore, various external validation studies showed that using the existing TAVI-specific MPMs on external populations had poor predictive performance (30). This limits their transportability to other settings.

Prediction models: performance measures

Model performance is measured by performance metrics or measures (27). These measures play a role in selecting the (complexity of) the final model and also in gauging the unbiased predictive model performance that is expected in new patients. These performance measures reflect some performance aspects. One aspect is discrimination and it is often measured by the Area Under Receiver Operating-Characteristic Curve (AU-ROC). Discrimination measures the model's ability to provide higher probabilities of the event (e.g. death) to subjects that had the event (e.g. died) than those who did not have the event. The AU-ROC measure, which is equal to the concordance statistic (c-statistic) (31, 32) is a proportion between 0 and 1. The higher the AU-ROC the better the discrimination.

Another important performance aspect is calibration. Calibration reflects the degree of agreement between the predicted probability of the outcome and the real observed outcome rates. Calibration can be either gauged by inspecting a calibration graph or via a statistical approach. In this thesis, we applied the Cox approach (33) to measure the calibration-intercept and calibration-slope for the prediction models. A model can be considered to have good calibration if the 95% confidence interval of the calibration-intercept includes the value 0, and the 95% confidence interval of the calibration-slope includes the value 1.

Another performance aspect is the predicted probability accuracy, which reflects the measure of distance between the individual predicted probability of a subject and the observed outcome. A measure of the predicted

probability accuracy is the Brier Score, which summarizes the deviation between the observed and predicted outcome at the subject (patient) level by averaging the squared mean difference between the predicted probability and the outcome. The lower the Brier score the better the accuracy. The Brier score includes both elements of discrimination and calibration. To "normalize" the Brier score, the Brier Skill Score (BSS) can be used. It quantifies how much the model improves on an uninformed reference model in which all predicted probabilities are equal to the prevalence of the outcome. The higher the BSS the better (34).

Finally, the interplay between the aspects of positive predictive value (PPV) and sensitivity can be measured by the Precision-Recall Curve. The term precision in information retrieval refers to PPV, and the term recall refers to sensitivity. PPV is the ratio of patients correctly predicted to have the event among all those predicted to have the event. Sensitivity is the proportion of correctly predicted patients in the whole population. The area under the Precision-Recall Curve (AU-PRC) is a measure of the balance between the PPV and sensitivity as it summarizes the trade-off between PPV and sensitivity for different probability thresholds (35). The higher the AU-PRC the better. This measure is especially useful when there is a large class imbalance in the dataset between the class labels (e.g., survivors vs non-survivors).

Why do we want to update prediction models?

Prediction models developed in one country might not be transportable to other countries, for example, due to changes over time in the patient population or because the model will be used in another geographical setting. The usage of prediction models without their adaptation to external datasets could be misleading and may provide suboptimal information for decision-making (29). Therefore, one should consider updating the model before using it on a new population. Updating an existing model can take advantage of the information on the existing models to improve model performance (36-40). In general, a model update can be achieved by several methods. A quite simple approach is adding more data to the original development set and using the same variables to refit the model. This approach might increase the power and the heterogeneity of the development sample. Another approach, which does not require the original dataset, is recalibrating the model on the new population. In this approach, and assuming a logistic regression model, the model's intercept in the linear predictor can be updated alone on the new dataset, or both the intercept and slope can be updated (model-recalibration). A more rigorous approach is updating all the coefficient estimates of the model on the new dataset, this is called model revision. Another approach allows for adding new variables and refitting the model to the new population, this approach is called model extension. This might decrease bias but at the cost of increasing the variance of the model (24).

When do we need to develop a new prediction model?

Developing a new prediction model might be necessary. For instance, if the model update did not satisfactory or did not improve the predictive performance as required. In other words, if the updated model is not transportable to the new setting. Another circumstance is when some variables (from the original model) are not available in the new dataset. Or if the characteristics and settings of the original population differ markedly from the new one. This might apply to our context, TAVI, where the existing MPMs were developed on either cardiac surgical patients or high-risk patients; the used prostheses are different from the currently used devices in our population; and not all the model variables are recorded in our national registry.

Collections of variables and the Strategic Information Management plan (SIM-plan) of a heart center

Aside from TAVI prediction models, this dissertation also addresses the more general topic of strategic information planning in a heart center. Data used for development, or updating a TAVI prediction model are generally obtained from a national registry. These data originate, in turn, from the hospital records, often residing in an electronic patient record (EPR) of various heart centers.

In general, healthcare providers collect and register quantitative and qualitative data about their patients in the (paper) medical patients' records. Data management in healthcare facilities is an ongoing systematic process to gather, analyze and interpret different kinds of information from numerous sources and devices (41).

Over the last few decades, patient data collection has been tremendously developed, especially, after introducing the electronic patient record (EPR) systems (42). These new systems help and facilitate data collection and extraction that can be used in several aspects and for different purposes. These include, among others: 1) offer tailored clinical decision-making, 2) improve quality of care and enhance health outcomes, 3) perform advanced scientific research, 4) develop advanced -personalized- treatment, 5) enhance marketing strategies, reduce healthcare costs, and proper resource allocations, and 6) facilitate efficient communication between healthcare providers and patients (41-46).

A good example of clinical data collection in healthcare facilities is using the collected data to predict a specific outcome after a specific procedure. A case in point is predicting the early-mortality after TAVI, which requires heart teams to use a selected set of variables from the EPR and other systems. These include baseline characteristics of a patient, current and past medical history, and the results of different diagnostic tests such as laboratory results, ECG, echocardiography, and cardiac MRI. This information can be entered into an MPM to give an estimate of the mortality risk after the TAVI procedure.

However, the increased data collection task itself has become a complex task for healthcare facilities and providers. This is especially true after digitizing -almost- all the used diagnostic systems, the vast variety of information sources, and the solo systems. System integration and enhancement of inter-system communication were introduced to alleviate this complexity and reduce the redundancy of the collected data (41, 47). However, the system integration process itself is not easy. A facility needs to properly understand and analyze its current information technology (IT) situation and its future needs. This will help the management team to plan good and achievable goals and objectives to manage their IT systems and resources. Nowadays, this can be achieved by implementing strategic information management planning (41, 48-51). A strategic information management plan (SIM-plan), is a document that provides a wide overview and specific analysis of the current organizational and hospital information system (HIS) situation (41, 48-51). It also provides the strategic organizational and IT goals for the coming five or ten years. Moreover, this document presents the plan to achieve these goals. Usually, the chief information officer (CIO) in the facility is the one who is responsible to develop the SIM-plan, supported by all the healthcare providers in the facility (41). The CIO will conduct the analysis and will provide the possible solutions to plan the future IT situation. Implementing the SIM-plan might provide the organization with a good HIS. A good HIS means a system that is timely accessible to the users, reliable in providing the necessary accurate and complete information, interoperable, highly integrated, allowing no data redundancy with a single recording and multiple usability of data, providing patient-centered information processing, and allowing cost-effective hospital functions (41, 50, 51).

Netherlands heart registration

The National Heart Registry (NHR) collects data on all cardiac interventions performed in the Netherlands including 16 Dutch heart centers. Each year, approximately 75,000 interventions are recorded in the NHR. The Registry follows patients with heart disease through all stages of the treatment process: from the moment of diagnosis up to many years after the intervention. The NHR serves the interests of cardiac patients by promoting quality monitoring and improvement of patient care in the Netherlands from an integrated approach.

Problem statements and research questions

As mentioned above, there are many prediction models used in predicting the mortality of TAVI patients, such as EuroSCORE-I and -II, STS-PRoM, OBSERVANT, FRANCE-2, and ACC-TAVI (11-14, 20, 21). However, these models have not been validated in our Dutch population. Therefore, one of the aims of this thesis is to scrutinize the performance of such models for predicting 30-day early-mortality for TAVI patients in our Dutch population.

Moreover, over the last years, the number of performed TAVI procedures has been markedly increasing, also in the Netherlands. However, as with any other intervention, TAVI has been associated with several complications. We, therefore, aim in this thesis to explore and analyze the incidence and trends of TAVI-related MACE in general, and in specific patient subgroups.

Outline of the thesis

In this thesis, we describe five studies that are designed to address our research questions. The thesis is structured as follows.

In chapter 2, we externally validated and compared the existing TAVI prediction models in predicting early-mortality (30-days) after TAVI. We used a large recent dataset of TAVI-patients from the Netherlands Heart Registration (NHR) and used a comprehensive set of predictive performance measures to evaluate the models.

In chapter 3, we updated the two best performing TAVI-specific models: the FRANCE-2 and ACC-TAVI for predicting the early-mortality. The update method depended on the closed-testing procedure. We performed internal-validation on the updated-models using a recent TAVI-cohort from the NHR. To understand the performance of the updated-models over time, which best mimics the model usage in clinical practice, we also performed temporal-validation in which the models are tested on a dataset collected prospectively after the models have been updated on earlier data.

In chapter 4, we developed and validated a novel TAVI-NHR prediction model for early-mortality after Transcatheter Aortic Valve Implantation (TAVI) based on data from the Netherlands Heart Registration (NHR). In contrast to the two previous chapters, in this chapter additional predictors that were relevant to the model for TAVI patients are eligible for selection.

In chapter 5, we described the recent incidence and trends of TAVI-related MACE. This includes 30-day and 1-year mortality, major vascular bleeding, pacemaker implantation, and stroke. We also investigated TAVI-related MACEs amongst specific patient subgroups based on age groups and operative-risk according to the EuroSCORE-II.

In chapter 6, we shared our experience in developing a strategic information management plan (SIM-plan) in a heart center. To develop this SIM-plan, we followed all the steps outlined in the empirical approach of Brigl et al., called the “Practical-Guideline SIM” (49). This starts by analyzing the organization, then identification of business and IT goals, assessing the current HIS situation, defining the future HIS situation, outlining the roadmap and migration-path, and finally SIM-plan approval and deployment.

Chapter 7 presents a general discussion of the main results of the thesis.

References:

1) Nkomo VT, Gardin JM, Skelton TN, Gottdiener JS, Scott CG, Enriquez-Sarano M. Burden of valvular heart diseases: a population-based study. Lancet. 2006;368(9540):1005-11.
2) Cribier A, Eltchaninoff H, Bash A, Borenstein N, Tron C, Bauer F, et al. Percutaneous transcatheter implantation of an aortic valve prosthesis for calcific aortic stenosis: first human case description. Circulation. 2002;106(24):3006-8.
3) Cribier A, Eltchaninoff H, Tron C, Bauer F, Agatiello C, Sebagh L, et al. Early experience with percutaneous transcatheter implantation of heart valve prosthesis for the treatment of end-stage inoperable patients with calcific aortic stenosis. J Am Coll Cardiol. 2004;43(4):698-703.
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 J Cardiothorac Surg. 2012;42(4):S1-44.
5) de Jaegere P, de Ronde M, den Heijer P, Weger A, Baan J. The history of transcatheter aortic valve implantation: The role and contribution of an early believer and adopter, the Netherlands. Neth Heart J. 2020;28(Suppl 1):128-35.
6) Badheka AO, Patel NJ, Panaich SS, Patel SV, Jhamnani S, Singh V, et al. Effect of Hospital Volume on Outcomes of Transcatheter Aortic Valve Implantation. Am J Cardiol. 2015;116(4):587-94.
7) Carroll JD, Vemulapalli S, Dai D, Matsouaka R, Blackstone E, Edwards F, et al. Procedural Experience for Transcatheter Aortic Valve Replacement and Relation to Outcomes: The STS/ACC TVT Registry. J Am Coll Cardiol. 2017;70(1):29-41.
8) Didier R, Eltchaninoff H, Donzeau-Gouge P, Chevreul K, Fajadet J, Leprince P, et al. Five-Year Clinical Outcome and Valve Durability After Transcatheter Aortic Valve Replacement in High-Risk Patients. Circulation. 2018;138(23):2597-607.
9) Vemulapalli S, Carroll JD, Mack MJ, Li Z, Dai D, Kosinski AS, et al. Procedural Volume and Outcomes for Transcatheter Aortic-Valve Replacement. N Engl J Med. 2019;380(26):2541-50.
10) Baumgartner H, Falk V, Bax JJ, De Bonis M, Hamm C, Holm PJ, et al. 2017 ESC/EACTS Guidelines for the Management of Valvular Heart Disease. Rev Esp Cardiol (Engl Ed). 2018;71(2):110.
11) Iung B, Laouenan 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.
12) 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):46-52.
13) 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.
14) Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, et al. EuroSCORE II. Eur J Cardiothorac Surg. 2012;41(4):734-44; discussion 44-5.
15) Roques F, Michel P, Goldstone AR, Nashef SA. The logistic EuroSCORE. Eur Heart J. 2003;24(9):881-2.
16) Nielsen HH. Transcatheter aortic valve implantation. Dan Med J. 2012;59(12):B4556.
17) Santangelo G, Ielasi A, Pellicano M, Latib A, Tespili M, Donatelli F. An Update on New Generation Transcatheter Aortic Valves and Delivery Systems. J Clin Med. 2022;11(3).
18) Carroll JD. TAVR Prognosis, Aging, and the Second TAVR Tsunami: Insights From France. J Am Coll Cardiol. 2016;68(15):1648-50.
19) Durko AP, Osnabrugge RL, Van Mieghem NM, Milojevic M, Mylotte D, Nkomo VT, et al. Annual number of candidates for transcatheter aortic valve implantation per country: current estimates and future projections. Eur Heart J. 2018;39(28):2635-42.
20) Michel P, Roques F, Nashef SA, Euro SPG. Logistic or additive EuroSCORE for high-risk patients? Eur J Cardiothorac Surg. 2003;23(5):684-7; discussion 7.
21) 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 Suppl):S23-42.
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(5):971-7.
23) Geisser S. Predictive Inference: An Introduction. New York: Chapman and Hall/CRC; 1993. 240 p.
24) Steyerberg E. Clinical Prediction Models, A Practical Approach to Development, Validation, and Updating. New York NY: Springer Science & Business Media, LLC; 2009.
25) Steyerberg EW, Eijkemans MJ, Harrell FE, Jr., Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med. 2000;19(8):1059-79.
26) Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014;35(29):1925-31.
27) Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Stat Med. 2004;23(16):2567-86.
28) Steyerberg EW, Eijkemans MJ, Harrell FE, Jr., Habbema JD. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. Med Decis Making. 2001;21(1):45-56.

29) Martin GP, Sperrin M, Ludman PF, de Belder MA, Gale CP, Toff WD, et al. Inadequacy of existing clinical prediction models for predicting mortality after transcatheter aortic valve implantation. Am Heart J. 2017;184:97-105.
30) 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 Assessment (IRRMA) study. Int J Cardiol. 2016;215:227-31.
31) Pepe MS. An interpretation for the ROC curve and inference using GLM procedures. Biometrics. 2000;56(2):352-9.
32) Pepe MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford: Oxford University Press; 2003 07 October 2004. 320 p.
33) Cox DR. Two further applications of a model for binary regression. Oxford University Press on behalf of Biometrika Trust. 1958;45:562-5 (4 pages).
34) Brier G. Verification of Forecasts Expressed in Terms of Probability. Monthly Weather Review. 1950;78:1-3.
35) Boyd K, Eng KH, Page CD, editors. Area under the Precision-Recall Curve: Point Estimates and Confidence Intervals2013; Berlin, Heidelberg: Springer Berlin Heidelberg.
36) Siregar S, Nieboer D, Versteegh MIM, Steyerberg EW, Takkenberg JJM. Methods for updating a risk prediction model for cardiac surgery: a statistical primer. Interact Cardiovasc Thorac Surg. 2019;28(3):333-8.
37) Vergouwe Y, Nieboer D, Oostenbrink R, Debray TPA, Murray GD, Kattan MW, et al. A closed testing procedure to select an appropriate method for updating prediction models. Stat Med. 2017;36(28):4529-39.
38) Su TL, Jaki T, Hickey GL, Buchan I, Sperrin M. A review of statistical updating methods for clinical prediction models. Stat Methods Med Res. 2018;27(1):185-97.
39) Wilson B, Tran DTT, Dupuis JY, McDonald B. External Validation and Updating of the Cardiac Surgery Score for Prediction of Mortality in a Cardiac Surgery Intensive Care Unit. J Cardiothorac Vasc Anesth. 2019;33(11):3028-34.
40) Lodi-Junqueira L, da Silva JL, Ferreira LR, Goncalves HL, Athayde GR, Gomes TO, et al. In-hospital mortality risk prediction after percutaneous coronary interventions: Validating and updating the Toronto score in Brazil. Catheter Cardiovasc Interv. 2015;86(6):E239-46.
41) Alfred Winter RH, Elske Ammenwerth, Birgit Brigl, Nils Hellrung, Franziska Jahn. Health Information Systems: Architectures and Strategies. 2nd ed. London: Springer Science & Business Media; 2011 oktober 2010. 340 p.
42) Bouayad L, Ialynytchev A, Padmanabhan B. Patient Health Record Systems Scope and Functionalities: Literature Review and Future Directions. J Med Internet Res. 2017;19(11):e388.
43) Jiang F, Liu Y, Hu J, Chen X. Understanding Health Empowerment From the Perspective of Information Processing: Questionnaire Study. J Med Internet Res. 2022;24(1):e27178.
44) Castro EM, Van Regenmortel T, Vanhaecht K, Sermeus W, Van Hecke A. Patient empowerment, patient participation and patient-centeredness in hospital care: A concept analysis based on a literature review. Patient Educ Couns. 2016;99(12):1923-39.
45) Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc. 2006;13(2):121-6.
46) Farmer R, Mathur R, Bhaskaran K, Eastwood SV, Chaturvedi N, Smeeth L. Promises and pitfalls of electronic health record analysis. Diabetologia. 2018;61(6):1241-8.
47) Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. Journal of Big Data. 2019;6(1):54.
48) Brigl B, Hubner-Bloder G, Wendt T, Haux R, Winter A. Architectural quality criteria for hospital information systems. AMIA Annu Symp Proc. 2005:81-5.
49) Brigl B, Ammenwerth E, Dujat C, Graber S, Grosse A, Haber A, et al. Preparing strategic information management plans for hospitals: a practical guideline SIM plans for hospitals: a guideline. Int J Med Inform. 2005;74(1):51-65.
50) Winter A, Brigl B, Buchauer A, Dujat C, Graber S, Hasselbring W, et al. Purpose and structure of strategic plans for information management in hospitals. Stud Health Technol Inform. 2000;77:880-4.
51) Winter AF, Ammenwerth E, Bott OJ, Brigl B, Buchauer A, Graber S, et al. Strategic information management plans: the basis for systematic information management in hospitals. Int J Med Inform. 2001;64(2-3):99-109.

Chapter 2:

External validation of existing prediction models of 30-day mortality after Transcatheter Aortic Valve Implantation (TAVI) in the Netherlands Heart Registration

International Journal of Cardiology, 2020

https://doi.org/10.1016/j.ijcard.2020.05.039

Hatem Al-Farra, Ameen Abu-Hanna, Bas A.J.M. de Mol, Willem Jan P.P. ter Burg, Saskia Houterman, José P.S. Henriques, Anita C.J. Ravelli; on behalf of the NHR THI Registration Committee[#]

Abstract

Background

Several mortality prediction models (MPM) are used for predicting early (30-day) mortality following transcatheter aortic valve implantation (TAVI). Little is known about their predictive performance in external TAVI populations. We aim to externally validate established MPMs on a large TAVI dataset from the Netherlands Heart Registration (NHR).

Methods

We included data from NHR-patients who underwent TAVI during 2013-2017. We calculated the predicted mortalities per MPM. We assessed the predictive performance by discrimination (Area Under Receiver Operating-characteristic Curve, AU-ROC); the Area Under Precision-Recall Curve, AU-PRC; calibration (using calibration-intercept and calibration-slope); Brier Score and Brier Skill Score. We also assessed the predictive performance among subgroups: tertiles of mortality-risk for non-survivors, gender, and access-route.

Results

We included 6,177 TAVI-patients with an observed early-mortality rate of 4.5% (n=280). We applied seven MPMs (STS, EuroSCORE-I, EuroSCORE-II, ACC-TAVI, FRANCE-2, OBSERVANT, and German-AV) on our cohort. The highest AU-ROCs were 0.64 (95%CI 0.61-0.67) for ACC-TAVI and 0.63 (95%CI 0.60-0.67) for FRANCE-2. All MPMs had a very low AU-PRC of ≤0.09. ACC-TAVI had the best calibration-intercept and calibration-slope. Brier Score values ranged between 0.043 and 0.063. Brier Skill Score ranged between -0.47 and 0.004. ACC-TAVI and FRANCE-2 predicted high mortality-risk better than other MPMs. ACC-TAVI outperformed other MPMs in different subgroups.

Conclusion

The ACC-TAVI model has relatively the best predictive performance. However, all models have poor predictive performance. Because of the poor discrimination, miscalibration and limited accuracy of the models there is a need to update the existing models or develop new TAVI-specific models for local populations.

Introduction

For a long time, surgical aortic valve replacement (SAVR) was the standard treatment for severe aortic valve stenosis. Patients with high mortality-risk were considered ineligible for SAVR and were treated medically. Transcatheter aortic valve implantation (TAVI) has emerged as an alternative procedure for medical treatment for those groups of patients (1). Patients' selection for SAVR or TAVI depends on proper identification of the post-procedural mortality-risk. In practice, heart teams use mortality prediction models (MPM) to support their decisions on patients' selection. Many MPMs have been developed for cardiac procedures for patient selection, risk stratification and benchmarking. The European System for Cardiac Operative Risk Evaluation (EuroSCORE-I, 2003; and the newer version EuroSCORE-II, 2012) (2-4), and the Society of Thoracic Surgeons Predicted Risk of Mortality (STS-PRoM, 2009) model (5), have been widely used as MPMs for early (30-day) mortality after cardiac surgery. These cardiac surgery MPMs have been also used for TAVI-patients. The guidelines for the management of valvular heart disease (version 2012) has suggested that high mortality-risk estimates of EuroSCORE-I ≥ 20% and/or STS-PRoM > 10% may serve as an appropriate indication for TAVI instead of SAVR (6). However, EuroSCORE-I, -II and STS-PRoM (STS) were developed and internally validated for predicting early-mortality on standard cardiac-surgery patients, and not specifically for TAVI-patients.

These MPMs are, therefore, missing essential TAVI-specific pre-procedural variables like access-route, balloon aortic valvuloplasty prior to TAVI and valve-type. Some studies reported that EuroSCORE overestimated the early-mortality probability after TAVI by 8% (7-10). Both EuroSCORE-II and STS have been reported to have poor discrimination for predicting early-mortality after TAVI (with AU-ROC of 0.66 and 0.58, respectively) (11). Also, other external-validation studies have reported their suboptimal predictive-performance (with poor AU-ROC and miscalibration) for TAVI early-mortality (12, 13).

Over time, TAVI-specific early-mortality MPMs have been developed, such as OBSERVANT (14), FRANCE-2 (15), and American College of Cardiology TAVI (ACC-TAVI) (16). The predictive performance of the TAVI-specific and the cardiac-surgery MPMs were externally validated in some studies (17-20).

In the IRRMA study (19), the TAVI-specific MPMs did not perform better than the cardiac surgery MPMs (had poor AU-ROCs and were miscalibrated). Contrariwise, TAVI-specific MPMs (ACC-TAVI and FRANCE-2) outperformed the other MPMs including the cardiac-surgery MPMs in the UK-study, although ACC-TAVI and FRANCE-2 had suboptimal predictive performance (both miscalibrated and had poor AU-ROC of 0.64 and 0.62, respectively) (20).

The most commonly used predictive performance measures in these external validation studies were discrimination (AU-ROC) and calibration (21). Besides these predictive performance measures, there are other measures like Area under the precision-recall curve (AU-PRC) and the Brier Skill Score (22-30) that provide additional aspects on the predictive performance, to better understand the MPMs' predictive behaviour.

The evidence about the external validity of the MPMs is limited and has not been investigated for TAVI-patients in the Netherlands. Therefore, we aimed in this study to externally validate and compare the existing MPMs in predicting early-mortality (30-days) after TAVI, using a large recent local dataset of TAVI-patients from the Netherlands Heart Registration (NHR) and deploy additional predictive performance measures.

Methods

Study design

This is a retrospective cohort study in which we used data from the Netherlands Heart Registration (NHR). Hence, instead of developing new models, we applied a set of currently used MPMs for external validation on our dataset. The study was approved by the institutional review board of the Catharina Hospital (Approval number: 2018-004). The used anonymized data conformed to the Declaration of Helsinki principles.

Selection of mortality prediction models (MPMs)

For this study we selected relevant MPMs by literature search on PubMed for published studies up to 2018. Using any of the following terms: TAVI, SAVR, mortality, early mortality, 30-day mortality, in-hospital mortality, clinical

prediction models, mortality prediction model, risk score, risk stratification with any of the following terms: performance measures, discrimination, or calibration. We also searched using the following Mesh-terms: aortic valve stenosis, transcatheter aortic valve replacement, TAVR, and ROC Curve.

An MPM was considered if it was published, internally validated, and used for early (30-day) mortality. MPMs with other end-points (long-term mortality) were not included in this study.

Definition of the primary outcome variable

The primary outcome variable of this study is the early post-procedural mortality, which we define as death within 30-days from the date of the TAVI procedure.

The Netherlands Heart Registration (NHR)

In the Netherlands, 16 heart centres perform TAVI for symptomatic aortic stenosis. Multi-disciplinary teams of cardiologists, surgeons and other healthcare professionals at each center decide on patients' eligibility for operation: SAVR or TAVI procedure. Data were extracted from the value-based healthcare (VBHC) program, which is a part of the Netherlands Heart Registration (NHR).

In the VBHC program, which focuses on measuring and improving outcomes that matter most to patients, 22 Dutch heart canters voluntarily submit patient demographics, clinical characteristics, intervention risk factors, procedural details, mortality-status, complications and follow-up after hospital discharge (31).

In total, 13 out of 16 Dutch heart canters participated in presenting the outcomes of TAVI. Each center obtained the mortality data from the regional municipal administration registry. For this study, all data on each TAVI-procedure from January 1, 2013, to December 31, 2017 (NHR-TAVI cohort) of these 13 centres were extracted. For each patient, to be included in this study, the outcome status (early-mortality) should be available.

To obtain reliable data, the NHR has an advanced, certified data-quality control system in place, and an audit was completed by the NHR on TAVI patient characteristics and outcomes in 2017. During that audit, NHR has examined a sample of 50 medical files among the participating centres.

Statistical analysis

For each selected MPM, the known and corresponding variables from the NHR-TAVI cohort were selected (e-component Table 2 presents the variables used from the NHR Registration to externally validate the candidate MPMs).

In the few cases in which a variable required by a model was not registered in the NHR-TAVI registration, the condition represented by the missing variable was assumed to be absent for all patients. This could theoretically induce a bias, though the same issue of non-registered variables had been described in previous external validation studies with a reported negligible bias (19, 20).

For missing values of variables registered in the NHR-TAVI cohort, we assumed there were missing at random. Therefore, multiple imputations with ten imputed datasets were applied for the missing values using Multiple Imputation by Chained Equations (MICE).

For each patient, the early-mortality probabilities obtained from the 10 imputed datasets were averaged. For each selected MPM, we used its logistic regression equation to predict early-mortality probability. In the equations, we used the regression coefficients as published in the original studies about the MPMs.

Predictive performance estimation

We used the following predictive performance aspects and their respective measures: discrimination by the Area Under Receiver Operating-Characteristic Curve (AU-ROC); the balance between the positive predictive value and the sensitivity by the Area Under Precision-Recall Curve (AU-PRC); calibration by the calibration-slope and -intercept; and accuracy by Brier Score and Brier Skill Score (BSS).

Discrimination measures the ability of the MPM to distinguish between survivors and non-survivors. It is quantified by the AU-ROC and is also equal to the concordance statistic (c-statistic) (24, 32). The closer the AU-ROC is to 1, the better the MPM is.

We compared AU-ROCs of various MPMs using the non-parametric method of Delong et al. (33). Furthermore, because some variables were imputed, the AU-ROCs of the MPMs were compared before and after imputation using the method described by Venkatraman (34).

The AU-PRC summarizes the trade-off between the precision and the recall for each MPM using different probability thresholds (35). The terms "recall" and "precision", originating from the discipline of Information Retrieval, correspond respectively to the sensitivity and the positive predictive value. AU-PRC evaluates the fraction of true positives among the positive predictions. In a dataset where the prevalence of the event is low (imbalanced dataset), the AU-ROC does not provide insight into the balance between the recall and the precision (30, 36, 37). Therefore, besides the AU-ROC, we also obtain AU-PRC. The closer the AU-PRC is to 1, the better the MPM is.

Calibration is the agreement between predicted and observed mortality rates across the full probability range. To assess calibration, we used the calibration approach formulated by Cox (38). In this approach, an existing MPM is first used to obtain the predicted log-odds of early mortality on our external cohort. Then, using a separate logistic regression model, these log-odds themselves are used as the sole predictor of (again) early-mortality. If the original probabilities based on the existing MPM were perfect, and hence the log-odds, then the coefficients of in the linear predictor of this logistic regression model would be 0 for the intercept and 1 for the slope. Specifically, the two coefficients correspond to 1) the calibration-intercept (Calibration-in-the-large), which indicates the extent that predictions are systematically too low or too high, and 2) the calibration-slope (regression slope of the linear predictor). Good calibration is observed if the 95% confidence interval (CI) for the calibration-intercept includes 0, and the 95% CI of the calibration-slope includes 1.

For measuring the accuracy, we use the Brier Score and Brier Skill Score (BSS), which summarize the deviations between the outcome and predicted probabilities at the patient level. Lower Brier Score and higher BSS indicate better accuracy. The Brier Score is the mean of the squared error and ranges between zero (perfect prediction) and one (the worst prediction) (25). For better interpretation, the Brier Score is transformed into the BSS. The BSS measures the proportional improvement of each model's predictions over a non-informative reference MPM that simply predicts the prior probability of the event for all patients. The maximum value for BSS is 1, which indicates a perfect deterministic prediction i.e., the model could exactly predict the observed outcomes (39). A BSS of zero means that there is no improvement compared to the predictions of the reference model. A negative BSS indicates poorer performance than the reference non-informative MPM.

For subgroup analysis in each MPM, we defined high, moderate and low mortality-risk subgroups based on the 33% and 66% probability tertiles for the non-survivors patients. The high, moderate, low subgroups of each MPM were plotted in a 100% stacked-column bar chart and compared. A good MPM would predict and allocate more cases from the non-survivors as high mortality-risk cases.

Another subgroup analysis was conducted on different subgroups defined by: age (≤75 and >75), gender (female), diabetes (yes and no), access-route (transfemoral and non-transfemoral), left-ventricular-ejection-fraction (LVEF) (<50% and ≥50%), NYHA (class-III and class-IV), and procedure-urgency (urgent, emergency and salvage).

In addition, for each MPM, we provide the density plots of the mortality probabilities for survivors and non-survivors. This chart is a variation on the histogram in which kernel smoothing is used for the plotting. A perfectly discriminating MPM will have non-overlapping density curves for survivors and non-survivors.

We use the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement for reporting (40) (E-component material; TRIPOD Checklist).

All statistical analyses were done in R software (version 3.5.1). Multiple imputations of the dataset were completed using the MICE package in R. The graphical plots were made using the ggplot2 package.

The package pROC was used for constructing and testing the AU-ROCs, and the package PRROC to construct AU-PRC. A 2-tailed p-value <0.05 was considered significant for all analyses.

Results

We found seven relevant MPMs, which is currently used for post-procedural in-hospital and early (30-day) mortality after TAVI (E-component Table 1). These MPMs were both internally validated in (3-5, 14-16, 41) and externally validated in (17-20) (E-component Table 4).

Generally, there were three types of MPMs used for predicting early-mortality after TAVI. The first type includes cardiac surgery MPMs that were developed on standard cardiac-surgery patients. These are EuroSCORE-I, EuroSCORE-II, and STS-PRoM, with 17, 19, and 41 variables in each model, respectively. The second type includes the TAVI-specific MPMs consisting of ACC-TAVI, FRANCE-2, and OBSERVANT, with 9, 10 and 7 variables in each model, respectively. The third type includes the MPMs developed on TAVI and SAVR patients. In this category, there was one MPM, the German-AV, with 16 variables. The originally reported internally validated AU-ROCs for these MPMs ranged between 0.59 (FRANCE-2) and 0.81 (EuroSCORE-II) (E-component Table 1). Data about 7,319 patients from the NHR-TAVI registration, were obtained for this study. We exclude 1,142 patients due to missing outcome mortality-status.

For this study, we obtained data of 7,319 patients from the NHR-TAVI registration, to be used as an external validation dataset. We excluded 1,142 patients due to missing outcome mortality-status. We included data of 6,177 patients, with a 4.5% (n=280) early-mortality rate. Table 1 presents summary statistics for the baseline and the procedural characteristics of the external-validation dataset of this study (our TAVI-NHR-patients). The mean age was 80.0 years (S.D. 7.0), 51.0% of the patients were female and 56.0% had NYHA class-III and 8% NYHA class-IV. About 37.0% of the patients had an LVEF <50.0%, and 9.0% from the TAVI-procedures were Urgent. Patients with critical-preoperative-state had the highest early-mortality risk of 21.1%, Dialysis with 9.0% early-mortality risk, NYHA class-IV with 9.4%, access-route (non-transfemoral) with 8.3%, and procedure-urgency (urgent) with 6.7%.

Table 1. Baseline characteristics of the population and early mortality in the TAVI-NHR cohort before implementing multiple imputation.

Variables	Total population (Total = 6,177)		Patients with early (30-day) mortality (Total = 280, early-mortality rate 4.5%)	
	Mean / Number (n)	S.D. / %	Mean / Number (n)	S.D. / %
Age (years)	80.0	± 6.9	80.9	± 6.6
Height (cm)	168.1	± 9.4	166.6	± 10.4
Weight (kg)	77.0	± 15.3	73.2	± 14.4
Body mass index (kg/m^2)	27.2	± 4.9	26.4	± 5.3
EuroSCORE I	16.3	± 10.5	19.6	± 13
EuroSCORE-II	6.1	± 5.7	7.8	± 6.8
Creatinine, µmol/L	108.2	± 69.2	115.0	± 67.9
eGFR mL/min/1.73 m^2	59.1	± 21.3	56.0	± 22.1
LVEF	50.2	± 11.2	48.2	± 11.8
sPAP mm Hg	31.1	± 10.9	33.7	± 13.1
sPAP >60 mmHg	86	1.4	6	2.1
LVEF <50%	2273	36.8	126	45.0
Female gender	3170	51.3	147	52.5
Chronic kidney disease	2764	44.7	112	40.0
Dialysis	87	1.5	8	2.9
Diabetes				
Diabetes, oral medication	789	13.2	29	10.4
Diabetes, insulin	420	7.0	17	6.1
Poor mobility	333	9.2	11	3.9
Chronic lung disease	1377	22.4	76	27.1

Extra-cardiac arteriopathy	1414	23.1	80	28.6
Previous cardiac surgery	1323	22.3	54	19.3
Recent myocardial infarction	119	2.0	9	3.2
Functional NYHA class				
Functional NYHA Class III	2991	56.1	140	50.0
Functional NYHA Class IV	405	7.6	38	13.6
Critical preoperative state	38	0.6	8	2.9
Procedure urgency				
Procedure urgency Elective	5415	90.8	215	76.8
Procedure urgency Urgent	536	9.0	41	14.6
Procedure urgency Emergency	15	0.3	1	0.4
Procedure weight (2 operations)	57	1.0	3	1.1
Anaesthesia	3671	62.9	202	72.1
Access route				
Access route Transfemoral	4926	80.7	182	65.0
Access route Non-transfemoral	1163	19.1	96	34.3
Balloon pre-TAVI	2738	51.8	118	42.1

Values are mean ± standard deviation (S.D.) or number (n) and percentage (%).
Abbreviations: eGFR = estimated glomerular filtration rate; sPAP = systolic Pulmonary Arterial Pressure; LVEF = Left Ventricular Ejection Fraction; NYHA = New York Heart Association functional Classification; Balloon pre-TAVI = Balloon aortic valvuloplasty prior to date of TAVI.

In the NHR-TAVI cohort, the variables hypertension and atrial-fibrillation (used in STS) were not registered in the NHR-TAVI registration. Also, the variable acute-pulmonary-oedema (used in FRANCE-2) was not registered (E-component Tables 2 and 3). Therefore, these variables were assumed to be absent for all patients. Some of the MPMs' variables had missing values. Most variables (total 16 variables) had <2.0% missing values and 6 variables had more than 5.0% missing values (NYHA class-III, class-IV, poor-mobility, and diabetes).

Details about the percentage of missing values are presented in E-component Table 3. These missing values were completed with multiple imputations. The AU-ROCs of all MPMs remained similar before and after imputations (E-component Table 5). ACC-TAVI with a predicted early-mortality of 4.4% came closest to the observed mortality (4.5%) in the NHR-TAVI cohort (Table 2).

The predicted early-mortalities of the MPMs ranged from 3.4% (underestimation) for STS to 16.2% for EuroSCORE-I, which indicates an overestimation of the early-mortality risk.

Table 2. Predicted early-mortality, discrimination (AU-ROC, (SD) 95% CI), area under the precision-recall curve (AU-PRC), calibration-intercept (95% CI), calibration-slope (95% CI), Brier score, and Brier skill score for each MPM in the whole NHR-TAVI cohort (N=6177).

Model (MPM)	Predicted early-mortality in NHR-TAVI	Discrimination AU-ROC (SD) 95% CI	AU-PRC	Calibration		Accuracy	
				Calibration-intercept (95% CI)[#]	Calibration-slope (95% CI)[#]	Brier score	Brier skill score
Surgical MPM							
STS	3.4%	0.62 (0.018) 0.58-0.65	0.08	0.31 (0.19 - 0.43)	0.90 (0.86-0.94)	0.043	0.004
EuroSCORE-I	16.2%	0.59 (0.018) 0.55-0.62	0.07	-1.49 (-1.61 - -1.37)	1.76 (1.68-1.84)	0.063	-0.47
EuroSCORE-II	5.5%	0.61 (0.017) 0.57-0.64	0.07	-0.21 (-0.34 - -0.09)	1.02 (0.98-1.07)	0.044	-0.03
TAVI-specific MPM							
ACC-TAVI	4.4%	0.64 (0.017) 0.61-0.67	0.09	0.04 (-0.08 - 0.16)	0.98 (0.94-1.01)	0.043	0.002
FRANCE-2	7.4%	0.63 (0.017) 0.60-0.67	0.09	-0.53 (-0.66 - -0.41)	1.21 (1.16-1.26)	0.044	-0.01
OBSERVANT	6.5%	0.58 (0.018) 0.55-0.62	0.08	-0.39 (-0.51 - -0.27)	1.11 (1.06-1.16)	0.044	-0.02
SAVR and TAVI MPM							
German-AV	9.0%	0.60 (0.018) 0.57-0.64	0.08	-0.76 (-0.88 - -0.64)	1.30 (1.25-1.36)	0.047	-0.09

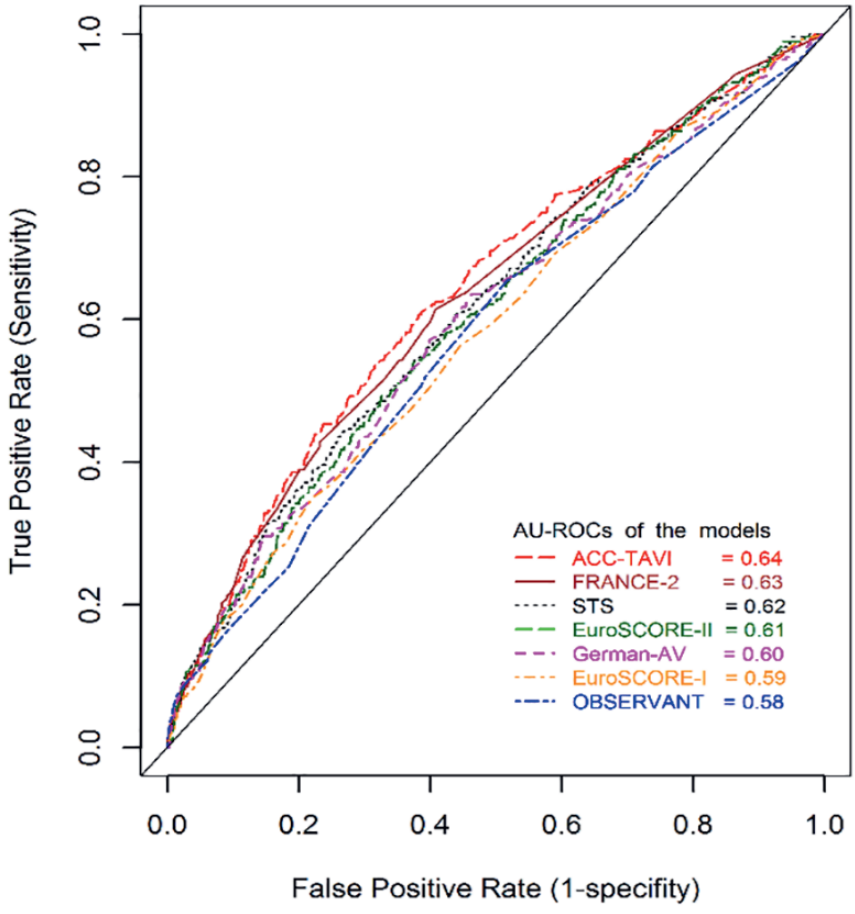
[#] Calibration-intercepts and -slopes for each model were estimated assuming the slope(s) and intercept(s) equal to one and zero respectively. A satisfactory calibration considered if the 95%CI for the calibration-intercept and-slope included the zero and one, respectively. Bold items represent having the best predictive-performance among the other models.

Abbreviations: MPM = Mortality prediction model, SD = standard deviation, AU-ROC = area under the receiver operating characteristic curve = concordance *(c) statistic*; AU-PRC = area under precision-recall curve; SAVR = Surgical Aortic Valve Replacement.

The AU-ROCs ranged between 0.64 (95%CI 0.61-0.67) for ACC-TAVI to 0.58 (95%CI 0.55-0.62) for OBSERVANT (Table 2), and the highest and lowest AU-ROCs differed significantly (p-value = 0.007). FRANCE-2 had the second-highest discriminative ability with AU-ROC of 0.63 (95%CI 0.60-0.67) (Table 2 and Figure 1).

There was no difference between ACC-TAVI and FRANCE-2 (p-value = 0.54). There was no significant statistical difference between AU-ROCs of each MPM in the entire cohort before and after imputation (E-component Table 5).

Figure 1. The area under receiver operative curve (AU-ROC) for each of the mortality prediction models in the TAVI-NHR cohort

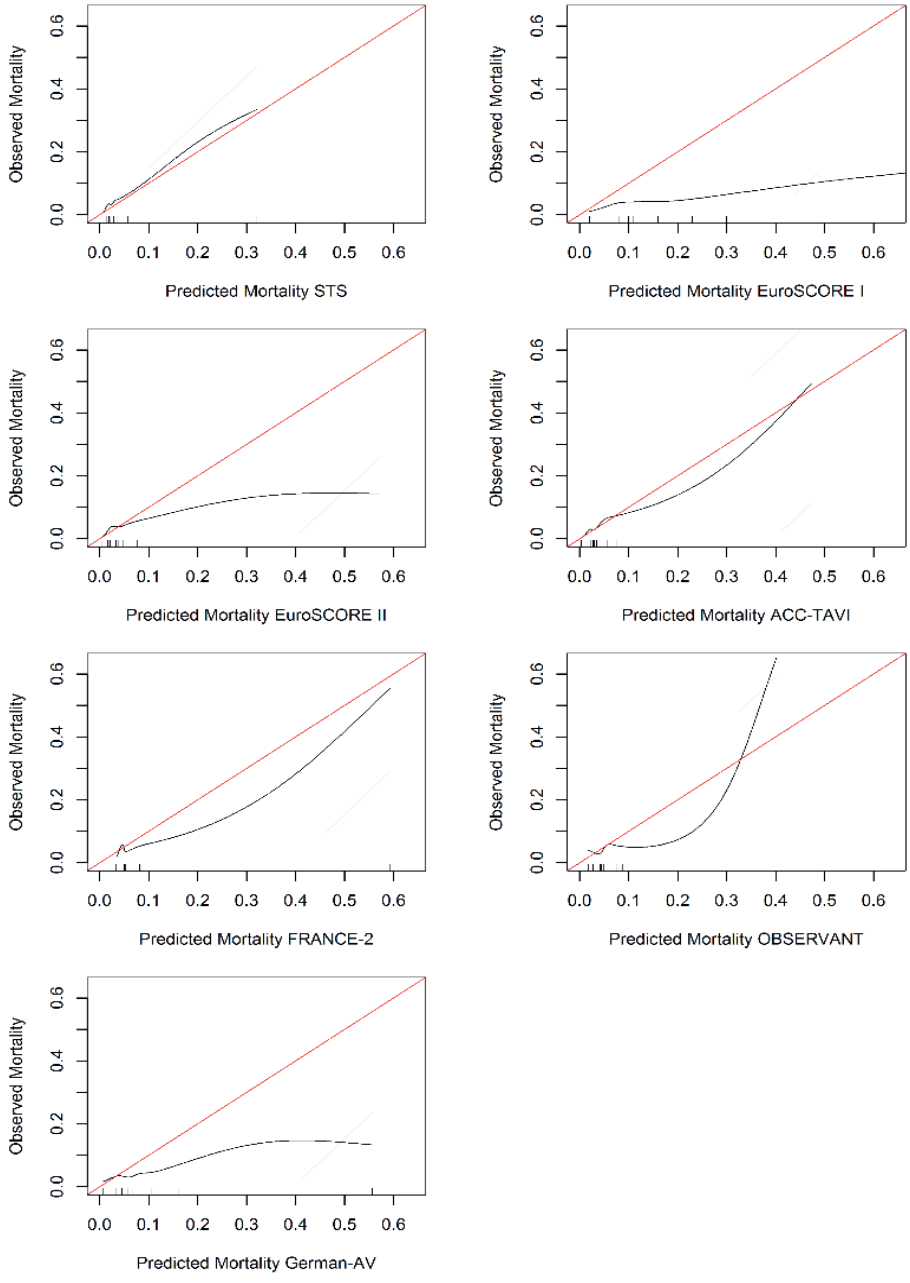


Abbreviations: AU-ROC = area under the receiver operating characteristic curve = concordance *(c) statistic*

For AU-PRC (trade-off between positive predictive value and sensitivity), both ACC-TAVI and FRANCE-2 had the highest AU-PRC values of 0.09.

Only for the model ACC-TAVI, the calibration-intercept 0.04 (95%CI -0.08 - 0.16) and calibration-slope 0.98 (95%CI 0.94 - 1.01) did not significantly deviated from their ideal values (Table 2 and Figure 2).

Figure 2. Calibration plots showing the predicted vs. observed early-mortality for the mortality prediction models in the NHR-TAVI cohort.



The diagonal red-line represents the perfect calibration for a perfect model (the predicted early-mortality being equal to the observed early-mortality). The black-line in each graph represents the calibration of each MPM. If the black-line is above the red-line (see STS graph), then the predicted early-mortality is lower than the observed early-mortality (i.e., underestimation). EuroSCORE I and II, FRANCE-2, German-AV overestimate the early-mortality; note the predicted early-mortality is consistently higher than the observed mortality. OBSERVANT overestimated the early-mortality in the low-risk (range x-axis from 0 – 0.33) and underestimated it for the high-risk cases (range > 0.33). ACC-TAVI overestimated early mortality, but with the best calibration-on-the-large (calibration intercept) and calibration–slope (see Table 3). Despite the high density of cases in the lower range of predicted mortality, 99% of the patients have predicted values in the depicted ranges of the x-axis.

In terms of accuracy, the Brier Score values were very low (<0.05) and similar for most of the MPMs of 0.04 except for EuroSCORE-I that had the worst Brier Score of 0.06. The BSS of ACC-TAVI is 0.002 and STS is 0.004 (Table 2). ACC-TAVI and FRANCE-2 predicted the high mortality-risk subgroup (among non-survivors) better than other MPMs. However, FRANCE-2 poorly classified moderate/low-risk subgroups (E-component Figure 1).

The ACC-TAVI had the best performance (in terms of AU-ROC, AU-PRC, calibration, and accuracy) among the subgroups (Age, Gender, Diabetes, Access-Route, LVEF, NYHA-classes, and Procedure-Urgency) (E-component Table 6).

To better explain the distribution of mortality probabilities of each MPM, we graphically constructed density plots. As shown in (E-component Figure 2), the curves for survivors and non-survivors overlapped on virtually all the probability range.

Discussion

This study showed that ACC-TAVI has relatively the best performance for predicting early-mortality in our TAVI-patients. However, the predictive performance of all validated MPMs in this study appears to be suboptimal. Hence these MPMs are unlikely to be useful for individual and personalized TAVI-mortality risk prediction outside their original populations. Therefore, their applicability in the clinical practice (for patient selection, shared decision-making or benchmarking) is questionable in the Netherlands, and possibly in other external populations.

This study showed that the ACC-TAVI and FRANCE-2 models have the highest AU-ROC of 0.64 and 0.63, respectively, which is comparable to the AU-ROC findings of a previous external validation (UK-study) (20) (E-component Table 4). However, an AU-ROC between 0.6 and 0.7 is often regarded as poor. The originally reported AU-ROC was 0.66 for ACC-TAVI and 0.59 for FRANCE-2 (15, 16).

ACC-TAVI is the only model in our study that had good calibration. This finding supports previous findings (20, 42, 43). This balanced performance might be due to the similarities between the populations in these external validation studies and the development population.

ACC-TAVI and FRANCE-2 have the highest Area Under Precision-Recall Curve AU-PRC of about 0.1, but models with such low AU-PRC value are considered inadequate and have poor performance (30, 37). We could not find previous publications reporting on the AU-PRC values for ACC-TAVI and FRANCE-2.

However, we believe the low AU-PRC obtained in the external validation is related to the generally low prevalence of the outcome measure in the TAVI patient population and the fact that the model does have a very good discrimination ability. Outcome prevalence is hence associated with a low positive predictive value and hence a low AU-PRC.

The BSS values ranged between -0.47 and 0.004. Both ACC-TAVI and STS have BSS of just above zero, meaning there is no marked prediction improvement compared to the non-informative reference model. For the other MPMs, the BSS had negative values, indicating predictions are even poorer than the reference model (Table 2).

When analysing the mortality-risk subgroups in early-mortality cases, the model FRANCE-2 classified 36% (102/280) of deaths as low mortality-risk patients. In contrast, ACC-TAVI classified (less and hence better) 21% (58/280) of deaths as low mortality-risk (E-component Figure 1). This difference, which is in favour of ACC-TAVI, is due to the ability of ACC-TAVI to predict more cases of high mortality-risk from the deaths. This is likely due to the presence of three variables Acuity-Category (2, 3 and 4) in the model ACC-TAVI, but not in FRANCE-2.

These variables correspond directly or by a combination of the NHR-cohort variables Procedure-Urgency (Urgent, Emergent, and Salvage), Critical-Preoperative-Status and Recent-Myocardial-Infarction. In FRANCE-2, only Critical-Preoperative-Status is used for the prediction. Therefore, it seems that FRANCE-2 ignores some potential mortality variables in the NHR-TAVI-cohort.

The density plots of probabilities for survivors and non-survivors show large overlap (E-component Figure 2), indicating the poor ability of any MPM to separate survivors from non-survivors in our population.

Strengths and limitations

The main strength of this study is the large sample size. This analysis is based on the contemporary and largest TAVI-population in the Netherlands. Nearly all heart centres in the Netherlands provided data on TAVI patients. To the best of our knowledge, this is the first study in the Netherlands that externally validated and compared the predictive-performance of seven existing MPMs on a TAVI-cohort. Besides, unlike earlier studies (17-20), we deployed additional predictive performance measures (area under precision-recall curve and Brier Skill Score).

A limitation of this study is that not all variables in the MPMs were registered in the NHR-TAVI registration for TAVI-patients. In E-component Table 2 it is visible that some variables of the STS and FRANCE-2 models are missing in the NHR registration. However, in line with other studies (19, 20) we assumed that the underlying conditions (e.g., acute-pulmonary-oedema) were absent for all patients for the corresponding missing variables.

In addition, we performed the analysis for FRANCE-2, one of the best performing MPMS, in which we simulated the values of the acute-pulmonary-oedema variable (the only variable missing for FRANCE-2 in our NHR dataset). In each simulation, we have randomly drawn values, with a probability of 0.5 of each outcome (absent/present) and calculated the performance measures. The performance estimates and their confidence intervals were essentially the same.

Another possible limitation is the missing values of some variables (E-component Table 3). However, we implemented multiple imputations to attenuate this limitation. Missing values and multiple imputations might introduce biases. Therefore, we calculated AU-ROCs of all MPMs before and after imputations, which remained unaffected (E-component Table 5).

The implication for future work

Cardiac surgery MPMs are used routinely to justify the indication for TAVI in high mortality-risk patients. Moreover, they are used for TAVI quality control and benchmarking. Our study showed that these MPMs have poor discrimination, miscalibration and overestimated TAVI-related early-mortality; hence, their use in patient selection, quality control and benchmarking is questionable.

In this study, ACC-TAVI and FRANCE-2 emerged as the best two performing MPMs in our cohort. However, they still seem relatively poor for predicting TAVI early-mortality outside their original populations. Using a univariate analysis, we found potential predictors that are not part of the set of variables in the two best performing MPMs (ACC-TAVI and FRANCE-2). Those variables are general anaesthesia (no/yes), body surface area (m2), diabetes on insulin (no/yes), LVEF (no/yes), peripheral artery disease (no/yes), age, and chronic pulmonary disease (no/yes). Including these models in a new TAVI prediction model could possibly improve the models for TAVI patients.

A new TAVI-specific MPM with better predictive performance is therefore required in order to stratify patients into high as well as moderate and low mortality risk subgroups. This is especially important as TAVI-procedures are becoming the standard therapy rather than conventional surgery. Until a new or updated TAVI-specific MPM will be available, we encourage participating heart centres in the Netherlands to enhance the data registry.

Conclusion

This external validation study showed that there are large differences between the ability of the MPMs to predict early-mortality after TAVI. The ACC-TAVI model has relatively the best predictive performance. However, all studied models had poor predictive performance. Because of the poor discrimination, poor calibration and the limited accuracy of the current models, their use in clinical practice and benchmarking, at least in the Netherlands and likely in other cohorts, is questionable. This study unveiled the unmet need for developing and validation of an appropriate TAVI-specific MPM.

Declaration of Competing Interest

Hatem Al-Farra, Ameen Abu-Hanna, Bas AJM de Mol, Willem Jan ter Burg, Saskia Houterman, José PS Henriques, and Anita CJ Ravelli declare no conflict of interest.

Addendum

The following physicians are the members of the NHR THI Registration Committee. They represent the hospitals that have provided the data for this study. Contact with NHR THI Registration Committee can be via the e-mail: info@nederlandsehartregistratie.nl

[#] The NHR THI Registration Committee:

M.M. Vis, Amsterdam University Medical Centers
J. Vos, Amphia Hospital
J. ten Berg, St. Antonius Hospital
W.A.L. Tonino, Catharina Hospital
C.E. Schotborgh, HagaHospital
V. Roolvink, Isala
F. Porta, Leeuwarden Medical Center
M. Stoel, Medisch Spectrum Twente
S. Kats, Maastricht University Medical Center
G. Amoroso, Onze Lieve Vrouwe Gasthuis
H.W. van der Werf, University Medical Center Groningen
P.R. Stella, University Medical Center Utrecht
P. de Jaegere, Erasmus University Medical Center

Supplementary materials (e-components)

E-components Table 1. The selected mortality prediction models, original mortality rate, discrimination (AU-ROC or C-statistics) for early-mortality and mortality risk threshold for each model

Model (MPM)	Year	Number of variables per model [a]	Primary mortality end-point	Number of patients (derivation: validation)	Original mortality rate	Originally-derived AU-ROC	Originally-validated AU-ROC	Threshold (risk cut-off point)
Surgical MPM								
STS [5]	2008	41	30-Days	109,759 (3:2)	3.4%	0.76	0.77	10.0%
EuroSCORE-I [2]	2003	17	In-hospital	19,030 (NA)	4.8%	0.79	0.68	20.0%
EuroSCORE-II [3]	2012	19	In-hospital	22,381 (3:1)	4.6%	0.80	0.81	8.0%
TAVI-specific MPM								
ACC-TAVI [16]	2016	9	In-hospital	20,586 (2:1)	5.3%	0.67	0.66	N.A.
FRANCE-2 [15]	2014	10	30-Days	3,833 (2:1)	10.0%	0.67	0.59	N.A.
OBSERVANT [14]	2014	7	30-Days	1,878 (2:1)	6.07%	0.73	0.71	6.0%
SAVR and TAVI MPM								
German-AV [41]	2013	16	In-hospital	11,794	3.7%	0.81	N.A.	N.A.

Abbreviations: AU-ROC = Area under the receiver operating characteristic curve; STS (STS-PRoM) = The Society of Thoracic Surgeons Predicted Risk of Mortality; EuroSCORE = European System for Cardiac Operative Risk Evaluation; ACC-TAVI= American College of Cardiology; TAVI = Transcatheter Aortic Valve Implantation; N.A. = not applicable, or not available; FRANCE-2=French Aortic National CoreValve and Edwards registry; OBSERVANT = Observational Study Of Appropriateness, Efficacy And Effectiveness of AVR-TAVR Procedures for the Treatment Of Severe Symptomatic Aortic Stenosis; AV = aortic valve; SAVR = Surgical Aortic Valve Replacement

[a] The variables for each model are described in supplementary table 2.

E-components Table 2. Presentation of variables in the seven selected mortality prediction models (MPM) and the status if the variable available in the Netherlands Heart Registration

Variables	STS	EuroSCORE-I	EuroSCORE-II	ACC-TAVI	FRANCE-2	OBSERVANT	German-AV	NHR
Patient characteristics								
Age (years)	✓	✓	✓	✓	✓		✓	✓
Sex	✓	✓	✓				✓	✓
Height	✓							✓
Weight	✓				✓			✓
Body mass index					✓		✓	✓
Body surface area	✓							NA
Ethnicity	✓							NA
Co-morbid conditions								
Creatinine clearance			✓			✓		✓
Serum creatinine	✓	✓						✓
Dialysis	✓				✓		✓	✓
eGFR				✓				✓
Diabetes	✓		✓			✓		✓
Hypertension	✓							NA
Chronic lung disease	✓	✓	✓	✓	✓		✓	✓
Extra-cardiac arteriopathy	✓	✓	✓				✓	✓
Peripheral vascular disease	✓							✓
Cerebrovascular accident	✓							✓
Neurological dysfunction	✓	✓						✓
Poor mobility			✓					✓
Immunosuppressive therapy	✓							NA
Cardiac history								
NYHA classifications	✓		✓	✓	✓	✓	✓	✓
CCS class IV angina			✓					✓
Unstable angina	✓	✓						✓
Recent myocardial infarction	✓	✓	✓				✓	✓
Atrial fibrillation	✓							NA
Previous valvular disease	✓							✓
Type and severity of valvular disease		✓						NA
Previous CABG	✓							✓
Previous cardiac surgery		✓	✓				✓	✓
Previous PCI	✓							✓
Balloon pre-TAVI						✓		✓
Number of diseased coronary vessels	✓							NA
Active endocarditis	✓	✓	✓				✓	✓
Arrhythmias	✓							✓
No sinus rhythm							✓	NA
Hemodynamic state								
LVEF	✓	✓	✓			✓	✓	✓
Pulmonary hypertension		✓	✓		✓	✓	✓	✓
Acute pulmonary oedema					✓			NA
Critical preoperative state		✓	✓		✓	✓	✓	✓
Cardiogenic shock	✓							NA
Resuscitation	✓							NA
Inotropic agents	✓							NA
Intra-aortic balloon pump	✓							NA
Procedure								
Emergency procedure	✓	✓	✓				✓	✓
Acuity category (2, 3 and 4)[a]				✓				
Thoracic aorta surgery	✓	✓	✓					✓
Aortic valve surgery	✓		✓					✓
Mitral valve surgery	✓							✓
Aortic and mitral valve surgery	✓							✓
Tricuspid valve surgery	✓							NA
Surgery for CHD	✓							NA
Concomitant CABG	✓							✓
Concomitant tricuspid valve surgery	✓							✓
Post-infarct septal rupture		✓						✓
TAVI-specific variables								
Transfemoral (TF) access route				✓				✓
Non-TF Trans-apical access route					✓			✓
Direct aortic, Subclavian, Other access					✓			✓

NA: not available or not registered in the NHR-TAVI registration, CABG: Coronary artery bypass surgery, CHD: Congenital Heart Disease

a Definitions from ACC-TAVI: Category 2 includes urgent procedure status and no pre-procedure shock, inotropes, mechanical assist device, or cardiac arrest; category 3, urgent or elective procedure status, pre-procedure shock, inotropes, or mechanical assist device, and no cardiac arrest less than 24 hours before the procedure; and category 4, emergent or salvage procedure status and/or cardiac arrest less than 24 hours before the procedure (4).

E-components Table 3. Predictor variables with missing (%) values in the NHR-TAVI cohort (6177 patients)

Predictor variables	Missing values in NHR-TAVI cohort	
	n	%
Creatinine, µmol/L, (mean (S.D.))	30.0	0.5
eGFR mL/min/1.73 m² (mean (S.D.))	30.0	0.5
Chronic lung disease	32	0.5
Extra-cardiac arteriopathy	44	0.7
Access route	83	1.3
Critical preoperative state	93	1.5
Recent Myocardial infarction	102	1.7
Height cm. (mean (S.D.))	125.0	2.0
Weight kg. (mean (S.D.))	124.0	2.0
LVEF (mean (S.D.))	161.0	2.6
Dialysis	201	3.3
Procedure urgency	211	3.4
Diabetes	218	3.5
Previous cardiac surgery	245	3.9
Anaesthesia	345	5.6
Procedure weight (2 operations)	417	6.7
Functional NYHA classes (I, II, III, and IV)	845	13.7
Balloon aortic valvuloplasty prior to the date of TAVI	898	14.5
Systolic Pulmonary Artery Pressure mm Hg (mean (S.D.))	2203.0	35.6
Poor mobility	2547	41.2

E-components Table 4. Discriminations (AU-ROC) of the mortality prediction models from external validation cohorts on TAVI patients in 3 countries

	NHR-TAVI cohort	IRRMA study (19)	UK-study (20)
Total population	6177	1327	6676
Mortality rate (%)	280 (4.5)	45 (3.3)	360 (5.4)
Cohort timeframe	2013-2017	2008-2014	2007-2014
Risk Model	**AU-ROC (95% CI)**	**AU-ROC**	**AU-ROC (95% CI)**
Surgical MPM			
STS	0.62 (0.58-0.65)	0.68	0.60 (0.57, 0.63)
EuroSCORE-I	0.59 (0.55-0.62)	0.70	0.57 (0.54, 0.61)
EuroSCORE-II	0.61 (0.57-0.64)	0.70	0.59 (0.55, 0.62)
TAVI-specific MPM			
ACC-TAVI	0.64 (0.61-0.67)	Not tested[$]	0.64 (0.60, 0.67)
FRANCE-2	0.63 (0.60-0.67)	0.71	0.62 (0.59, 0.65)
OBSERVANT	0.58 (0.55-0.62)	0.63	0.57 (0.54, 0.60)
SAVR and TAVI MPM			
German-AV	0.60 (0.57-0.64)	0.52	0.59 (0.56, 0.62)

[$] ACC-TAVI is developed after the publication of the IRMMA study

Abbreviations: NHR = the Dutch Heart Registry (De Nederlandse Hart Registratie); IRRMA = Israeli TAVR Registry Risk Model Accuracy study; SAVR = Surgical Aortic Valve Replacement

E-components Table 5. Discrimination (AU-ROC) of all MPMs in the entire cohort before and after imputation

MPM	AU-ROC before imputation (95% CI)	AU-ROC after imputation (95% CI)	P-Value
Surgical MPM			
STS	0.61 (0.57-0.65)	0.62 (0.58-0.65)	0.716
EuroSCORE-I	0.53 (0.49-0.59)	0.59 (0.55-0.62)	0.122
EuroSCORE-II	0.60 (0.54-0.66)	0.61 (0.57-0.64)	0.813
TAVI-specific MPM			
ACC-TAVI	0.64 (0.60-0.68)	0.64 (0.61-0.67)	0.929
FRANCE-2	0.64 (0.59-0.68)	0.63 (0.60-0.67)	0.912
OBSERVANT	0.57 (0.51-0.62)	0.58 (0.55-0.62)	0.653
SAVR and TAVI MPM			
German-AV	0.57 (0.52-0.63)	0.60 (0.57-0.64)	0.398

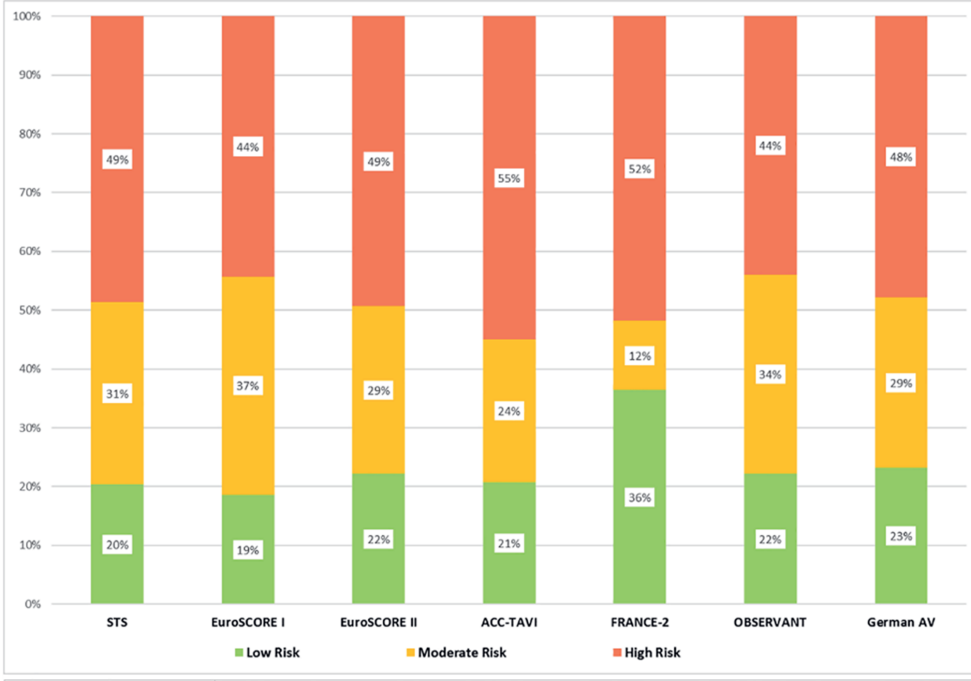
Abbreviations: SAVR = Surgical Aortic Valve Replacement

E-components Table 6. Discrimination (AU-ROC) (95%CI) analysis for early-mortality by subgroups age, gender access route, LVEF, NYHA class and procedure urgency

	Sub-group					
	Age		Gender		Access route	
Risk Model	Age ≤ 75 years	Age > 75 years	Female	Diabetic	TF	Non-TF
STS	0.66 (0.58-0.73)	0.61 (0.57-0.64)	0.65 (0.60-0.69)	0.62 (0.58-0.65)	0.60 (0.56-0.64)	0.62 (0.56-0.68)
EUROSCORE-I	0.62 (0.54-0.70)	0.57 (0.53-0.61)	0.59 (0.55-0.64)	0.58 (0.55-0.62)	0.57 (0.53-0.62)	0.56 (0.51-0.62)
EUROSCORE-II	0.61 (0.53-0.69)	0.61 (0.57-0.64)	0.63 (0.58-0.67)	0.60 (0.57-0.64)	0.60 (0.56-0.64)	0.58 (0.52-0.64)
ACC-TAVI	0.68 (0.60-0.76)	0.63 (0.59-0.67)	0.66 (0.61-0.70)	0.64 (0.60-0.67)	0.60 (0.55-0.63)	0.64 (0.58-0.69)
FRANCE-2	0.67 (0.59-0.67)	0.63 (0.59-0.66)	0.66 (0.61-0.70)	0.63 (0.59-0.66)	0.58 (0.54-0.62)	0.59 (0.53-0.65)
OBSERVANT	0.58 (0.50-0.67)	0.58 (0.54-0.62)	0.61 (0.56-0.66)	0.58 (0.55-0.62)	0.56 (0.52-0.60)	0.61 (0.55-0.67)
German-AV	0.62 (0.53-0.71)	0.60 (0.55-0.64)	0.62 (0.57-0.67)	0.60 (0.57-0.64)	0.59 (0.55-0.63)	0.61 (0.55-0.67)
	Sub-group					
	LVEF		NYHA		Procedure urgency	
Risk Model	LVEF <50%	LVEF ≥50%	NYHA class 3	NYHA class 4	Elective	Non-Elective
STS	0.60 (0.55-0.66)	0.62 (0.58-0.67)	0.59 (0.54-0.63)	0.59 (0.51-0.68)	0.60 (0.56-0.64)	0.63 (0.55-0.71)
EUROSCORE-I	0.59 (0.53-0.64)	0.56 (0.51-0.60)	0.56 (0.52-0.61)	0.58 (0.49-0.57)	0.58 (0.54-0.61)	0.59 (0.51-0.69)
EUROSCORE-II	0.61 (0.56-0.66)	0.59 (0.55-0.63)	0.58 (0.53-0.62)	0.57 (0.48-0.66)	0.59 (0.56-0.63)	0.63 (0.53-0.71)
ACC-TAVI	0.62 (0.57-0.68)	0.65 (0.60-0.69)	0.62 (0.58-0.66)	0.63 (0.54-0.72)	0.63 (0.59-0.67)	0.60 (0.51-0.69)
FRANCE-2	0.61 (0.56-0.66)	0.64 (0.60-0.68)	0.61 (0.58-0.66)	0.59 (0.49-0.69)	0.62 (0.59-0.66)	0.63 (0.55-0.71)
OBSERVANT	0.57 (0.52-0.63)	0.60 (0.56-0.64)	0.58 (0.54-0.62)	0.60 (0.50-0.70)	0.57 (0.53-0.61)	0.58 (0.49-0.67)
German-AV	0.57 (0.52-0.63)	0.61 (0.56-0.66)	0.57 (0.52-0.61)	0.56 (0.47-0.65)	0.60 (0.56-0.64)	0.58 (0.49-0.67)

Abbreviations: LVEF = Left Ventricular Ejection Fraction, TF = Transfemoral Access route

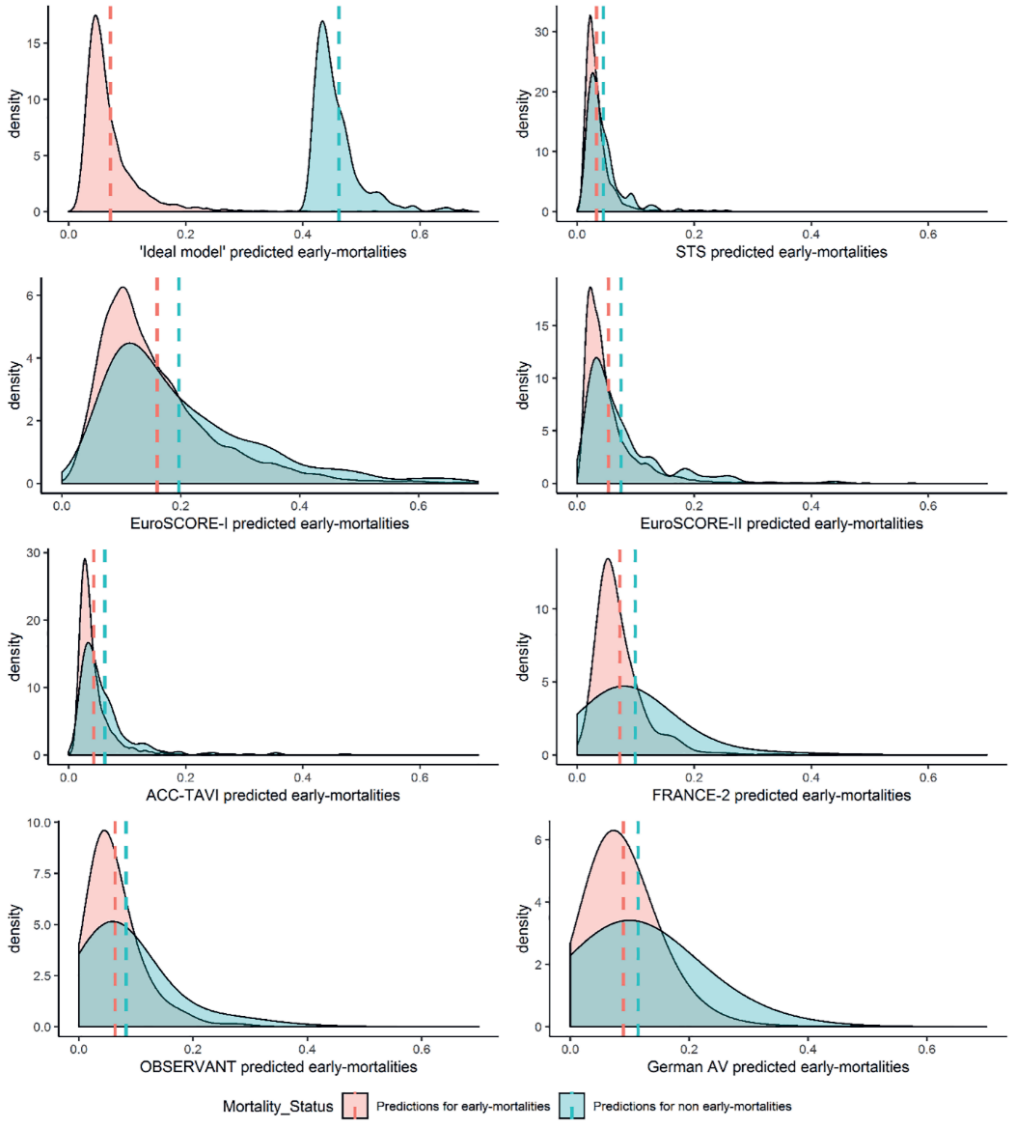
E-components Figure 1. Categorization of early (30-day) mortality into three risk subgroups. The 100% stacked column bar chart showing the categorization of the early-mortality (N=280) into three risk subgroups (low-, moderate- and high-risk mortality) based on the 0, 33, and 66 probability tertiles for each prediction model. Each column represents one model. Numbers in the bars represent the percentage of patients in that subgroup. The best model will predict and allocate most of early-mortality cases in the high-risk subgroup, less in the moderate-risk and least in the low-risk subgroup. ACC-TAVI predicted 55% of early-mortality as high-risk. FRANCE-2 has the second-best high-risk, however, it has the worst moderate-risk and low-risk subgroups.



MPM	Dividing the range of the probabilities distribution (mortality risk) of each MPM into three risk subgroups (categories) by tertiles with equal probabilities in each category		
	Low Risk	Moderate Risk	High Risk
STS	0.38 - 2.28	> 2.28 - 3.51	> 3.51 - 32.09
EuroSCORE-I	2.0 - 10.0	> 10.0 - 17.0	> 17.0 - 76.00
EuroSCORE-II	0.53 - 2.87	> 2.87 - 5.36	> 5.36 - 57.20
ACC-TAVI	0.29 - 2.81	> 2.81 - 4.24	> 4.24 - 47.23
FRANCE-2	3.49 - 5.17	> 5.17 - 7.66	> 7.66 - 59.39
OBSERVANT	1.80 - 4.31	> 4.31 - 6.60	> 6.60 - 40.13
GERMAN AV	0.64 - 6.04	> 6.04 - 9.94	> 9.94 - 55.63
	Total patients predicted in each risk subgroups		
	Low Risk	Moderate Risk	High Risk
STS	57	87	136
EuroSCORE-I	52	104	124
EuroSCORE-II	62	80	138
ACC-TAVI	58	68	154
FRANCE-2	102	33	145
OBSERVANT	62	95	123
GERMAN AV	65	81	134

Abbreviations: MPM= mortality prediction models, STS = Society of Thoracic Surgeons; EuroSCORE = European System for Cardiac Operative Risk Evaluation; ACC-TAVI = American College of Cardiology score for Transcatheter Aortic Valve Implantation; FRANCE-2 = FRench Aortic National CoreValve and Edwards, OBSERVANT = Observational Study of Appropriateness, Efficacy And Effectiveness of AVR-TAVR Procedures For the Treatment Of Severe Symptomatic Aortic Stenosis, AV = Aortic valve.

E-components Figure 2. Density distribution plotting of the early-mortality predicted probabilities for each class (early-death class vs. no early-death class) for each model. The dash lines represent means. The good model (example in the upper left corner) would have a separate plotting density distribution for each predicted class, irrespective of whether the density distribution is thick, thin, long or short. No model had distinguished the predicted classes from each other (noticeable overlapping).



TRIPOD Checklist: Prediction Model Validation

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1*
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and **rationale** for developing or validating the multivariable prediction model, including references to existing models.	3
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	3
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	✓
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	✓
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	✓
	5b	Describe eligibility criteria for participants.	✓
	5c	Give details of treatments received, if relevant.	NA
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	✓
	6b	Report any actions to blind assessment of the outcome to be predicted.	✓
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	✓
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	✓
Sample size	8	Explain how the study size was arrived at.	✓
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	✓
Statistical analysis methods	10c	For validation, describe how the predictions were calculated.	✓
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	✓
	10e	Describe any model updating (e.g., recalibration) arising from the validation, if done.	NA
Risk groups	11	Provide details on how risk groups were created, if done.	✓
Development vs. validation	12	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	✓
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	✓
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	✓
	13c	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	✓
Model performance	16	Report performance measures (with CIs) for the prediction model.	✓
Model-updating	17	If done, report the results from any model updating (i.e., model specification, model performance).	✓
Discussion			
Limitations	18	Discuss any limitations of the study (such as non-representative sample, few events per predictor, missing data).	✓
Interpretation	19a	For validation, discuss the results with reference to performance in the development data, and any other validation data.	✓
	19b	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	✓
Implications	20	Discuss the potential clinical use of the model and implications for future research.	✓
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	✓
Funding	22	Give the source of funding and the role of the funders for the present study.	✓

References:

1) Cribier A, Eltchaninoff H, Bash A, Borenstein N, Tron C, Bauer F, et al. Percutaneous transcatheter implantation of an aortic valve prosthesis for calcific aortic stenosis: first human case description. Circulation. 2002; 106(24):3006-8.
2) Michel P, Roques F, Nashef SA, Euro SPG. Logistic or additive EuroSCORE for high-risk patients? Eur J Cardiothorac Surg. 2003; 23(5):684-7; discussion 7.
3) Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, et al. EuroSCORE II. Eur J Cardiothorac Surg. 2012; 41(4):734-44; discussion 44-5.
4) Roques F, Michel P, Goldstone AR, Nashef SA. The logistic EuroSCORE. Eur Heart J. 2003; 24(9):881-2.
5) 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 Suppl):S23-42.
6) 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 J Cardiothorac Surg. 2012; 42(4):S1-44.
7) Genereux P, Head SJ, Van Mieghem NM, Kodali S, Kirtane AJ, Xu K, et al. Clinical outcomes after transcatheter aortic valve replacement using valve academic research consortium definitions: a weighted meta-analysis of 3,519 patients from 16 studies. J Am Coll Cardiol. 2012; 59(25):2317-26.
8) Tamburino C, Capodanno D, Ramondo A, Petronio AS, Ettori F, Santoro G, et al. Incidence and predictors of early and late mortality after transcatheter aortic valve implantation in 663 patients with severe aortic stenosis. Circulation. 2011; 123(3):299-308.
9) Osnabrugge RL, Speir AM, Head SJ, Fonner CE, Fonner E, Kappetein AP, et al. Performance of EuroSCORE II in a large US database: implications for transcatheter aortic valve implantation. Eur J Cardiothorac Surg. 2014; 46(3):400-8; discussion 8.
10) 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(18):1706-15.
11) Durand E, Borz B, Godin M, Tron C, Litzler PY, Bessou JP, 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.
12) Ben-Dor I, Gaglia MA, Jr., 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.
13) Piazza N, Wenaweser P, van Gameren M, Pilgrim T, Tzikas 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.
14) 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.
15) Iung B, Laouenan 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.
16) 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):46-52.
17) Zbronski K, Huczek Z, Puchta D, Paczwa K, Kochman J, Wilimski R, et al. Outcome prediction following transcatheter aortic valve implantation: Multiple risk scores comparison. Cardiol J. 2016; 23(2):169-77.
18) Collas VM, Van De Heyning CM, Paelinck BP, Rodrigus IE, Vrints CJ, Bosmans JM. Validation of transcatheter aortic valve implantation risk scores in relation to early and mid-term survival: a single-centre study. Interact Cardiovasc Thorac Surg. 2016; 22(3):273-9.
19) 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 Assessment (IRRMA) study. Int J Cardiol. 2016; 215:227-31.
20) Martin GP, Sperrin M, Ludman PF, de Belder MA, Gale CP, Toff WD, et al. Inadequacy of existing clinical prediction models for predicting mortality after transcatheter aortic valve implantation. Am Heart J. 2017; 184:97-105.
21) Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ. 2009; 338:b605.
22) Tang M, Hu P, Wang CF, Yu CQ, Sheng J, Ma SJ. Prediction Model of Cardiac Risk for Dental Extraction in Elderly Patients with Cardiovascular Diseases. Gerontology. 2019:1-8.
23) Linden A. Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis. J Eval Clin Pract. 2006; 12(2):132-9.
24) Pepe MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford: Oxford University Press; 2003 07 October 2004. 320 p.
25) Brier G. Verification of Forecasts Expressed in Terms of Probability. Monthly Weather Review. 1950;78:1-3.
26) Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. BMJ. 2013; 346:e5595.

27) Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. BMJ. 2013; 346:e5793.
28) Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. PLoS Med. 2013; 10(2):e1001380.
29) Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med. 2013; 10(2):e1001381.
30) Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 2015; 10(3):e0118432.
31) The Netherlands Heart Registration (NHR). https://nederlandsehartregistratie.nl/, 2019 (accessed 14 December 2019).
32) Dodd LE, Pepe MS. Partial AUC estimation and regression. Biometrics. 2003; 59(3):614-23.
33) DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988; 44(3):837-45.
34) Venkatraman ES. A permutation test to compare receiver operating characteristic curves. Biometrics. 2000; 56(4):1134-8.
35) Boyd K, Eng KH, Page CD, editors. Area under the Precision-Recall Curve: Point Estimates and Confidence Intervals2013; Berlin, Heidelberg: Springer Berlin Heidelberg.
36) Ozenne B, Subtil F, Maucort-Boulch D. The precision--recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. J Clin Epidemiol. 2015; 68(8):855-9.
37) Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd international conference on Machine learning; Pittsburgh, Pennsylvania, USA. 1143874: ACM; 2006. p. 233-40.
38) Cox DR. Two further applications of a model for binary regression. Oxford University Press on behalf of Biometrika Trust. 1958; 45:562-5 (4 pages).
39) Murphy AH. A New Vector Partition of the Probability Score. . J Appl Meteorol, National Center for Atmospheric Research, Boulder, Colo. 1973:595-600.
40) Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.
41) 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(5):971-7.
42) Arsalan M, Weferling M, Hecker F, Filardo G, Kim WK, Pollock BD, et al. TAVI risk scoring using established versus new scoring systems: role of the new STS/ACC model. EuroIntervention. 2018; 13(13):1520-6.
43) Pilgrim T, Franzone A, Stortecky S, Nietlispach F, Haynes AG, Tueller D, et al. Predicting Mortality After Transcatheter Aortic Valve Replacement: External Validation of the Transcatheter Valve Therapy Registry Model. Circ Cardiovasc Interv. 2017; 10(11).


Update...

Chapter 3:

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 & Vasculature, 2021

https://doi.org/10.1016/j.ijcha.2021.100716

Hatem Al-Farra, Bas A.J.M. de Mol, Anita C.J. Ravelli, Willem Jan P.P. ter Burg, Saskia Houterman, José P.S. Henriques, Ameen Abu-Hanna; on behalf of the NHR THI Registration Committee[#]

Abstract

Background

The predictive performance of the models FRANCE-2 and ACC-TAVI for early-mortality after Transcatheter Aortic Valve Implantation (TAVI) can decline over time and can be enhanced by updating them on new populations. We aim to update and internally and temporally validate these models using a recent TAVI-cohort from the Netherlands Heart Registration (NHR).

Methods

We used data of TAVI-patients treated in 2013-2017. For each original-model, the best update-method (model-intercept, model-recalibration, or model-revision) was selected by a closed-testing procedure. We internally validated both updated models with 1000 bootstrap samples. We also updated the models on the 2013-2016 dataset and temporally validated them on the 2017-dataset. Performance measures were the Area-Under ROC-curve (AU-ROC), Brier-score, and calibration graphs.

Results

We included 6177 TAVI-patients, with 4.5% observed early-mortality. The selected update-method for FRANCE-2 was model-intercept-update. Internal validation showed an AU-ROC of 0.63 (95%CI 0.62-0.66) and Brier-score of 0.04 (0.04-0.05). Calibration graphs show that it overestimates early-mortality. In temporal-validation, the AU-ROC was 0.61 (0.53-0.67).

The selected update-method for ACC-TAVI was model-revision. In internal-validation, the AU-ROC was 0.63 (0.63-0.66) and Brier-score was 0.04 (0.04-0.05). The updated ACC-TAVI calibrates well up to a probability of 20%, and subsequently underestimates early-mortality. In temporal-validation the AU-ROC was 0.65 (0.58-0.72).

Conclusion

Internal-validation of the updated models FRANCE-2 and ACC-TAVI with data from the NHR demonstrated improved performance, which was better than in external-validation studies and comparable to the original studies. In temporal-validation, ACC-TAVI outperformed FRANCE-2 because it suffered less from changes over time.

Introduction

Since 2002, the Transcatheter Aortic Valve Implantation (TAVI) was introduced as a less invasive treatment for patients with severe aortic stenosis at high-mortality risk and not candidate for surgical aortic valve replacement (SAVR) (1-2). Over the last years, TAVI emerged as a safe and efficacious alternative treatment also for intermediate and low-risk patients with severe aortic stenosis (3-4).

Proper risk estimation of post-operative early (30-day) mortality following TAVI using mortality prediction models (MPM) may help heart teams in getting insight into the outcome of TAVI procedures and may help to improve the quality of care. In the past, the classical cardiac surgery MPMs, such as the European System for Cardiac Operative Risk Evaluation (EuroSCORE-I and EuroSCORE-II) (5-7), and the Society of Thoracic Surgeons Predicted Risk of Mortality (STS-PRoM) (8), were used to predict early-mortality after TAVI. However, those classical MPMs had significant limitations in early-mortality prediction after TAVI (9). Therefore, several TAVI-specific MPMs (such as FRANCE-2 and ACC-TAVI) have been developed for preoperative risk estimation (10-11). These TAVI-specific MPMs were externally validated on different TAVI-populations (12-15).

The models FRANCE-2 and ACC-TAVI have been shown to outperform other validated MPMs on their discrimination performance. However, in those external-validation studies, the predictive performance of both MPMs (FRANCE-2 and ACC-TAVI) was still poor. The discrimination in terms of the Area Under the Receiver-Operating-Characteristics (AU-ROC) was 0.63 and 0.64, respectively; calibration was poor; and accuracy was limited (12-13).

Besides, MPMs in general may also lose their predictive performance over time due to performance drift (16-18). Poor predictive performance could be due to deficiencies in the development methods of the original-models, changes in the population's characteristics over time (e.g., expanding TAVI indication to low-risk patients), or due to improvements in the intervention procedure. For these reasons, using such MPMs without their adaptation on an external population is suboptimal (18-22).

Although developing a new MPM from scratch on new datasets is a common practice, especially when the performance of pre-existing models is poor, updating these models can capitalize on information in the pre-existing models (16, 23-26). Updating existing prediction models can indeed improve their performance in new populations as demonstrated in various studies (24-25, 27-28) and enables reusing the MPMs for their original purposes (13, 29-30).

Generally, there are three common updating-methods for logistic regression models: updating the intercept, updating the intercept and slope (model-recalibration), or updating all estimated coefficients (model-revision). The closed-testing procedure described by Vergouwe et al. (23) selects the best updating-method. In spite of the fact that different existing and recently developed TAVI-specific models are available, only a few models were externally validated. For this study, we selected the models FRANCE-2 and ACC-TAVI because they have been externally validated in three external-validation studies and have shown the best performance (12-14). One of these three external-validation studies was recently performed on our own NHR population (14). We hence sought to update the best two externally performing models on our population, and for simplicity did not attempt to update all proposed models.

In this study, we aim to update two TAVI-specific models (FRANCE-2 and ACC-TAVI) for predicting the early-mortality depending on the closed-testing procedure (23). We perform internal-validation on the updated models using a recent TAVI-cohort from the NHR. To understand the performance of the updated models over time, which best mimics the model's envisioned usage in clinical practice; we also perform temporal-validation in which the models are tested on a dataset collected prospectively after the models have been updated on earlier data.

Methods

Study population

In the Netherlands, 16 heart centres perform TAVI procedures. The Dutch heart centres submit patients' data (including demographics, clinical characteristics, intervention risk factors, procedural details, mortality status, complications, and follow-up data after hospital discharge) to the NHR registry (31). In total, data from 13 Dutch heart centres, who had the outcome measurement "30-day mortality", were included in this study. Data from three heart centres were excluded as they did not timely present the outcome measurements. All variables used in each model were obtained from the NHR, including the outcome (early-mortality status), from January 1, 2013, to December 31, 2017. Although the obtained data originate from several centres, in our sample we have no information about which center a patient belongs to due to local privacy regulations.

The prediction models FRANCE-2 and ACC-TAVI

The model FRANCE-2 (French Aortic National CoreValve and Edwards) is an early-mortality risk score. It was developed in 2014, based on the TAVI French registry with 3883 TAVI patients to predict early-mortality after TAVI (10). As reported on the internal-validation of this model, the AU-ROC of FRANCE-2 was 0.59 (95% CI 0.54-0.64), and both calibration-intercept and calibration-slope did not deviate from their ideal values of zero and one (10). This MPM was externally validated (13-14), where the AU-ROC was 0.63 (95% CI 0.60-0.67), the Area Under Precision-Recall Curve (AU-PRC) was 0.09, the Brier-score was 0.044, the Brier-skill score (BSS) was -0.01 and both the calibration-intercept and calibration-slope did significantly deviate from 0 and 1, respectively (13-14) (see E-component Table 1).

ACC-TAVI was developed in 2016 by the society of thoracic surgeons and the American college of cardiology to predict in-hospital mortality in TAVI patients (n=20586) in the United States (11). As reported on the internal-validation of this model, the AU-ROC of ACC-TAVI was 0.66 (95% CI 0.62-0.69), and the calibration-intercept and calibration-slope did not deviate from the ideal values of zero and one, respectively (11). The stated purpose of this MPM is TAVI patient counselling, quality-of-care improvement, and national monitoring for appropriateness of the selection of patients for TAVI. This MPM was externally validated (13-14), where the AU-ROC was 0.64 (95% CI 0.61-0.67), the AU-PRC was 0.09, the Brier-score was 0.043, the BSS was 0.002 and only the calibration-intercept did not deviate from zero (14), (E-component Table 1).

Definition of the primary outcome and the used variable predictors

In this study, the primary outcome is the 30-day mortality or early (post-procedural) mortality, which we defined as death within 30-days from the TAVI procedure date. The variables used in each MPM and the definition of the predictor variables and their corresponding variables in the TAVI-NHR cohort are given in the supplementary material (E-component Tables 2, 3, and 4).

Statistical analysis

Continuous predictor variables are summarized as mean (standard deviation) or median (inter-quartile-range) and were compared using Students' t-test or the Mann-Whitney test as appropriate. Categorical predictor variables are summarized as counts and percentages and were compared using the chi-squared test or Fisher exact test as appropriate. A p-value <0.05 of a 2-tailed test was considered significant for all analyses. For bootstrapping the 95% confidence interval was calculated using the percentile method.

Missing predictors and missing values

There could be missing predictors and missing values of existing predictors. For missing predictors, and in line with the approach used in other studies (13-14, 32), if a variable predictor required by an MPM was not registered

in the NHR-TAVI cohort, the condition represented by this predictor was assumed to be absent for all patients. However, in addition, we performed a sensitivity analysis by simulating the values of the missing predictor variable. In each simulation, we have randomly drawn values, with a probability of 0.5 of each outcome (absent/present), and calculated the predictive performance measures.If the registered existing variable predictors of FRANCE-2 and ACC-TAVI in the NHR-TAVI cohort have missing values then we assumed that they were missing at random, as we have no specific reason to assume otherwise. Therefore, and in line with the approach in other studies (13-14, 32), multiple imputations with ten imputed datasets were generated for the missing values using Multiple Imputation by Chained Equations (MICE) (33). The outcome measure early-mortality was included, as methodologically recommended, in the imputation models of missing variable predictors. The flow diagram in E-component Figure 1 summarizes the following statistical analysis methods.

Selection of the update-method strategy by the closed-testing procedure, and model updating

To select the most appropriate update-method for the two MPMs, we applied the closed-testing procedure of Vergouwe et al. on the whole NHR-TAVI cohort (23). Application of this procedure will decide on one out of the four following update options: no update; update only the intercept (calibration-in-the-large), update both intercept and slope (logistic calibration); or revise the coefficients of the underlying predictors. Details about these methods appear in E-component Methods 1.

The four update-methods imply an increasing number of estimated parameters. Accordingly, the closed-testing procedure allows the extensiveness of the update to increase progressively from a minimum (no update) to a maximum (model-revision). The procedure involves multiple testing with maintaining approximately the chosen type I error rate by implementing a series of likelihood ratio tests of the updated models against the original-model. The procedure will only select the model-revision method if there is enough evidence that the new regression coefficients are significantly different in the updating population (23). The update-method that is referred to as model-extension, which considers adding variable predictors other than the original estimated variable predictors, is outside the scope of this paper.

Internal-validation of the update-method strategy

We repeated the multiple imputations and the closed-testing procedure in each of the 1000 bootstrap samples to choose the update-methods in each sample (34). Specifically, we updated each of the two models (FRANCE-2 and ACC-TAVI) in the bootstrap sample using the corresponding chosen update-method. Consequently, we assessed the optimism corrected performance for our performance measures. The optimism corrected performance was calculated by subtracting the optimism from the apparent model performance, where optimism was based on the difference in the performance of the models trained on the bootstrap samples and tested on the original dataset. We also calculated the proportions of times in which an update method was selected, and the average performance of that chosen method.

Temporal and cross-validation of the updated models

We also validated the predictive performance of the updated MPMs by temporal-validation. Specifically, we updated the models, with the respective selected update-method, on the NHR-TAVI January-2013 up to December-2016 cohort and validated them onward on the cohort from January-2017 up to and including December-2017. This approach reflects the envisioned real-life behavior of the model when facing new patients.

We also performed cross-validation with four folds. This size was chosen so that each fold, and hence the corresponding test set, is about equal to the test set in the temporal-validation approach. Unlike in the temporal-validation approach, in cross-validation, we do not take changes over time (which can denote performance drift) into account. Comparing the results of temporal-validation with cross-validation helps understand whether a model in temporal-validation suffered from changes over time due to drift.

Performance measures

For each of the validation approaches, we used the following performance aspects and their respective measures: discrimination by the Area Under Receiver Operating-Characteristic Curve (AU-ROC); the balance between the positive predictive value (PPV) and the sensitivity by the Area Under Precision-Recall Curve (AU-PRC) (35); calibration by calibration graphs and the Cox method for inspecting the calibration-intercepts and calibration-slopes (36); and prediction accuracy by Brier-score and the Brier-skill score (BSS) (37).

For each updated model, we measured the Youden's index (J statistic), which allows to identify the optimal cut-off point of the early-mortality risk probabilities (38) to strike a balance between sensitivity and specificity. Details about these performance measures appear in E-component Methods 2.

All statistical analyses were performed in the R statistical environment version 3.5.1 (39). Multiple imputations of the dataset were completed using the MICE package. The graphical plots were made using the ggplot2 package. The package pROC was used for constructing the ROC plots and testing the AU-ROCs. The package PRROC was used to construct the PRC plot and obtain the AU-PRCs. The reporting in this study adheres to the TRIPOD checklist for the reporting of multivariable prediction models, the checklist is the E-component Table 8 (40).

Results

General results

To update the existing FRANCE-2 and ACC-TAVI model, we included 6177 TAVI patients from the NHR-TAVI registration (2013-2017) with an observed early-mortality rate of 4.5% (n=280) (Table 1). The mean age of the patients was 80.0 years, 51.0% of the patients were female, 7.6% had NYHA class-IV, and 56.0% had NYHA class-III. Urgent TAVI-procedures were 9.0% and emergency procedures were 0.3%. Patients with critical preoperative state had the highest early-mortality risk of 21.1%, followed by patients with NYHA class-IV 9.4%, dialysis with 9.2%, non-transfemoral access route with 8.2%, and urgent procedure-acuity with 7.6% (Table 1). The mean EuroSCORE-II (the estimated early-mortality risk) for the whole population was 5.5%. The mean EuroSCORE-II in the first year (2013) was slightly higher with 5.8%, while in the last year (2017) it was lower with 5.1%. The same pattern has been observed for the mean estimated early-mortality risk when measured by FRANCE-2 (8.2% in 2013, which gradually dropped over the years to 6.9% in 2017), and when measured by ACC-TAVI (4.8% in 2013, which gradually dropped to 4.1% in 2017).

In the NHR-TAVI cohort, only the predictor variable acute-pulmonary-oedema, which is used in the FRANCE-2, was not registered. The variables systolic pulmonary artery pressure and NYHA class had 35.6% and 13.7% missing values in the TAVI NHR cohort, respectively. The rest of the missing values of predictors were less than 2%. Nine predictors with missing values were imputed with 10 multiple imputations. E-component Table 5 provides details about the percentage of missing values before imputation.

Table 1. Patient baseline and procedural characteristics of the study population (n=6177) stratified according to 30-day postprocedural early mortality

Variable	Total cases N (%)	Alive (n=5897) n (%)	Early-mortality (n=280) n (%)	Risk of early-mortality No. of death/total cases (%)	P-value
Age (year) (mean (SD))	80.0 (6.90)	79.9 (6.9)	80.9 (28.9)	-	0.036
Female gender (yes)	3170 (51.3)	3023 (51.3)	147 (52.5)	4.6	0.731
BMI kg/m^2 (mean (SD))	27.2 (4.88)	27.3 (4.9)	26.4 (9.4)	-	0.006
eGFR (mean (SD))	59.1 (21.31)	59.2 (21.7)	55.9 (20.0)	-	0.014
SPAP (mean (SD))	31.1 (10.93)	30.9 (10.8)	33.7 (12.0)	-	0.002
SPAP >60 mm Hg (yes)	86	80 (1.3)	6 (2.1)	6.9	
Chronic lung disease (yes)	1377 (22.4)	101 (22.2)	76 (27.1)	5.5	0.037
Critical preoperative state (yes)	38 (0.6)	30 (0.5)	8 (2.9)	21.1	<0.001
Dialysis (yes)	87 (1.5)	79 (1.4)	8 (2.9)	9.2	0.057
Functional status NYHA (yes)					<0.001
NYHA class I	666 (12.5)	646 (12.7)	20 (7.1)	4.5	

NYHA class II	1270 (23.8)	1232 (24.2)	38 (13.6)	2.9	
NYHA class III	2991 (56.1)	2851 (55.9)	140 (50.0)	4.7	
NYHA class IV	405 (7.6)	367 (7.2)	38 (13.6)	9.4	
Procedure acuity (yes)					<0.001
Procedure acuity Elective	5415 (90.8)	5200 (91.1)	215 (76.8)	3.9	
Procedure acuity Urgent	536 (9.0)	495 (8.7)	41 (14.6)	7.6	
Procedure acuity Emergency	15 (0.3)	14 (0.2)	1 (0.4)	6.7	
Unstable angina (yes)	10 (0.2)	10 (0.2)	0 (0)	0	$
TAVI access route (yes)					<0.001
Transfemoral (TF) access route (yes)	4926 (79.7)	4744 (80.4)	182 (65)	3.7	
TF Surgical	770 (12.6)	745 (12.8)	25 (8.9)	3.2	
TF Per-cutaneous	2793 (45.8)	2691 (46.2)	102 (36.4)	3.7	
TF Unknown ϕ	1363 (22.3)	1308 (22.5)	55 (19.6)	4.0	
Non-transfemoral access (yes)	1165 (18.8)	1067 (18.1)	96 (34.3)	8.2	
Subclavian access	103 (1.7)	94 (1.6)	9 (3.2)	8.7	
Transapical access	506 (8.3)	462 (7.9)	44 (15.7)	8.7	
Direct aortic access	554 (9.1)	511 (8.8)	43 (15.4)	7.9	
Acute pulmonary oedema (yes)	N.A.	N.A.	N.A.		

Abbreviations: BMI: Body mass index, eGFR: estimated Glomerular Filtration Rate, SPAP: Systolic pulmonary arterial pressure, NYHA: New York Heart Association, TF: Transfemoral.
$: Not applicable. ϕ: the TAVI access route is transfemoral, but the sort (surgical or per-cutaneous) is not registered. N.A.: Not Available in the NHR-TAVI-cohort.

Performance of FRANCE-2 before and after update

The predicted Mortality of the FRANCE-2 model -as measured before updating the model in our population- was 7.4%. The AU-ROC was 0.60 (95% CI 0.58-0.63). The original model overestimates the early-mortality, as shown in the calibration graph (Figure 1). The Brier-score was 0.044. The selected update-method after applying the closed-testing was model-intercept-update (Table 2). Performing this update method on the whole dataset resulted in the corresponding final updated model (E-component Table 6a for the final updated intercept of the model).

Table 2. Internal-validation of the model update and fitting strategy in each of 1000 bootstrap samples that were applied on the original models ACC-TAVI and FRANCE-2.

	Internal-validation of the update-method strategy in 1000 bootstrap samples on the model FRANCE-2			
Performance measures	**FRANCE-2 No-update model**	**FRANCE-2 Model-intercept-update**	**FRANCE-2 Model-recalibration**	**FRANCE-2 Model-revision**
Total number of selected update-methods$	0	339	33	628
AU-ROC# (95% CI)	0	0.64 (0.63-0.67)	0.67 (0.66-0.68)	0.63 (0.61-0.66)
Brier-score (95% CI)	0	0.043 (0.041-0.47)	0.043 (0.041-0.05)	0.043 (0.041-0.05)
	Internal-validation of the update-method strategy in 1000 bootstrap samples on the model ACC-TAVI			
Performance measures	**ACC-TAVI No-update model**	**ACC-TAVI Model-intercept-update**	**ACC-TAVI Model-recalibration**	**ACC-TAVI Model-revision**
Total number of selected update-methods$	9	0	33	958
AU-ROC# (95% CI)	0.64 (0.63-0.66)	0	0.65 (0.64-0.68)	0.63 (0.62-0.66)
Brier-score (95% CI)	0.042 (0.03-0.05)	0	0.043 (0.041-0.05)	0.043 (0.041-0.05)

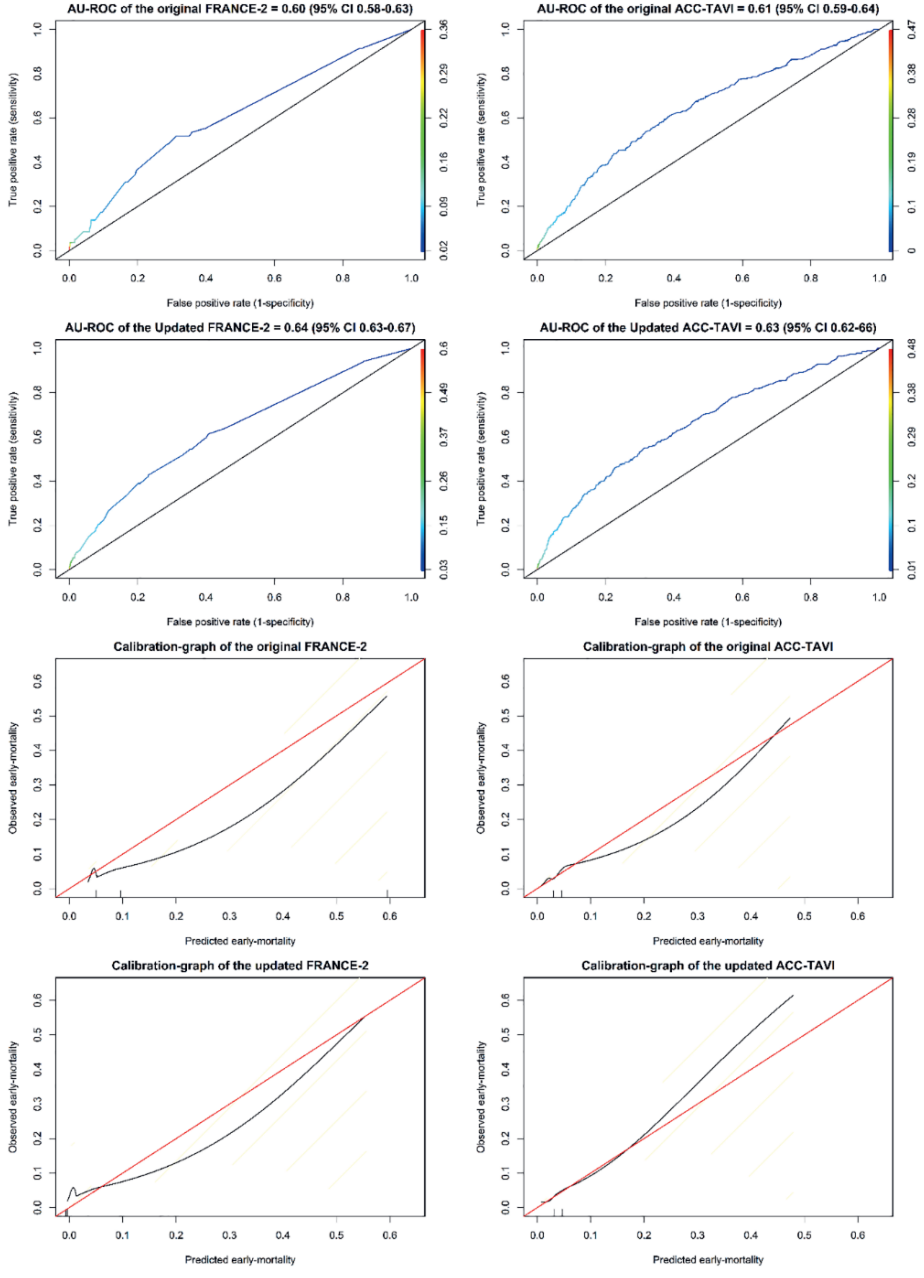
Abbreviations: AU-ROC: Area under the Receiver operating characteristic curve, ACC-TAVI: (ACC TVT) American College of Cardiology Transcatheter Valve Therapy, FRANCE-2: French Aortic National CoreValve and Edwards, N.A. Not applicable.
$ Total number of the selected methods from the 1000 bootstrap drawn with replacement from the whole NHR-TAVI cohort and having the same size.
# The presented AU-ROC is after adjustment for in-sample optimism

The predicted mortality of the updated model was 4.8%. The optimism-corrected AU-ROC was 0.64 (0.63-0.67). The updated FRANCE-2 model initially underestimate up to 5% probability then overestimates the early-mortality as shown in the updated FRANCE-2 calibration graph (Figure 1). Only the calibration-slope did not deviate from its expected value of 1 (Table 2). Brier-score was 0.043 (0.041-0.47).

Repeating the whole model-updating strategy involving the multiple imputations and update-method selection on 1000 bootstrap samples revealed that model-revision was the most selected method (in 62% of the bootstrap samples) (Table 2). This is unlike the model-intercept update method that happened to be selected when deploying the close-test on only the whole cohort. However, the optimism-corrected AU-ROC via bootstrapping

of the updated FRANCE-2 (with model-revision) was 0.63 (95% CI 0.61-0.66). Only the calibration-slope did not deviate from its expected value of 1 (Table 2). Details of the results about the Area Under the Precision-Recall-curve, Brier-skill score, and Calibration-in-the-large and calibration-slope appear in E-component Results 1 and 2.

Figure 1. Area under receiver operating characteristic curves and calibration graphs of the original and the updated models France-2 and ACC-TAVI. In the calibration graphs: the vertical axes represent the observed early-mortality, while the horizontal axes represent the predicted probabilities of the early-mortality. Note that there is a high density of cases in the lower range of probabilities.



Performance of ACC-TAVI before and after update

The predicted mortality of the original ACC-TAVI model on our population before being updated was 4.4%. The AU-ROC was 0.61 (95% CI 0.59-0.64). The model calibrates well up to predictions of 5%, were most of the predications are. Subsequently, overestimates the early-mortality as shown in the calibration plot (Figure 1). The Brier-score was 0.043.

Applying the closed-testing method resulted in selecting model-revision as update-method (Table 2). Performing the selected update method on the data results in the corresponding final updated model (E-component Table 6b for the final updated model).

The predicted mortality of the updated model was 4.4%. The optimism-corrected AU-ROC via bootstrapping was 0.63 (95% CI 0.62-0.66). The updated model calibrates well up to predictions of 20%, subsequently, it underestimates the early-mortality proportion as shown in the calibration graph (Figure 1).

Both calibration-intercept and calibration-slope deviated from their ideal values of 0 and 1 (Table 2). The Brier-score was 0.043 (0.041-0.05). Details of the results about the Area Under the Precision-Recall-curve, Brier-skill score, and Calibration-in-the-large and calibration-slope appear in E-component Results 1 and 2.

Results of temporal-validation for both updated models

In the temporal-validation, the training set (years 2013-2016) included data of 4345 patients with an observed early-mortality rate of 5.1%. We updated in this training set each of the original MPMs using the respective update methods that have been selected earlier by applying the closed-testing procedure on the whole cohort. On the validation set (the cohort of the year 2017, n= 1832, observed early-mortality rate = 3.2%).

For the updated model FRANCE-2 (intercept-update), the predicted early-mortality rate in the validation set was 4.6%, the AU-ROC was 0.61 (95% CI 0.53-0.67) (Table 3). For the updated model ACC-TAVI (model revision) the predicted early-mortality rate was 4.6%, and the AU-ROC was 0.65 (95% CI 0.58–0.72) (Table 3). Details of the results about the Area Under the Precision-Recall-curve, Brier-skill score, and Calibration-in-the-large and calibration-slope appear in E-component Results 3.

Table 3. Results of temporal-validation and cross-validation for the updated-models ACC-TAVI (updated with model-revision) and FRANCE-2 (updated with model-intercept-update). The development sample of the temporal-validation (cohort 2013-2016) n= 4345. The validation sample (cohort 2017) n=1832. Table is showing the results of the 4-folds cross-validation (n= 1544 per fold).

Performance measures	ACC-TAVI Model-revision	FRANCE-2 Model-intercept-update
Temporal-validation		
AU-ROC# (95% CI)	0.65 (0.58-0.72)	0.61 (0.53-0.67)
Brier score	0.031	0.031
4-folds cross-validation		
AU-ROC# (95% CI)	0.65 (0.64-0.68)	0.63 (0.62-0.67)
Brier-score (95% CI)	0.043 (0.041-0.05)	0.043 (0.04-0.5)

Abbreviations: AU-ROC: Area under the Receiver operating characteristic curve, ACC-TAVI: (ACC TVT) American College of Cardiology Transcatheter Valve Therapy, FRANCE-2: French Aortic National CoreValve and Edwards,
The presented AU-ROC is after adjustment for in-sample optimism

Results of cross-validation for both updated models

In the 4-folds cross-validation (per fold n=1544), the AU-ROC updated FRANCE-2 via the intercept-update was 0.63 (95% CI 0.62-0.67). The AU-ROC for the ACC-TAVI (updated with model-revision) was 0.65 (95% CI 0.64-0.68) (Table 3). For both updated models, the calibration-intercept and calibration-slope did not deviate from their ideal values in any of the folds (Table 3).

Discussion

In this study, we updated and internally and temporally validated the FRANCE-2 and ACC-TAVI prediction models for early-mortality for TAVI patients with contemporary data of TAVI patients from the Netherlands Heart Registration. The update-method for FRANCE-2 was intercept-update and the internally validated AU-ROC was 0.63. The update-method for ACC-TAVI was model-revision and the internally validated AU-ROC was 0.63. After updating the models and on temporal-validation, ACC-TAVI did not have a significantly better AU-ROC than FRANCE-2 (0.65 vs. 0.61, p = 0.06).

Internal-validation of both updated models FRANCE-2 and ACC-TAVI showed AU-ROC (discrimination) to be comparable to the reported performance of their original models (10-11), and better than the performance reported in the external-validation studies (12-14). (Table 2 and E-component Table 1).

Temporal-validation showed improvement in the discrimination ability of the updated models, which was comparable to the original model (11) (Table 3 and E-component Table 1). However, both calibration-intercept and calibration-slope for both updated models have significantly deviated from their ideal values of 0 and 1, respectively.

The updated model FRANCE-2 calibrated poorly in this study. The calibration plot of FRANCE-2 in Figure 1 shows significant deviations from the ideal calibration in the whole risk range. It underpredicted early-mortality in lower-risk classes (up to 5%) and after that overpredicted early-mortality.

On the other hand, the updated ACC-TAVI calibrated better than the updated FRANCE-2 in this study. The calibration plot of ACC-TAVI in Figure 1 did not show significant deviations from ideal calibration in the first two deciles of risk (up to 20%). However, the updated ACC-TAVI has underpredicted early-mortality in the higher-risk range, above 20%. Miscalibration in external-validation studies is common (12-15). A common reason could be the improvement of care and/or procedural techniques that took place between the development time of the original MPMs and the time of external-validation. Both FRANCE-2 and ACC-TAVI were developed in the years 2014 and 2016, respectively. Besides, TAVI became a common procedure and the learning curve is likely to have flattened. In addition, the TAVI population's characteristics have been changing over time (e.g., expanding TAVI indication to intermediate and low-risk patients) instead of the initial predominance of high-risk cases. This is likely the reason behind the reduced mortality risk in this study. A noticeably decrease in early-mortality after TAVI-procedures during that period has been reported (41-43).

It is worth mentioning that the included patients' data in this study were collected in a period where most of the candidate patients originated from the high-risk category. However, in 2017 (the last year in our cohort), the European Society of Cardiology guidelines suggested offering TAVI procedures for intermediate-risk patients. This could explain the gradually decreasing mean early-mortality risk, which we measured over the years by the EuroSCORE-II, FRANCE-2 and ACC-TAVI. The inclusion of a relatively lower risk group of patients in the last year increases the heterogeneity of the patient sample in that year. This heterogeneity, in turn, has likely contributed to the improved discriminatory ability of the models in the temporal-validation. All these factors might affect the predictive performance and calibration of logistic prediction models. Therefore, and unlike in cross-validation, in temporal-validation, both updated models did not perform well when facing population drift. This underscores the importance of implementing a periodic dynamic model update for TAVI-specific MPMs.

Individual updates for the prediction models FRANCE-2 and ACC-TAVI were performed previously by Martin et al. in 2018 (32). In that study, a "hybrid method" was proposed for updating and aggregating multiple MPMs. A method was used that re-calibrates multiple MPMs using stacked regression while simultaneously revising specific covariates in the final model. They updated both models (FRANCE-2 and ACC-TAVI, in addition to other MPMs) for comparison purposes with their proposed new method. Both updated models (FRANCE-2 and ACC-TAVI) had an AU-ROC of 0.63 and 0.64, respectively (32), which are fairly comparable with our findings.

To the best of our knowledge, our study is new in its use of the closed-testing procedure for selecting appropriate update methods for these two TAVI-specific early-mortality models FRANCE-2 and ACC-TAVI. Apart from the study of Martin et al. (32), we could not find updating studies that report on updating TAVI-specific early-mortality models such as FRANCE-2 or ACC-TAVI. There are, however, model updating studies in cardiology and cardiac surgery for MPMs other than the TAVI-specific MPMs (25-26, 30, 44). These studies used either one of the update-methods (intercept update, intercept and slope update, model-revision, or model extension), or apply all of them and choose the model with the best performance without using a formal testing procedure (25-26, 44). The utility of deploying the formal closed-testing procedure for selecting an appropriate update method has been motivated by van Calster et al (45). Of note, in the study (45) and unlike our study, the predictive performance of the updated model was measured instead of repeating the whole update strategy itself as we did in this paper, and as we would recommend.

Our study has several strengths. We used a large multi-center dataset of more than 6000 TAVI patients from a recent national registry dataset. We also used comprehensive predictive performance measures (including the area under the precision-recall curve and Brier-skill Score) to quantify the predictive performance of the updated models. In addition to the internal validation in which the update strategy was repeated in 1000 bootstrap samples, we also performed temporal-validation to inspect the real-life behaviour of the updated models when facing new patients; and cross-validation to understand whether this behaviour is ascribed to performance drift.

A limitation of this study is that the predictor variable acute-pulmonary-oedema in FRANCE-2 is not registered in the NHR-TAVI registration (E-component Tables 2, 4, and 5). Therefore, in line with other studies (13-14, 32), we assumed that acute-pulmonary-oedema was absent for all patients. However, to understand whether there is a risk of bias, we performed a sensitivity analysis by simulating the values of the acute-pulmonary-oedema predictor and calculated the performance measures of the updated model. The analysis revealed essentially the same predictive performance measures (E-component Table 7). Another limitation is that the generalisability of the updated models is unknown since we were unable to externally validate the predictive performance in external data. Thus, we recommend researchers to externally validate the models.

We found that ACC-TAVI had the best predictive performance for early-mortality for TAVI patients. For clinical practice, although most of the existing MPMs for TAVI patients are still far from having a good performance, updating the models on new populations does improve their predictive performance, and hence improves their applicability for supporting clinical decision-making.

This study also showed that the updated MPMs suffer from performance drift over time. One should hence, in general, consider a dynamic strategy for updating prediction models, to maintain their relevance to contemporary patient populations. This is a topic that is becoming more pertinent as interventions are increasingly given to lower risk patients (17).

A strategy using statistical process control (SPC) to detect structural deviations from the natural variability in a prediction model's behaviour over time has been suggested as a possible solution to correct for population, and hence, performance drift (46). In addition, the implications of performance drift on benchmarking have been demonstrated (47), which is useful for quality-of-care officers.

Different lines of work merit future research. First, one may consider model-extension techniques for updating prediction models, whereby additional predictors (such as anatomical features and dynamic and continuous parameters from ECG, MRI or ECHO) are considered beyond those used in the original model. Second, instead of early-mortality, researchers may consider updating, extending and validating such MPMs with long-term (1-year) mortality as a primary end point. Third, one may consider comparing the updated (and extended) models, with new MPMs developed using different machine learning methods. Fourth, because the estimation of low prevalent outcomes like early-mortality (14, 41-43) is challenging, one might also look at more prevalent outcomes, such as combining several adverse outcomes (post-operative mortality and complications such as

paravalvular leak, major vascular bleeding, stroke and permanent pacemaker implantation) or other patient-relevant outcomes like quality of life. Using additional variables in model extension and update, and by applying machine learning approaches to develop new models, might help identify the best treatment to offer (TAVI vs. SAVR) with the lowest predicted post-operative complication rate (48), assuming that the patient is readily eligible to the given alternatives. Finally, there is a need for more external- and temporal-validation and model updating studies in other countries (30, 49).

Conclusion

Applying the update-methods and the internal-validation methods on the FRANCE-2 and ACC-TAVI prediction models with data from the NHR-TAVI registration improved the performance of the models to the extent of their original internal validation. Currently, the updated ACC-TAVI with model-revision proved to be the best current tool for early-mortality risk prediction in TAVI patients. However, the predictive performance of the updated models is still suboptimal. The updated models FRANCE-2 and ACC-TAVI are not guaranteed to improve performance in new populations, and hence we recommend that, if possible, other countries and centres consider model updates in their populations as well. Moreover, findings from temporal-validation reinforce the need for implementing a periodic dynamic model update strategy to overcome the effect of performance drift.

Declaration of Competing Interest

Hatem Al-Farra, Bas A.J.M. de Mol, Anita C.J. Ravelli, Willem Jan P.P. ter Burg, Saskia Houterman, José P.S. Henriques, and Ameen Abu-Hanna declare no conflict of interest.

Author Contributions

HF, AAH, AR, and BM contributed to the conception, design, and coordination of the research. The registry commission supervised the collected national TAVI registry data. HF performed all formal analyses with close involvement of AHH and AR. HF wrote the first draft, and AAH and AR contributed to writing the first manuscript. All authors contributed to the critical revision of the article, and the final version was approved by all the authors.

Acknowledgments

The authors are grateful to the participants of the NHR registry for their important contributions to data collection and the members of the NHR THI Registration Committee of the NHR registry.

Addendum

The following physicians are the members of the NHR THI Registration Committee of the NHR registry. They represent the hospitals that have provided the TAVI data for this study. Contact with NHR THI Registration Committee can be via the e-mail: info@nederlandsehartregistratie.nl

The NHR THI Registration Committee

M.M. Vis, Amsterdam University Medical Centers.
J. Vos, Amphia Hospital.
L. Timmers, St. Antonius Hospital.
W.A.L. Tonino, Catharina Hospital.
C.E. Schotborgh, Haga Hospital.
V. Roolvink, Isala.
F. Porta, Leeuwarden Medical Center.
M.G. Stoel, Medisch Spectrum Twente.
S. Kats, Maastricht University Medical Center.
G. Amoroso, Onze Lieve Vrouwe Gasthuis.
H.W. van der Werf, University Medical Center Groningen.
P.R. Stella, University Medical Center Utrecht.
P. de Jaegere, Erasmus University Medical Center.

E-components Supplementary Methods, Results, Figures and Tables:

E-component Methods:

E-component Methods 1: Details about the update methods

1) **No update**: in this case, no update is performed and one just uses the original-model as such (49).

2) **Model intercept update** (calibration-in-the-large)**:** In general, this method is used when there is only a difference in the prevalence of the outcome (early-mortality) between the original development and the new population. Updating the model means that a new logistic regression model is created with the following linear predictor $lp_{updated\_model}$

$$lp_{updated\_model} = \alpha_{new} + lp_{original}$$

Where $lp_{original}$ is the linear predictor of the original model, which is equal to the natural logarithm of the predicted probabilities. We obtained these probabilities using the published regression coefficients of each original MPM.

Hence, only the intercept, α_{new}, is fitted to make the average predicted early-mortality probability equal to the observed overall early-mortality prevalence in the external dataset. In our analysis, we fitted a logistic regression model in each of our 10 imputed datasets with the intercept α_{new} as the only free parameter and the $lp_{original,}$ as an offset predictor (i.e., the slope is fixed at unity). For each patient, we averaged the predicted early-mortality obtained from the 10 imputed datasets.

3) **Model intercept and slope update (logistic calibration):** This method is used when the regression coefficients of the original model are over-fitted or under-fitted for the new population. This method will update both the original intercept and the overall calibration-slope. To this end, we fitted logistic regression models in our 10 imputed datasets with $lp\ original$ as the only covariable, as follow:

$$lp_{updated\_model} = \alpha_{new} + \beta_{overall} * lp_{original}$$

4) **Model-revision**: This method updates all the individual regression estimate coefficients of the original model. We used all variable predictors to fit logistic regression models of the following form on the 10 imputed datasets.

$$lp_{updated\_model} = \alpha_{new} + \beta_{new,1} * X_1 + \ldots + \beta_{new,n} * X_n$$

Where α_{new} and $\beta_{new,i}$ indicate the intercept and the *n* regression coefficients. Subsequently, we pooled the estimated coefficients and standard errors across the 10 models according to Rubin's rules (49, 50).

E-component Methods 2: Definitions of the used performance measures:

AU-ROC: The Area Under the Receiver Operating Characteristic Curve is a discrimination measure that gauges the ability of the model to assign higher probability of mortality to those who die than those who live. A value of 0.5 indicates lack of discrimination and a value of 1 indicates perfect discrimination. The AU-ROC is equal in value to the c-index.

AU-PRC: The Area Under the Precision-Recall Curve is an aggregate measure of the balance between the positive predictive value (also referred to as precision in the information retrieval community) and sensitivity (also referred to as recall). The higher the value the better the model. The maximum value is 1.

Brier-score: This score measures the error in the precision of the predicted probabilities. A perfect model will have a score of 0 and a model that provides the probability of 0.5 to all subjects will have a score of 0.25.

BSS: The Brier Skill Score measures the room of improvement in the Brier Score of a model compared to the Brier Score of a non-informative model that provides all

Calibration-graphs: These are graphs showing the agreement between the predicted probabilities and the proportion of the events. A well calibrated model will have a graph that is on or close to the diagonal line extending from the point (0, 0) to (1, 1).

Youden's index: A statistic equal to sensitivity + specificity -1. The maximum value of the index can be used as a criterion for selecting the optimum cut-off point on the ROC curve. This threshold corresponds to maximizing the distance in the ROC plot from the identity (diagonal) line of the ROC plot. A perfect model would have a Youden's index of 1.

E-component Results:

E-component Results 1:

Results of Area Under the Precision-Recall-curve, Brier-skill score and Calibration-in-the-large and calibration-slope. Validating the update-method strategy in each of 1000 bootstrap samples that were applied the original models ACC-TAVI and FRANCE-2.

Performance measures	Internal validation of the update-method strategy in 1000 bootstrap samples on the model FRANCE-2			
	FRANCE-2 Original model (No-update model)	**FRANCE-2 Model-intercept-update**	**FRANCE-2 Model-recalibration**	**FRANCE-2 Model-revision**
Total number of selected update-methods$	0	339	33	628
AU-PRC (95% CI)	0	0.09 (0.08-0.11)	0.08 (0.07-0.11)	0.11 (0.10-0.14)
Brier-skill score (95% CI)	0	0.005 (0.001-0.01)	0.004 (0.001-0.02)	0.015 (0.01-0.03)
Calibration-on-the-large "Calibration-intercept"	N.A.	Not calibrated -0.41 (-0.69 - -0.13)	Not calibrated -0.44 (-0.59 - -0.10)	Not calibrated -0.39 (-0.56 - -0.10)
Logistic-calibration "Calibration slope"	N.A.	Calibrated 1.00 (0.93-1.04)	Calibrated 0.99 (0.93-1.04)	Calibrated 0.99 (0.92-1.03)

Performance measures	Internal validation of the update-method strategy in 1000 bootstrap samples on the model ACC-TAVI			
	ACC-TAVI Original model (No-update model)	**ACC-TAVI Model-intercept-update**	**ACC-TAVI Model-recalibration**	**ACC-TAVI Model-revision**
Total number of selected update-methods$	9	0	33	958
AU-PRC (95% CI)	0.09 (0.04-0.11)	0	0.08 (0.04-0.1)	0.11 (0.1-0.14)
Brier-skill score (95% CI)	0.002 (0.001-0.004)	0	0.004 (0.002-0.01)	0.014 (0.01-0.03)
Calibration-on-the-large "Calibration-intercept"	Not calibrated -0.39 (-0.61 - -0.12)	N.A.	Not calibrated -0.38 (-0.62 - -0.09)	Not calibrated -0.37 (-0.49 - -0.12)
Logistic-calibration "Calibration slope"	Not calibrated 1.16 (1.08-1.32)	N.A.	Not calibrated 1.11 (1.06-1.29)	Not calibrated 1.13 (1.01-1.39)

Abbreviations: AU-PRC: Area Under the Precision-Recall-curve, ACC-TAVI: (ACC TVT) American College of Cardiology Transcatheter Valve Therapy, FRANCE-2: French Aortic National CoreValve and Edwards, N.A. Not applicable.
$ Total number of the selected methods from the 1000 bootstrap drawn with replacement from the whole NHR-TAVI cohort and having the same size.
# The presented AU-ROC is after adjustment for in-sample optimism

E-component Results 2: Results of the Youden's index (J statistic) from the temporal validation:

For the updated-model FRANCE-2 (intercept-update), the best threshold of the early-mortality probabilities was 0.05. At this point, the model's specificity was 0.69, the accuracy was 0.68, the precision (PPV) was 0.05, and the sensitivity (recall) was 0.52. For the updated-model ACC-TAVI (model-revision) the best threshold (the optimal cut-off point) of the early-mortality probabilities was 0.04. At this point, the model's specificity was 0.55, the accuracy was 0.56, the precision (PPV) was 0.05, and the sensitivity (recall) was 0.69.

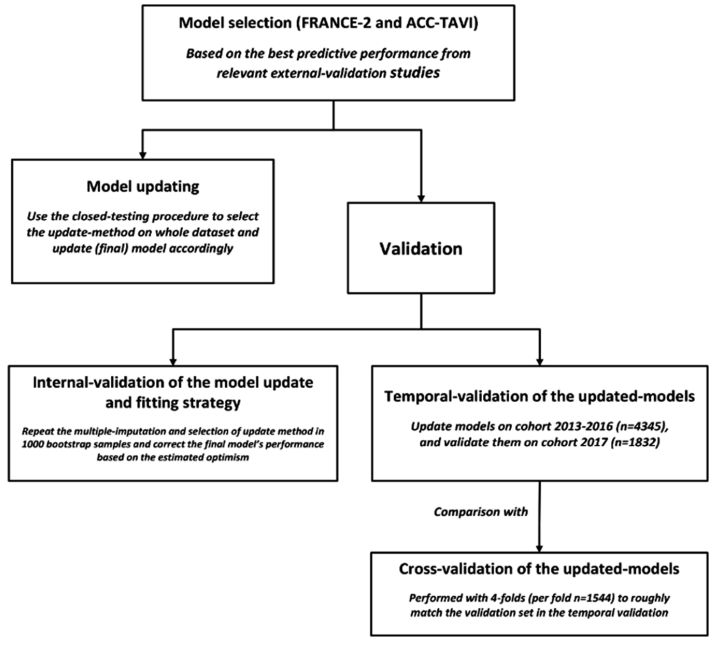
E-component Results 3:

Results of Area Under the Precision-Recall-curve, Brier-skill score and Calibration-in-the-large and calibration-slope. Table is showing the temporal-validation's results of the updated-models ACC-TAVI (updated with model-revision) and FRANCE-2 (updated with model-intercept-update). The development sample of the temporal-validation (cohort 2013-2016) n= 4345. The validation sample (cohort 2017) n=1832. Table is showing the results of the 4-folds cross-validation (n= 1544 per fold).

Performance measures	FRANCE-2 Model-intercept-update	ACC-TAVI Model-revision
Temporal validation		
AU-PRC (95% CI)	0.07 (0.023-0.35)	0.05 (0.01-0.28)
Brier-skill score	-0.03	-0.02
Calibration-on-the-large **"Calibration-intercept" (95% CI)**	Not calibrated -0.43 (-0.68 - -0.15)	Not calibrated -0.39 (-0.67 - -0.15)
Logistic-calibration **"Calibration slope" (95% CI)**	Not calibrated 1.12 (1.06-1.23)	Not calibrated 1.13 (1.04-1.22)
4-folds cross validation		
AU-PRC (95% CI)	0.09 (0.08-0.11)	0.099 (0.094-0.11)
Brier-skill score (95% CI)	0.007 (0.003-0.02)	0.005 (0.003-0.01)
Calibration-on-the-large **"Calibration-intercept" (95% CI)**	All folds calibrated	All folds calibrated
Fold 1	-0.19 (-0.45-0.06)	-0.02 (-0.27-0.22)
Fold 2	0.10 (-0.14-0.33)	-0.02 (-0.27-0.23)
Fold 3	-0.05 (-0.30-019)	0.07 (-0.18-0.29)
Fold 4	0.12 (-0.12-0.35)	-0.03 (-0.29-0.21)
Logistic-calibration **"Calibration slope" (95% CI)**	All folds calibrated	All folds calibrated
Fold 1	1.06 (1.15-0.97)	0.99 (0.92-1.08)
Fold 2	0.97 (1.05-0.89)	0.99 (0.91-1.07)
Fold 3	1.01 (1.09-0.93)	0.99 (0.92-1.08)
Fold 4	0.95 (1.03-0.88)	1.00 (0.92-1.08)

Abbreviations: AU-PRC: Area Under the Precision-Recall-curve, ACC-TAVI: (ACC TVT) American College of Cardiology Transcatheter Valve Therapy, FRANCE-2: French Aortic National CoreValve and Edwards,
The presented AU-ROC is after adjustment for in-sample optimism

E-component Figure 1: Flow diagram of the statistical analysis methods



E-component Tables:

E-component Table 1: Comparison of the predictive performance of the models ACC-TAVI and FRANCE-2 in the original studies, in the external validation studies (Martin et. al) (1) and (Al-Farra et. al) (14), and in this study.

FRANCE-2					
		External validation studies		Updated FRANCE-2 Model-intercept-update	
Performance measures	**Original study (10) n= 3833**	**UK-STUDY (1) n= 6676**	**NHR-STUDY (14) n= 6177**	**Internal validation n= 6177**	**Temporal validation validation set (year 2017) n= 1832**
Observed early mortality risk (n)	10.0% (382)	5.4% (360)	4.5% (280)	4.5% (280)	3.2% (58)
Estimated early-mortality risk	Reported as good concordance from observed risk	9.2% (overestimation)	7.4% (overestimation)	-	4.6%
AU-ROC (95%CI)	0.59 (0.54-0.64)	0.62 (0.59-0.65)	0.63 (0.60-0.67)	0.64 (0.63-0.67)	0.61 (0.53-0.67)
AU-PRC	N.A.	N.A.	0.09	0.09 (0.08-0.11)	0.07 (0.023-0.35)
Brier-score	N.A.	0.053	0.044	0.043 (0.041-0.47)	0.031
Brier-skill score	N.A.	N.A.	-0.01	0.005 (0.001-0.01)	-0.03
Calibration-intercept (95%CI)	Calibrated	−0.60 (−0.71- −0.49)	-0.53 (-0.66- -0.41)	Not calibrated -0.41 (-0.69 - -0.13)	Not calibrated -0.43 (-0.68 - -0.15)
Calibration-slope (95%CI)	Calibrated	0.69 (0.53-0.86)	1.21 (1.16-1.26)	Calibrated 1.00 (0.93-1.04)	Not calibrated 1.13 (1.04-1.22)

ACC-TAVI					
		External validation studies		Updated ACC-TAVI Model-revision	
Performance measures	**Original study (11) n= 20586 Validation set n = 6868**	**UK-STUDY (1) n= 6676**	**NHR-STUDY (14) n= 6177**	**Internal validation n= 6177**	**Temporal validation validation set (year 2017) n= 1832**
Observed early mortality risk (n)	4.4% (300)	5.4% (360)	4.5% (280)	4.5% (280)	3.2% (58)
Estimated early-mortality risk	Reported with no significant difference from observed risk	5.2%	4.4%	4.6%	4.6%
AU-ROC (95% CI)	0.66 (0.62-0.69)	0.64 (0.60-0.67)	0.64 (0.61-0.67)	0.63 (0.62-0.66)	0.65 (0.58-0.72)
AU-PRC	N.A.	N.A.	0.09	0.11 (0.1-0.14)	0.05 (0.01-0.28)
Brier-score	N.A.	0.051	0.043	0.043 (0.041-0.05)	0.031
Brier-skill score	N.A.	N.A.	0.002	0.014 (0.01-0.03)	-0.02
Calibration-intercept (95%CI)	Calibrated	0.04 (−0.07-0.15)	0.04 (-0.08-0.16)	Not calibrated -0.37 (-0.49 - -0.12)	Not calibrated -0.39 (-0.67 - -0.15)
Calibration-slope (95%CI)	Calibrated	0.67 (0.52-0.82)	0.98 (0.94-1.01)	Not calibrated 1.13 (1.01-1.39)	Not calibrated 1.13 (1.04-1.22)

Abbreviations: AU-ROC: Area under the Receiver operating characteristic curve, AU-PRC: Area Under the Precision-Recall-curve, ACC-TAVI: (ACC TVT) American College of Cardiology Transcatheter Valve Therapy, FRANCE-2: French Aortic National CoreValve and Edwards, N.A. Not applicable.

E-component Table 2: The variables used in the original models ACC-TAVI and FRANCE-2

Variable	FRANCE-2 (10 variables)	ACC-TAVI (9 variables)
Age (mean (SD))	✔	✔
Gender (female) (yes)		
sPAP[a] (mean (SD))	✔	
sPAP[a] >60 mm Hg (yes)	✔	
Chronic lung disease (yes)	✔	✔
Critical preoperative state (yes)	✔	
Dialysis (yes)	✔	✔
Functional status NYHA (yes)		
NYHA class IV	✔	✔
Procedure acuity (yes)		
Procedure acuity Elective		✔
Procedure acuity Urgent		✔
Procedure acuity Emergency		✔
TAVI access route (yes)		
Non Transfemoral access (yes)		✔
Subclavian access	✔	
Transapical access	✔	
Direct aortic access	✔	
Other access	✔	
BMI (mean (SD))		
BMI <18 (yes)	✔	
BMI 18-30 (yes)	✔	
eGFR (mean (SD))		✔
Acute pulmonary oedema	✔	

Abbreviations: sPAP: Systolic pulmonary arterial pressure, NYHA: New York Heart Association, BMI: Body mass index, eGFR: estimated Glomerular Filtration Rate, TF: Transfemoral.

E-component Table 3: Variable matching between the ACC-TAVI model and the 2013-2017 NHR-TAVI registration

ACC-TAVI Variable	NHR TAVI	Mapped TAVI Values
Age per 5-year increments	Age at operation	Age divided by 5 rounded down to whole number
eGFR Calculated based on age, sex, race, pre-procedure creatinine and requirement of preprocedural dialysis.	Age at operation Sex, Creatinine	Calculated by the Modification of Diet in Renal Disease formula
Dialysis vs no dialysis The patient is currently undergoing either haemodialysis or peritoneal dialysis on an ongoing basis as a result of renal failure.	On dialysis	Yes
NYHA class IV The patient has cardiac disease with dyspnoea at rest that increases with any physical activity, resulting in inability to perform any physical activity without discomfort.	NYHA status	Symptoms at rest or minimal activity
Severe chronic lung disease The patient has a history of severe chronic lung disease, defined as FEV1 <50% predicted, and/or room air pO2 < 60 or Room Air pCO2 > 50.	History of pulmonary disease	COAD/emphysema Asthma Other significant pulmonary disease
Non-femoral access site The procedure was performed using a femoral access site for the valve sheath.	Delivery approach	**Not**: "Femoral – percutaneous" or "Femoral – surgical"
Procedure acuity status category:		
Acuity status category 2: The patient meets both of the following criteria: 1. Procedure status is urgent. 2. No pre-procedure shock, inotropes, mechanical assist device, or cardiac arrest are required.	Critical Pre-Operative Status Previous MI Procedure acuity	Yes if Procedure acuity = "Urgent" AND Critical pre-operative status = "No" AND Previous MI is No MI
Acuity status category 3: The patient meets all three of the following criteria: 1. Procedure status is elective or urgent. 2. Patient had pre-procedure shock, inotropes, or mechanical assist device. 3. No prior cardiac arrest within 24 hours of procedure.	Critical Pre-Operative Status Previous MI Procedure acuity	Yes if Procedure acuity = "Urgent"/ "Elective" AND Critical pre-operative status = "Yes" AND Previous MI is not recorded as 6-24 hours or <6 hours
Acuity status category 4: The patient meets either one or both of the following: 1. Procedure status is emergency or salvage; or 2. Patient had prior cardiac arrest within 24 hours of operation.	Critical Pre-Operative Status Previous MI Procedure acuity	Yes if Procedure acuity = "Emergency"/ "Salvage" OR Previous MI = MI 6-24 hours or MI <6 hours

E-component Table 4: Variable matching between the FRANCE-2 model and the 2013-2017 NHR-TAVI registration

FRANCE-2 Variable	NHR-TAVI	Mapped TAVI Values
Age	Age	Age split into categories as per model
BMI	Weight and Height	((Weight)/(Height^2)) split into categories as per model
Respiratory Insufficiency	History of pulmonary disease	COAD/emphysema, Asthma, Other significant pulmonary disease
Acute pulmonary oedema±	N/A	N/A
NYHA Class IV	NYHA dyspnoea status	Symptoms at rest or minimal activity
Dialysis	On Dialysis	Yes
Pulmonary hypertension	PA Systolic > 60mmHg	Yes
Critical preoperative state	Critical Pre-Operative Status	Yes
Transapical Access	Delivery Approach	Transapical
Other Access	Delivery Approach	Any option other than Transapical or Transfemoral access

± Assume not present for all patients

E-component Table 5: Predictor variables with missing (%) values, baseline characteristics before imputation in the NHR-TAVI cohort (6177 patients)

Variable	Missing values in NHR-TAVI cohort	
	Number of missing	%
estimated Glomerular Filtration Rate (eGFR)	30	0.5
Chronic lung disease	32	0.5
Access route	83	1.3
Critical preoperative state	93	1.5
Body mass index (BMI)	125	2.0
Dialysis	201	3.3
Procedure acuity	211	3.4
Functional NYHA class (I, II, III, and IV)	845	13.7
Systolic Pulmonary Artery Pressure (mm Hg)	2203	35.6
Acute pulmonary oedema	Not registered in the TAVI-NHR cohort	--

E-component Table 6 a: The new estimated new updated intercept of the model FRANCE-2

Variables predictors of the updated model FRANCE 2	Coefficients
New updated intercept from this study, to be used with the original coefficients	-3.84
Original published coefficients (10)	
Intercept (old)	-3.32
Age	0.42
BMI <18	0.82
BMI 18-30	0.41
NYHA class IV	0.58
Acute pulmonary oedema	0.47
SPAP >60 mm Hg	0.37
Critical preoperative state	0.87
Chronic lung disease	0.5
Dialysis	1.06
Transapical access	0.7
Direct aortic, Subclavian or Other access	0.78

Abbreviations: BMI: Body mass index, eGFR: estimated Glomerular Filtration Rate, SPAP: Systolic pulmonary arterial pressure, NYHA: New York Heart Association, TF: Transfemoral.

E-component Table 6 b: The new estimated coefficients for the final updated model ACC-TAVI

Variables predictors of the updated model ACC-TAVI	New coefficients	Std. Error	Std. Error 95%CI		OR	OR 95%CI		P. value	Published coefficients (4)
			2.5 %	97.5 %		2.5 %	97.5 %		
Intercept	-5.93	0.822	-7.57	-4.35	0.00	0.00	0.01	5.360E-13	-4.72976
Age per 5-y increments	0.17	0.049	0.07	0.26	1.18	1.07	1.30	0.001	0.12185
eGFR per 5-U increments	-0.02	0.015	-0.05	0.01	0.98	0.95	1.01	0.247	-0.06933
Dialysis	0.54	0.393	-0.29	1.26	1.72	0.75	3.54	0.168	1.17932
NYHA class IV	0.59	0.182	0.22	0.94	1.80	1.25	2.55	0.001	0.22304
Chronic lung disease	0.29	0.140	0.01	0.56	1.34	1.01	1.75	0.037	0.51084
Procedure access site									
Non-femoral access site	0.84	0.133	0.57	1.10	2.31	1.77	2.99	0.000	0.67347
Procedure Acuity									
Acuity category 2	0.61	0.185	0.23	0.96	1.84	1.26	2.61	0.001	0.4507
Acuity category 3	1.47	0.485	0.42	2.35	4.33	1.52	10.52	0.003	0.99269
Acuity category 4	0.83	0.314	0.16	1.40	2.28	1.17	4.06	0.009	1.20737

Abbreviations: BMI: Body mass index, eGFR: estimated Glomerular Filtration Rate, SPAP: Systolic pulmonary arterial pressure, NYHA: New York Heart Association, TF: Transfemoral.

E-component Table 7: Sensitivity analysis by simulating the values of the absent variable predictor (acute-pulmonary-oedema) and calculated the performance measures of the updated-model in each of 1000 bootstrap samples

Performance measures	FRANCE-2-NHR Model-intercept-update
AU-ROC[#] (95% CI)	0.61 (0.52-0.65)
AU-PRC	0.06
Brier score	0.031
Brier-skill score	-0.04
Calibration-on-the-large "Calibration-intercept" (95% CI)	Not calibrated -0.5 (-0.7 - -0.2)
Logistic-calibration "Calibration slope" (95% CI)	Not calibrated 1.13 (1.04-1.22)

E-component Table 8: TRIPOD Checklist: Prediction Model Validation

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1*
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and **rationale** for developing or validating the multivariable prediction model, including references to existing models.	3
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	3
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	✓
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	✓
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	✓
	5b	Describe eligibility criteria for participants.	✓
	5c	Give details of treatments received, if relevant.	NA
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	✓
	6b	Report any actions to blind assessment of the outcome to be predicted.	✓
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	✓
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	✓
Sample size	8	Explain how the study size was arrived at.	✓
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	✓
Statistical analysis methods	10c	For validation, describe how the predictions were calculated.	✓
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	✓
	10e	Describe any model updating (e.g., recalibration) arising from the validation, if done.	✓
Risk groups	11	Provide details on how risk groups were created, if done.	✓
Development vs. validation	12	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	✓
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	✓
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	✓
	13c	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	✓
Model performance	16	Report performance measures (with CIs) for the prediction model.	✓
Model-updating	17	If done, report the results from any model updating (i.e., model specification, model performance).	✓
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	✓
Interpretation	19a	For validation, discuss the results with reference to performance in the development data, and any other validation data.	✓
	19b	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	✓
Implications	20	Discuss the potential clinical use of the model and implications for future research.	✓
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	✓
Funding	22	Give the source of funding and the role of the funders for the present study.	✓

References:

1) Martin GP, Sperrin M, Mamas MA. Pre-procedural risk models for patients undergoing transcatheter aortic valve implantation. J Thorac Dis. 2018;10(Suppl 30):S3560-S7.
2) 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 J Cardiothorac Surg. 2012;42(4):S1-44.
3) Fang F, Tang J, Zhao Y, He J, Xu P, Faramand A. Transcatheter aortic valve implantation versus surgical aortic valve replacement in patients at low and intermediate risk: A risk specific meta-analysis of randomized controlled trials. PLoS One. 2019;14(9):e0221922.
4) Carnero-Alcazar M, Maroto LC, Cobiella-Carnicer J, Vilacosta I, Nombela-Franco L, Alswies A, et al. Transcatheter versus surgical aortic valve replacement in moderate and high-risk patients: a meta-analysis. Eur J Cardiothorac Surg. 2017;51(4):644-52.
5) Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, et al. EuroSCORE II. Eur J Cardiothorac Surg. 2012;41(4):734-44; discussion 44-5.
6) Roques F, Michel P, Goldstone AR, Nashef SA. The logistic EuroSCORE. Eur Heart J. 2003;24(9):881-2.
7) Michel P, Roques F, Nashef SA, Euro SPG. Logistic or additive EuroSCORE for high-risk patients? Eur J Cardiothorac Surg. 2003;23(5):684-7; discussion 7.
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 Suppl):S23-42.
9) Silaschi M, Conradi L, Seiffert M, Schnabel R, Schon G, Blankenberg S, et al. Predicting Risk in Transcatheter Aortic Valve Implantation: Comparative Analysis of EuroSCORE II and Established Risk Stratification Tools. Thorac Cardiovasc Surg. 2015;63(6):472-8.
10) Iung B, Laouenan 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.
11) 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):46-52.
12) 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 Assessment (IRRMA) study. Int J Cardiol. 2016;215:227-31.
13) Martin GP, Sperrin M, Ludman PF, de Belder MA, Gale CP, Toff WD, et al. Inadequacy of existing clinical prediction models for predicting mortality after transcatheter aortic valve implantation. Am Heart J. 2017;184:97-105.
14) Al-Farra H, Abu-Hanna A, de Mol B, Ter Burg WJ, Houterman S, Henriques JPS, Ravelli, A. C. J. et al. External validation of existing prediction models of 30-day mortality after Transcatheter Aortic Valve Implantation (TAVI) in the Netherlands Heart Registration. Int J Cardiol. 2020.
15) Wolff G, Shamekhi J, Al-Kassou B, Tabata N, Parco C, Klein K, et al. Risk modeling in transcatheter aortic valve replacement remains unsolved: an external validation study in 2946 German patients. Clin Res Cardiol. 2020.
16) Siregar S, Nieboer D, Versteegh MIM, Steyerberg EW, Takkenberg JJM. Methods for updating a risk prediction model for cardiac surgery: a statistical primer. Interact Cardiovasc Thorac Surg. 2019;28(3):333-8.
17) Hickey GL, Grant SW, Caiado C, Dunning J, Poullis M, et al.Dynamic prediction modeling approaches for cardiac surgery.Circ Cardiovasc Qual Outcomes.2013;6(6):649-58.
18) Hickey GL, Grant SW, Murphy GJ, Bhabra M, Pagano D, McAllister K, et al. Dynamic trends in cardiac surgery: why the logistic EuroSCORE is no longer suitable for contemporary cardiac surgery and implications for future risk models. Eur J Cardiothorac Surg. 2013;43(6):1146-52.
19) Genereux P, Head SJ, Van Mieghem NM, Kodali S, Kirtane AJ, Xu K, et al. Clinical outcomes after transcatheter aortic valve replacement using valve academic research consortium definitions: a weighted meta-analysis of 3,519 patients from 16 studies. J Am Coll Cardiol. 2012;59(25):2317-26.
20) Tamburino C, Capodanno D, Ramondo A, Petronio AS, Ettori F, Santoro G, et al. Incidence and predictors of early and late mortality after transcatheter aortic valve implantation in 663 patients with severe aortic stenosis. Circulation. 2011;123(3):299-308.
21) Durand E, Borz B, Godin M, Tron C, Litzler PY, Bessou JP, 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.
22) Mack M, Hamandi M. Why Surgical Risk Algorithms Are Not Predictive of Transcatheter Aortic Valve Replacement Outcomes! Circ Cardiovasc Interv. 2019;12(1):e007560.
23) Vergouwe Y, Nieboer D, Oostenbrink R, Debray TPA, Murray GD, Kattan MW, et al. A closed testing procedure to select an appropriate method for updating prediction models. Stat Med. 2017;36(28):4529-39.
24) Su TL, Jaki T, Hickey GL, Buchan I, Sperrin M. A review of statistical updating methods for clinical prediction models. Stat Methods Med Res. 2018;27(1):185-97.

25) Wilson B, Tran DTT, Dupuis JY, McDonald B. External Validation and Updating of the Cardiac Surgery Score for Prediction of Mortality in a Cardiac Surgery Intensive Care Unit. J Cardiothorac Vasc Anesth. 2019;33(11):3028-34.
26) Lodi-Junqueira L, da Silva JL, Ferreira LR, Goncalves HL, Athayde GR, Gomes TO, et al. In-hospital mortality risk prediction after percutaneous coronary interventions: Validating and updating the Toronto score in Brazil.Catheter Cardiovasc Interv. 2015;86(6):E239-46.
27) Steyerberg E. Clinical prediction models: a practical approach to development, validation, and updating. NewYork NY: Springer; 2009.
28) Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. J Clin Epidemiol. 2008;61(1):76-86.
29) Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Stat Med. 2004;23(16):2567-86.
30) Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98(9):691-8.
31) Netherlands Heart Registration: NHR; 2019 (Cited 2020 June 15). Available from: https://nederlandsehartregistratie.nl/wp-content/uploads/2020/01/NHR-Rapportage-2019-per-spread-230120.pdf.
32) Martin GP, Mamas MA, Peek N, Buchan I, Sperrin M. A multiple-model generalisation of updating clinical prediction models. Stat Med. 2018;37(8):1343-58.
33) Sterne JA, 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:b2393.
34) Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. Second Edition ed. New York: Springer Science & Business Media; 2001 June 15, 2001.
35) Boyd K, Eng KH, Page CD, editors. Area under the Precision-Recall Curve: Point Estimates and Confidence Intervals2013; Berlin, Heidelberg: Springer Berlin Heidelberg.
36) Cox DR. Two further applications of a model for binary regression. Oxford University Press on behalf of Biometrika Trust. 1958;45:562-5 (4 pages).
37) Murphy AH. A New Vector Partition of the Probability Score. . J Appl Meteorol, National Center for Atmospheric Research, Boulder, Colo. 1973:595-600.
38) Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32-5.
39) R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing,Vienna, Austria. 2019.
40) Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.
41) Barili F, Freemantle N, Pilozzi Casado A, Rinaldi M, Folliguet T, Musumeci F, et al. Mortality in trials on transcatheter aortic valve implantation versus surgical aortic valve replacement: a pooled meta-analysis of Kaplan-Meier-derived individual patient data. Eur J Cardiothorac Surg. 2020.
42) Takagi H, Hari Y, Nakashima K, Kuno T, Ando T, Group A. Mortality after transcatheter versus surgical aortic valve replacement: an updated meta-analysis of randomised trials. Neth Heart J. 2020.
43) Jakobsen L. Corrigendum for Jakobsen et al. "Short- and Long-term Mortality and Stroke Risk after Transcatheter Aortic Valve Implantation" Am J Cardiol 2018;121:78-85. Am J Cardiol. 2018;121(3):395.
44) Iqbal J, Vergouwe Y, Bourantas CV, van Klaveren D, Zhang YJ, Campos CM, et al. Predicting 3-year mortality after percutaneous coronary intervention: updated logistic clinical SYNTAX score based on patient-level data from 7 contemporary stent trials. JACC Cardiovasc Interv. 2014;7(5):464-70.
45) Van Calster B, Van Hoorde K, Vergouwe Y, Bobdiwala S, Condous G, Kirk E, et al. Validation and updating of risk models based on multinomial logistic regression. Diagn Progn Res. 2017;1:2.
46) Minne L, Eslami S, de Keizer N, de Jonge E, de Rooij SE, Abu-Hanna A. Statistical process control for validating a classification tree model for predicting mortality--a novel approach towards temporal validation. J Biomed Inform. 2012;45(1):37-44.
47) Minne L, Eslami S, de Keizer N, de Jonge E, de Rooij SE, Abu-Hanna A. Effect of changes over time in the performance of a customized SAPS-II model on the quality of care assessment. Intensive Care Med. 2012;38(1):40-6.
48) Pollari F, Hitzl W, Claes M, Grossmann I, Fischlein T. Machine Learning for Making Aortic Valve Interventions Complementary and Not Competitive. JACC Cardiovasc Interv. 2019;12(20):2112.
49) Steyerberg E. Clinical Prediction Models, A Practical Approach to Development, Validation, and Updating. New York NY: Springer Science & Business Media, LLC; 2009.
50) Rubin DB. Multiple Imputation for Nonresponse in Surveys: Wiley; 2004
51) Barnard J, Rubin DB. Small-Sample Degrees of Freedom with Multiple Imputation. Biometrika. 1999;86(4):948-55.

Chapter 4:

Development and validation of a prediction model for early-mortality after Transcatheter Aortic Valve Implantation (TAVI) based on the Netherlands Heart Registration (NHR): The TAVI-NHR risk model

Catheterization and Cardiovascular Interventions, 2022

https://doi.org/10.1002/ccd.30398

Hatem Al-Farra, Anita C.J. Ravelli, José P.S. Henriques, Saskia Houterman, Bas A.J.M. de Mol, Ameen Abu-Hanna; on behalf of the NHR THI Registration Committee[#]

Abstract

Background

The currently available mortality prediction models (MPM) have suboptimal performance when predicting early-mortality (30-days) following Transcatheter Aortic Valve Implantation (TAVI) on various external populations. We developed and validated a new TAVI-MPM based on a large number of predictors with recent data from a national heart registry.

Methods

We included all TAVI-patients treated in the Netherlands between 2013-2018, from the Netherlands Heart Registration. We used logistic-regression analysis based on the Akaike Information Criterion for variable selection. We multiply imputed missing values, but excluded variables with >30% missing values. For internal-validation, we used ten-fold cross-validation. For temporal (prospective) validation, we used the 2018-dataset for testing. We assessed discrimination by the c-statistic, predicted probability accuracy by the Brier-score, and calibration by calibration-graphs, and calibration-intercept and calibration-slope. We compared our new model to the updated ACC-TAVI and IRRMA MPMs on our population.

Results

We included 9144 TAVI-patients. The observed early-mortality was 4.0%. The final MPM had ten variables, including: critical-preoperative state, procedure-acuteness, body surface area, serum creatinine and diabetes-mellitus status. The median c-statistic was 0.69 (IQR 0.646-0.75). The median Brier-score was 0.038 (IQR 0.038-0.040). No signs of miscalibration were observed. The c-statistic's temporal-validation was 0.71 (95% confidence intervals 0.64-0.78). Our model outperformed the updated currently available MPMs ACC-TAVI and IRRMA (p-value <0.05).

Conclusion

The new TAVI-model used additional variables and showed fair discrimination and good calibration. It outperformed the updated currently available TAVI-models on our population. The model's good calibration benefits pre-procedural risk-assessment and patient counselling.

Introduction

Transcatheter aortic valve implantation (TAVI) was introduced as a treatment of choice for patients suffering from severe symptomatic aortic stenosis who are no candidates for surgical aortic valve replacement (SAVR) or considered to have an increased surgical risk. Heart teams estimate the post-procedural (30-day) early-mortality risk for heart procedures (including TAVI) using general surgical mortality prediction models (MPM) such as the European System for Cardiac Operative Risk Evaluation (EuroSCORE-II, 2012) (1, 2), and the Society of Thoracic Surgeons Predicted Risk of Mortality (STS-PRoM) (3). These surgical MPMs are known to have poor performance in predicting early-mortality on TAVI-populations (4-6). On the other hand, among the currently available TAVI-specific models (TAVI-MPMs), the ACC-TAVI (7) and FRANCE-2 (8) were shown to have, relatively, the best performance when externally-validated on TAVI-populations. However, this performance is still poor (4, 5), even when these MPMs are first updated on various external dataset (9). Therefore, development of a new validated TAVI-MPM in large registries is important. In this paper, we describe and investigate the performance of a newly developed TAVI-MPM using the Netherlands Heart Registration (NHR) cohort.

Unlike earlier model-updates of the currently available TAVI-MPMs, this new TAVI-MPM estimates coefficients for each separate variable anew, and considers additional candidate variable predictors like body surface area (BSA), diabetes mellitus (DM) status, serum creatinine and frailty-status. The currently available TAVI-MPMs use different sets of predictors. These predictors include age, gender, left ventricular ejection fraction (LVEF), body mass index (BMI), procedure-acuteness status and TAVI access-route (transfemoral and non-transfemoral) (7, 8). None of the currently available TAVI-MPMs use frailty-status, DM or BSA as variable predictors in their models.

Using a comprehensive set of evaluation measures, we perform internal-validation, and also temporal (prospective) validation to understand the behaviour of the prediction model over time, which best mimics what the model would face in clinical practice.

Methods

Study population

In the Netherlands, 16 heart-centres submit patients' demographics, clinical characteristics, procedural details mortality-status, complications and follow-up data for several cardiac procedures to the NHR (10). For this study, we obtained the anonymized data on each performed TAVI-procedure from the NHR in the period between January, 1 2013 and December, 31 2018. For each patient, to be included in this study, the outcome early-mortality should be available. Therefore, we could use data of 13 heart-centres out of all 16 heart-centres (see flowchart for patients' selection in e-supplementary figure 1).

Statistical analysis

The primary outcome is the post-procedural (30-day) early-mortality, which is defined as death within 30-days from the TAVI-procedures' date. Each centre obtained mortality data from the Municipal Personal Records Database.

Candidate variables (predictors) were selected based on the currently available MPMs, reviewing literature and availability of the predictors in the NHR-registry. As candidate predictors we used patient demographics, past medical history, and procedural details. These predictors were measured and registered before and during any TAVI-procedure (Table 1 and e-supplement table 1). Continuous predictors are presented as mean (with standard deviation, SD) or median (with interquartile range, IQR) and were compared between early-mortality status groups using Students' t-test or the Mann-Whitney test, respectively. Categorical predictors are presented as counts and percentages, and were compared using the chi-squared test or Fisher exact test as appropriate. A 2-tailed p-value <0.05 was considered statistically significant for all analyses.

For missing values of some potential predictors (e-supplementary table 2), we assumed that they were missing at random, as we have no a priori reason to consider them otherwise. Therefore, five imputation datasets were generated based on the chain equations (MICE) (11). In the imputation models for each predictor, we used all other available candidate predictors including, as recommended, the mortality-status as covariates. We excluded a predictor that has more than 30% missing values in the primary analysis and performed a sensitivity analysis where this predictor was imputed and included.

Prediction model development strategy

To avoid having a model that overfits the data, we applied a variables selection strategy. Specifically, we used stepwise backward elimination based on the Akaike Information Criterion (AIC) (12). The development strategy is applied on the whole dataset for obtaining the final model, but model validation was applied only on the test portions of the datasets, as described in the next section.

The development strategy goes through four steps. Step one: generate five imputed datasets containing no missing values. Step two: on each of the five imputed datasets, fit a logistic-regression model with the outcome and the potential predictors. Predictors were selected with a stepwise backward elimination based on the AIC (12). In this step, we fit the model with all continuous and categorical predictors as defined in the NHR-registry but we re-grouped (re-clustered) some values of three non-binary categorical variables, as following: 1) DM status was re-grouped into two groups: a. the reference category, which includes "No DM" for patients with no DM, together with "DM on medication treatment" for patients who have DM and are using medication treatment (oral medication and / or insulin); and b. "DM without medication treatment" for patients with the status of pre-existing diagnosis of DM that is controlled with diet or life-style modifications only; 2) NYHA class I and class II were grouped in one category, and because this variable reflects a patient's functional capacity, and has a a monotonous association with mortality, we represented it as a continuous variable; and 3) TAVI access-route, which was regrouped into "transfemoral", "subclavian", "transapical" and "direct aortic". We registered the selected predictors from each model on each of the imputed datasets. The predictors that were selected in all five imputed datasets, were considered as the final set of predictors for the prediction strategy.

Step three: refit a logistic-regression model with the selected final predictors on each imputed dataset to predict early-mortality. Step Four: obtain the final prediction model by pooling the estimates (β coefficients), their 95% confidence intervals (CI) and standard errors (SE), and the odds ratio (with their 95% CI, and P-value) across the five models using Rubin's rules (13). The final model is represented by its linear predictor (LP). The predicted probability is calculated from this LP using this formula:

$$Predicted\ probability = \frac{1}{1 + \mathrm{e}^{-\mathrm{LP}}}$$

Internal and prospective-validation

For internal-validation, the entire development strategy was evaluated with a stratified 10-fold cross-validation. In particular, the whole strategy is repeated on the corresponding training datasets and tested only on the test datasets. For example, in the first cross-validation fold, 90% of the dataset is used for the multiple imputation and for deciding on the pooled model of that fold, then this model is evaluated on the tenth test datasets each consisting of the 10% remaining dataset. We use the median to aggregate the performance over the ten test datasets. The training and test datasets were imputed separately.

To check for possible consequences to population (or calibration) drift, we also performed temporal (prospective) validation. This enables understanding the behaviour of the MPM over time, which best mimics what the MPM would face over time in clinical practice. In the prospective validation we developed a model

using the prediction strategy on the dataset of 2013-2017. Then we validated the developed MPM on the imputed dataset of 2018.

Comparison of the new model with the currently available TAVI-models (model update and external validation)

To compare the performance of our new model with the currently available TAVI-specific MPM, we selected two of them. The first selected model is the ACC-TAVI model (7) because in external validation on our population it showed the best relative performance among other models (5). Note that although the FRANCE-2 model also showed a relatively high performance in the Dutch population, it was not selected as we do not have the variable acute pulmonary oedema registered in our cohort. Based on the approach used in (9) for prediction models' update, we updated the ACC-TAVI model (7) with our current cohort.

The second selected model, is the model Israeli TAVR Registry Risk Model Accuracy Assessment (IRRMA) (6). We selected this model as it has very few predictors. Conform the approaches for models' external validation in (4, 5) and for models' update in (9), we externally validated and then updated the model IRRMA (6) with our current cohort.

Consequently, we compared these two updated models' predictive performances with the performance of our new model.

Predictive performance measures

The predictive performance of the model developmental strategy was measured in terms of discrimination with the c-statistic, which is equal to the Area under the Receiver-Operating-Characteristic curve (AU-ROC) (14); the balance between the positive predictive value and sensitivity with the Area Under the Precision-Recall-Curve (AU-PRC) (15, 16); accuracy of predictions with the Brier-score, and the Brier-skill Score (BSS) (17); and calibration with calibration graphs, and the calibration-intercept and calibration-slope based on the Cox approach (14, 18). All performance measures were based on the aggregated cross-validation predictions on the test folds of the five imputed datasets. Details on all these measures appear in e-supplementary text 1.

We also used a forest plot to show the incremental improvement in the AU-ROC of the main prediction strategy. We also constructed a nomogram (19) and provided a computer-based dynamic nomogram graphical interface for the final model, which allows clinicians to easily calculate the linear predictor (LP) and to estimate the TAVI early-mortality risk for each patient. Details appear in e-supplementary text 1.

Sensitivity analysis

We performed various sensitivity analyses. We allowed the AIC predictor selection to use restricted cubic splines (RCS) for continuous predictors; the categorical variables were also provided without re-grouping their values; we allowed predictors to be included in the final model if they appeared three or four out of the five imputed datasets (20). We also applied the same development strategy but using LASSO instead of the AIC for variable selection. Moreover, we also imputed the whole dataset before splitting it to training and testing datasets in the 10-fold cross-validation. Finally, we also included the predictor that has a very large proportion of missing values. Details about all these sensitivity analyses appear in e-supplementary text 2.

All analyses in this study were performed in the R statistical environment (R Foundation for Statistical Computing for Windows V. 3.6.1 (http://www. Rproject.org)) (21). Multiple imputation of the dataset was completed using the mice package (11). We used the package pROC for constructing ROC curves (22). The package rms, DynNom (R Shiny) (23) were used for constructing the nomogram and dynamic nomogram, respectively. The reporting on this model adheres to the TRIPOD checklist for the reporting of multivariable prediction models (24) (e-supplementary table 9).

Results

We included all patients that underwent TAVI (N=9144) in the period between January, 1 2013 and December, 31 2018 from the NHR, after excluding patients with unknown outcome status (N=510) (see flowchart for patients' selection in e-supplementary figure 1).

The observed early-mortality was 4.0% (n=368). Table 1 shows the early-mortality percentages for the predictor variables. The mean age of non-survivors was 80.2 (SD 7.3) years, and 50.6% of the patients were female. Only 9.3% of the procedures were non-electively performed. Most of the procedures were performed via transfemoral access (79.5%). E-supplement table 1 describes the basic characteristics of the total study population (N = 9144), stratified by the early-mortality status after TAVI (survivors = 8776 vs. non-survivors = 368), along with the results of the univariate analysis of each of the variable predictors, before as well after missing value imputation. Early-mortality had a significant univariate positive association with lower BSA (mean 1.8 m2 non-survivals vs. 1.9 m2 in survivals, p-value <0.05), lower BMI (mean 26.5 kg/m2 vs. 27.3 kg/m2, p-value <0.05), lower LVEF (mean 47.3% vs. 50.5%, p-value <0.05) and higher critical-preoperative status (4.2% vs. 0.5%, p-value <0.05) (e-supplement table 1 details the univariate analysis).

Table 1. Early (30-day) mortality rates after TAVI in total population (n=9144), and the univariate analysis of the variable predictors

A. Continuous variable predictors (unit)	Mean (S.D.)* in non-survivors	Univariate analysis* Odds ratio (95% CI), P-value
Age (years)	80.2 (7.3)	1.01 (0.99 - 1.03), 0.22
Body Surface Area (m^2)	1.83 (0.2)	0.27 (0.16 - 0.44), < 0.001
LVEF (%)	47.3 (11.8)	0.98 (0.97 - 0.98), < 0.001
Serum creatinine (µmol/L)	116.7 (69.1)	1.001 (1.00 - 1.002), 0.004
sPAP (mm Hg)	32.9 (12.7)	1.02 (1.01 - 1.03), 0.001

B. Categorical variable predictors	Number of cases (N)* N=9144	No of non-survivors (n) n=368	Mortality risk % (n/N x 100) 4.0%	Univariate analysis* Odds ratio (95% CI), P-value
Gender = Female	4630	190	4.1	1.04 (0.85 – 1.28), 0.69
Gender = Males	4514	178	3.9	1.0 (ref)
Chronic lung disease (yes)	1970	95	4.8	1.30 (1.02 - 1.65), 0.03
Chronic lung disease (No)	7139	267	3.7	1.0 (ref)
Extra-cardiac arteriopathy (yes)	2020	98	4.9	1.32 (1.04 - 1.67), 0.02
Extra-cardiac arteriopathy (No)	7073	263	3.7	1.0 (ref)
Neurological dysfunction (yes)	319	11	3.4	0.92 (0.47 - 1.61), 0.78
Neurological dysfunction (No)	7519	282	3.8	1.0 (ref)
Previous cardiac surgery (yes)	1909	74	3.9	1.01 (0.77 - 1.30), 0.96
Previous cardiac surgery (No)	6987	269	3.9	1.0 (ref)
Critical preoperative state (yes)	59	15	25.4	8.70 (4.64 - 15.43), < 0.001
Critical preoperative state (No)	8991	339	3.8	1.0 (ref)
Recent myocardial infarction (yes)	168	14	8.3	2.27 (1.25 - 3.83), 0.004
Recent myocardial infarction (No)	8870	341	3.8	1.0 (ref)
Dialysis (yes)	110	10	9.1	2.47 (1.20 - 4.53), 0.007
Dialysis (No)	8824	344	3.9	1.0 (ref)
Poor mobility (yes)	565	26	4.6	1.46 (0.94 - 2.18), 0.08
Poor mobility (No)	5789	185	3.2	1.0 (ref)
CCS class IV angina (yes)	193	8	4.1	1.17 (0.52 - 2.25), 0.67
CCS class IV angina (No)	6988	249	3.6	1.0 (ref)
Previous CVA (yes)	1009	40	4.0	0.99 (0.70 - 1.37), 0.96
Previous CVA (No)	8038	321	4.0	1.0 (ref)
Previous aortic valve surgery (yes)	432	26	6.0	1.61 (1.04 - 2.38), 0.02
Previous aortic valve surgery (No)	8289	317	3.8	1.0 (ref)
Previous permanent pacemaker (yes)	731	27	3.7	0.96 (0.63 - 1.40), 0.83
Previous permanent pacemaker (No)	7715	297	3.8	1.0 (ref)
Anaesthesia (yes)	5457	256	4.7	1.79 (1.40 - 2.30), < 0.001
Anaesthesia (No)	3207	86	2.7	1.0 (ref)
Balloon pre-TAVI (yes)	4090	152	3.7	1.02 (0.81 - 1.29), 0.85
Balloon pre-TAVI (No)	4014	146	3.6	1.0 (ref)

PABV (yes)	1127	40	3.5	0.95 (0.67 - 1.33), 0.79
PABV (No)	6681	248	3.7	1.0 (ref)
Procedure weight (2 operations) (yes)	98	3	3.1	0.78 (0.19 - 2.09), 0.68
Procedure weight (2 operations) (No)	8627	335	3.9	1.0 (ref)
Procedure urgency				
Elective (ref)	8097	278	3.4	1.0 (ref)
Urgent	805	61	7.6	2.31 (1.72 – 3.05), < 0.001
Emergency	26	6	23.1	8.44 (6.07 – 19.99), < 0.001
Procedure urgency (continue variable)				2.57 (1.97 – 3.35), <0.001
Diabetes Mellitus (DM) status				
No DM (ref)	6502	267	4.1	1.0 (ref)
DM without medication treatment	394	32	8.1	2.18 (1.46 - 3.15), < 0.001
DM on medication treatment	2056	69	3.3	0.85 (0.65 - 1.11), 0.25
Functional NYHA class				
NYHA class I and II (ref)	3011	78	2.6	1.0 (ref)
NYHA class III	4472	197	4.4	1.73 (1.33 - 2.27), <0.001
NYHA class IV	569	47	8.2	3.39 (2.32 – 4.90), < 0.001
Functional NYHA class (continue variable)				1.81 (1.52 – 2.17), <0.001
Access route				
Access route Transfemoral (ref)	7075	232	3.3	1.0 (ref)
Access route subclavian artery	485	23	4.7	1.47 (0.92 - 2.23), 0.09
Access route Transapical	650	49	7.5	2.40 (1.73 - 3.28), < 0.001
Access route direct aortic	694	49	7.1	2.24 (1.61 - 3.05), < 0.001
Frailty status category	1296	41	N.A.	
Not fragile, category 1 (0-3) (ref)	691	17	2.5	1.0 (ref)
Mild fragile, category 2 (4-5)	367	15	4.1	1.80 (0.87 - 3.70), 0.11
Moderate fragile, category 3 (6-8)	201	8	4.0	1.75 (0.70 - 4.04), 0.21
Severe fragile, category 4 (9-14)	36	1	2.8	1.21 (0.07 - 6.17), 0.86

* Values and frequencies presented in this table were calculated before imputation of the missing data.
Abbreviations: Balloon pre-TAVI = Balloon aortic valvuloplasty prior to date of TAVI; CCS class = Canadian Cardiovascular Society grading of angina pectoris; CVA = cerebrovascular accident; DM = Diabetes mellitus; LVEF = Left Ventricular Ejection Fraction; N.A. = not applicable; NYHA = New York Heart Association functional Classification; PABV = Percutaneous Aortic Balloon Valvuloplasty (TAVI post-dilation), Ref = reference in the univariate analysis; sPAP = systolic Pulmonary Arterial Pressure; VSR = poet myocardial infarction ventricular septal rupture.

Only the predictor frailty-status has >30.0% missing values. Therefore, this predictor was excluded from the model development analysis. All other predictors with missing values were multiply imputed (in five datasets) based on the assumption that data was missing at random (e-supplementary table 2). Two predictors (endocarditis and post-myocardial infarction ventricular septal rupture) had a very low occurrence in our cohort, and had no association with the outcome. Hence, they were excluded from the model development analysis.

In total, 28 variables were included as potential predictors. The AIC predictor selection (where each predictor appeared in all five imputed datasets) resulted in a final model (referred to as TAVI-NHR) with ten predictor variables. Table 2 shows the predictors of the TAVI-NHR, the β coefficients along with their 95% CI, the SE of each of the coefficients, and their odds ratio with 95% CI and P-values.

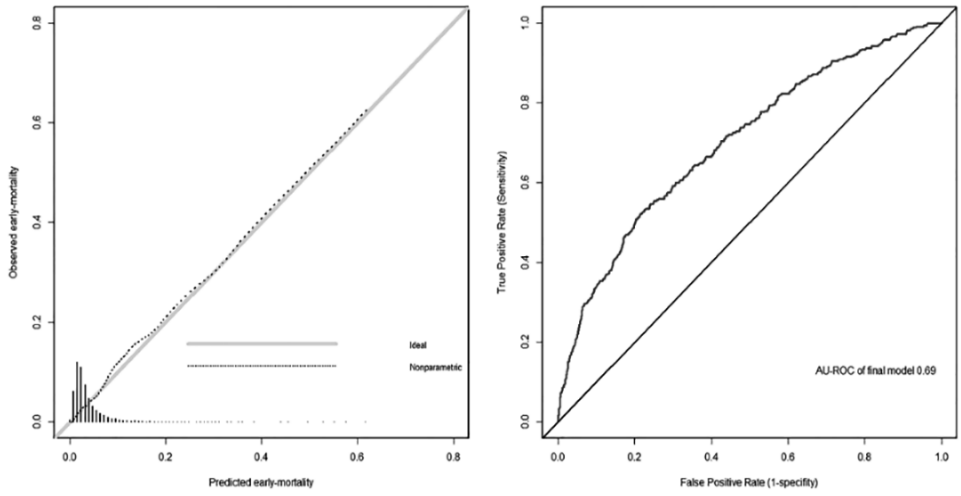
Table 2. The predictor variables of the NHR-TAVI model, with the pooled regression coefficients, standard errors of the regression coefficients, the odds ratios and p-values

Variable predictor	ß Coefficients (95% CI)	SE	Odds ratio (95% CI), P-value
Intercept	-2.28 (-4.17 - -0.39)	0.97	OR: N.A., 0.02
Age (years)	0.02 (0.01 - 0.04)	0.01	1.02 (1.01 - 1.04), 0.01
Serum creatinine (µmol/L)	0.002 (0.001 – 0.003)	0.001	1.001 (1.000 – 1.003), 0.011
Left Ventricular Ejection Fraction (LVEF) (%)	-0.02 (-0.03 - -0.01)	0.01	0.98 (0.97 - 0.99), <0.001
Body surface area (BSA) (m^2)	-1.49 (-2.03 - -0.95)	0.28	0.23 (0.13 - 0.39), <0.001
NYHA class continues (I-II, III, and IV)	0.36 (0.18 – 0.55)	0.09	1.44 (1.19 – 1.73), <0.001
Procedure acuteness (Yes)	0.61 (0.30 – 0.92)	0.15	1.84 (1.36 – 2.51), <0.001
Chronic lung disease (Yes)	0.24 (-0.01 - 0.49)	0.13	1.27 (0.99 - 1.63), 0.06
Critical preoperative state (Yes)	1.64 (0.98 - 2.30)	0.34	5.15 (2.66 – 9.97), <0.001
Diabetes Mellitus without medication treatment (Yes)	0.92 (0.49 - 1.31)	0.21	2.47 (1.64 - 3.71), <0.001
TAVI access route Transfemoral (reference)			1.00
TAVI access route Subclavian artery (Yes)	0.49 (0.05 - 0.94)	0.23	1.64 (1.05 - 2.56), 0.03
TAVI access route Transapical (Yes)	0.90 (0.57 - 1.22)	0.16	2.45 (1.78 - 3.39), <0.001
TAVI access route Direct aortic (Yes)	0.74 (0.41 - 1.07)	0.17	2.09 (1.51 – 2.90), <0.001

Abbreviations: CI = confidence interval; N.A. not applicable; NYHA = New York Heart Association functional Classification; OR = odds ratio; SE = standard errors of the regression coefficients.

In internal-validation with the ten-fold cross-validation of the prediction strategy, the median AU-ROC was 0.69 (IQR 0.64 - 0.75) (Figure 1, right). All models evaluated on the test sets of the cross-validation were well calibrated. The calibration-intercepts and calibration-slopes did not significantly deviate from their ideal values of 0 and 1, respectively. Figure 1 left, shows the calibration plot of the TAVI-NHR, with no signs of miscalibration. Figure 1 (lower part) summarizes the performance of the TAVI-NHR on the test datasets in cross-validation.

Figure 1. Calibration graph (left) and AU-ROC (right) of the final model. Performance measures of the internal validation of the prediction strategy in 10-fold cross-validation, and performance measures of the temporal validation of the prediction strategy on 2018-dataset (n= 2289) (below)



Performance measure	Internal validation (cross-validation)[§]	Temporal validation
	Median (IQR) of the predictive performance on the whole cohort n= 9144	Mean (95% CI) of predictive performance on the 2018 cohort n= 2289
AU-ROC	0.69 (0.64 - 0.75)	0.71 (0.64 - 0.78)
AU-PRC	0.11 (0.08 - 0.14)	0.12 (0.09 - 0.14)
Brier score	0.038 (0.038 - 0.040)	0.027 (0.022 - 0.032)
Brier-skill score	0.009 (-0.004 - 0.039)	0.005 (-0.002 - 0.009)
Calibration-intercept	-0.01 (-0.03 - 0.06)	-0.30 (-0.56 - 0.05)
Calibration-slope	1.00 (0.99 - 1.02)	1.09 (0.98 - 1.17)

Abbreviations: AU-ROC = area under the receiver operating characteristic curve = concordance c-statistic; AU-PRC = area under precision-recall curve; IQR = Interquartile range; 95% CI = 95% confidence intervals

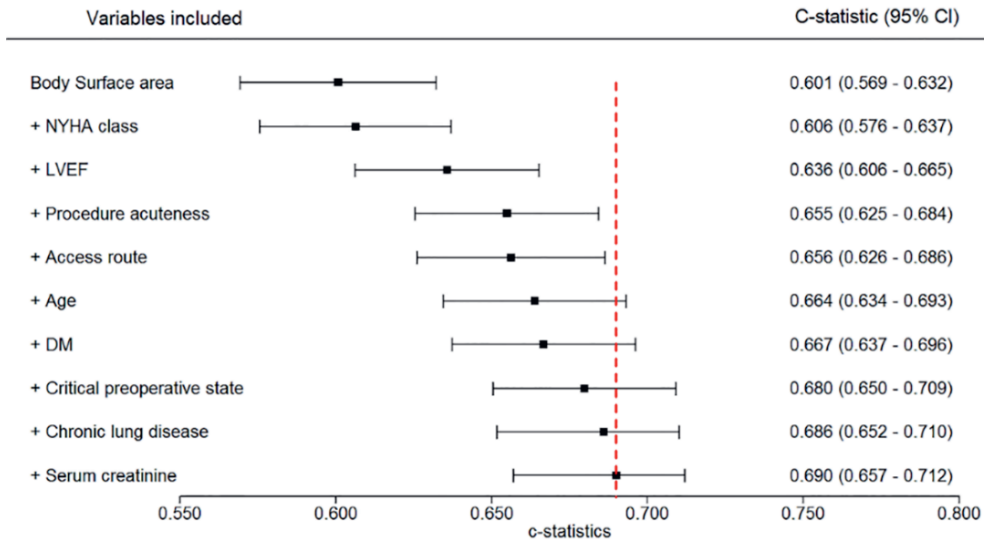
[§] On each of the training-folds, we repeated the prediction strategy (including generating five multiple imputation sets, fitting logistic regression models, selecting variables by stepwise AIC, selecting the variables that appeared five times out of the five imputed datasets, then pooling the coefficients' estimates for the final model). We tested the final model from each fold on the corresponding test-fold. We evaluated the predictive performance in terms of discrimination (AU-ROC), AU-PRC, Brier score, Brier-skill score, calibration-intercept and calibration-slope.

The forest plot (figure 2) shows the incremental improvement in the AU-ROC of the main prediction strategy. The constructed nomogram is presented in e-supplementary figure 2. Also, a computer-based dynamic nomogram graphical interface is available online via the link https://nhr-tavi.shinyapps.io/TAVI-NHR/

In temporal-validation, we developed a model on the cohort 2013-2017 (n=6855), and tested it on the 2018-cohort (n=2289), the AU-ROC was 0.71 (IQR 0.64-0.78). The model was well calibrated with a calibration-intercept and calibration-slope not significantly deviating from 0 and 1, respectively. Figure 1 (lower part) summaries the

performance of the temporal-validation. The median AU-ROC of the internally validated AU-ROC of the updated model ACC-TAVI in cross-validation was 0.64 (IQR 0.60-0.73). Detailed results and calibration graph in e-supplementary table 3 and e-supplementary figure 3.

Figure 2. Forest plot showing the AU-ROC for different models, using different sets of variables from the final selected predictor variables by Akaike Information Criterion (AIC).



Abbreviations: NYHA = New York Heart Association functional Classification; LVEF = Left Ventricular Ejection Fraction; DM = Diabetes mellitus.

The estimated early-mortality risks derived from these models were evaluated in the test sets. Note that the five imputed datasets were stacked in one dataset (where each observation gets a weight of 0.2), and then we split the dataset into training and test sets. The variable list shown builds downwards, adding to the existing variables; for example, the c-statistic (AU-ROC) shown in the fifth row is for the model that includes body surface area, LVEF, NYHA, procedure acuteness status, and access route.

The AU-ROC of the externally validated original IRRMA model was 0.59. The AU-ROC of the internally validated updated IRRMA was 0.60. Detailed results and calibration graphs appear in e-supplementary table 4 and e-supplementary figures 4 and 5.

The predictive performance of our model, in terms of discrimination ability, compared to the performance of these two updated models was statistically significantly higher (p-value <0.05).

Sensitivity analysis showed there were no qualitative changes in the results, with one exception. When the variable frailty-status (86% missing values) was imputed and included as a continuous predictor in the development of a prediction model, it was selected in all five imputed datasets. The median AU-ROC of this model in cross-validation was 0.71 (IQR 0.69-0.80). Results of all other sensitivity analyses appear in e-supplementary (text 2, tables 5, 6, 7A, 7B, 8A, 8B, and 8C).

Discussion

In this study, we developed a new TAVI-MPM for early-mortality (30 days), based on a large TAVI-cohort from the NHR. The TAVI-NHR includes nine predictors. The TAVI-NHR had a fair discrimination with a median AU-ROC of 0.69 (IQR 0.64 - 0.75).

The calibration plot shows that the predicted early-mortality and the observed early-mortality agreed over almost the whole range of probabilities (figure 1, left). When the predicted probability is between 0.15 and 0.35, the prediction slightly underestimates the proportion of observed early-mortality (25, 26).

The model has a statistically significant superior performance to the currently available TAVI-MPMs even after updating them on our population (p-value <0.05). Also, in temporal-validation the AU-ROC of TAVI-NHR was higher than the reported temporally validated AU-ROC of the updated ACC-TAVI and FRANCE-2 in (9). The calibration-intercept and calibration-slope of TAVI-NHR did not deviate from their ideal values in contrast to the updated models ACC-TAVI or FRANCE-2 in (9).

Most of the currently available TAVI-MPMs reported original AU-ROCs in internal-validation ranging from 0.59 to 0.66 (6, 8, 27). External-validation studies for TAVI-MPMs had shown that the currently available models have limited performance on various TAVI-populations (4, 5). In previously published study, the models ACC-TAVI and FRANCE-2 showed the best performance in the Dutch population (5). However, this performance was still limited.

In various update-studies, these TAVI-MPMs showed unsatisfactory performance (limited discriminative ability and miscalibration) (9, 28). Unlike the currently available TAVI-MPMs and their updated versions, the TAVI-NHR has a higher AU-ROC and a good calibration in cross-validation and in temporal validation, and thus better predictions.

Furthermore, in this study, and for comparison, we used our cohort to update the ACC-TAVI (7), and externally validate and update the IRRMA model (6). We noted that the performance of TAVI-NHR is statistically significantly higher than these updated models. Hence, TAVI-NHR has a superior performance in TAVI-population, at least in the Netherlands.

The numbers of the used predictors in the currently available TAVI-MPMs range from four predictors for IRRMA (6) up to 12 in FRANCE-2 (8). The TAVI-NHR has ten predictors, which lies in that range. The prediction strategy used in this study has identified three new predictors, which are not used in the currently available MPMs. These are: LVEF (%), BSA (m2), and DM status (DM without medication treatment).

Several studies (29-33) have shown that frailty-status is a major risk factor for mortality, and adding it to MPMs resulted in improved prediction for early-mortality. The presence of this vulnerable state has been associated with high mortality rate in TAVI-patients (34, 35). In our sensitivity analysis, including the imputed frailty-status, which had a high number of missing values, in the prediction strategy has shown noticeable improvement in the performance in comparison to the TAVI-NHR. Therefore, our study adds to the current evidence that allowing frailty-status as a predictor could improve the performance. However, this requires further assessment in large registries. Routinely collecting the frailty-status and including it in MPMs has hence the potential to improve models and better aid heart teams in risk identification.

Our study has the following strengths. We used a recent and large national sample of >9000 TAVI-patients over 6-years from the NHR. Another strength is that we used AIC with backward selection of potential predictors thus refrained from the common use of predictor selection based on univariate analysis. Our study reports, in addition to discrimination and calibration, also on AU-PRC, Brier-score and BSS. Finally, we performed a range of sensitivity analyses to gain insight into the stability of the results and findings.

The main limitation of this study may stem from the fact that the analysis is based on a registry of routinely collected data. In such studies, in general, data collection and patients' selection for undergoing the procedure may not be standardized among the different participating centres; there is limited data verification; there are missing data in the centres, and cases are not strictly monitored like in prospective studies and randomized controlled studies. However, the quality of the used data in our current study seems to be reasonable, and it had few missing values for almost all the used candidate predictors in the model development analysis. Moreover, the NHR requires standard collection of variables from all the heart centres; the data validity is automatically

checked upon upload in the registry; and the NHR performs annual quality checks and audits of the data. Another limitation is that in our cohort of TAVI-patients, both old and new generation TAVI- prostheses were used. However, this heterogeneous cohort also reflects current clinical practice. Another important limitation of this study is related to the significant proportion of patients with missing values of the variable frailty-status. However, we showed in a sensitivity analysis that Including this variable when imputed has improved the predictive performance of the respective model, with a median AU-ROC in cross-validation of 0.71 (IQR; 0.69-0.80) (details of the other model and the performance measures appear in e-supplementary table 8A, 8B and 8C). Multiple imputation should at least partly reflect the variability of this variable, and the use of multiple imputations for such highly missing variables under the assumption of missing at random may be justified (36). On the other hand, in the daily clinical practices, it seems that this variable were only measured and collected if the patient was at higher-risk for surgical aortic valve replacement. We, therefore, recommend that clinicians and national registries measure this variable for each TAVI-candidate patient.

Our validated MPM may provide useful feedback for heart teams to identify patients who are at high-risk for early-mortality. Our model has a good calibration and may support patient selection and counselling. An implementation of the TAVI-NHR model as a dynamic nomogram TAVI-NHR can be used as an easy-to-use tool (See e-supplementary figure 2). This allows users, at least in the Netherlands, to easily calculate the risk of early-mortality after TAVI. We also provided a computer-based dynamic nomogram graphical interface for the TAVI-NHR. The resulting dynamic nomogram gives (graphically and numerically) the predicted early-mortality risk (and the corresponding 95% CI) for any chosen set of values of the independent predictors.

The evaluation of our model on an independent external TAVI-population, in other countries, merits future work. Based on this experience the necessity of periodically updating the model should be evaluated and we should obtain a better understanding of calibration drift over time. We also advise researchers to develop prediction models for other outcome, especially long-term mortalities after-TAVI, and TAVI-related complications; such as major vascular bleeding, stroke, permanent pacemaker implantation and renal failure. Finally, it seems that the frailty is a promising predictor for early-mortality after TAVI, therefore, the impact of this variable on the model should still be confirmed.

Conclusion

Using a large and recent TAVI-cohort we developed and validated a new TAVI-MPM with improved discrimination and good calibration. The new model (TAVI-NHR) outperformed the currently available TAVI-MPMs and their updated versions. TAVI-NHR and the provided nomogram implementing the TAVI-NHR model may be useful for heart teams during patient counselling for risk-assessment before the TAVI procedure.

Contributors

HF, AAH, AR and BM contributed to the conception, design and coordination of this study. HF performed the statistical analysis, using existing data. HF, AAH, and AR contributed to the analysis. HF wrote the first draft. HF, AR and AAH, contributed to drafting the manuscript. HF, AAH, JH, BM, AR, SH contributed to the critical revision of the article. The final version was approved by all the authors. The first author HF and principal investigators AAH and BM, take primary responsibility for the paper.

Funding

The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Declaration of competing interest

Hatem Al-Farra, Anita CJ Ravelli, José PS Henriques, Saskia Houterman, Bas AJM de Mol, and Ameen Abu-Hanna declare no conflict of interest.

Ethics approval

The study was approved by the institutional review board of the Catharina Hospital (Approval number: W19.270). The used anonymized data conformed to the Declaration of Helsinki principles.

Addendum

The following physicians are the members of the NHR THI Registration Committee. They represent the hospitals that have provided the data for this study. Contact with NHR THI Registration Committee can be established via the e-mail:
info@nederlandsehartregistratie.nl

[#] The NHR THI Registration Committee:

M.M. Vis, Amsterdam University Medical Centres.
P. den Heijer, Amphia Hospital.
L. Timmers, St. Antonius Hospital.
W.A.L. Tonino, Catharina Hospital.
C.E. Schotborgh, Haga Hospital.
V. Roolvink, Isala.
F. Porta, Leeuwarden Medical Centre.
M.G. Stoel, Medisch Spectrum Twente.
S. Kats, Maastricht University Medical Centre.
G. Amoroso, Onze Lieve Vrouwe Gasthuis.
H.W. van der Werf, University Medical Centre Groningen.
P.R. Stella, University Medical Centre Utrecht.
N.M.D.A. van Mieghem, Erasmus University Medical Centre.

Appendix. E-components text, tables and figures

E-supplementary text 1. Definitions of the used predictive performance measures

Discrimination (c-statistic), the area under the receiver operating characteristic (AU-ROC)

Discrimination measures the ability of the MPM to discriminate between patients with the outcome (non-survivals) from those patients without the outcome (survivals). The closer the AU-ROC is to one, the better the discrimination is. We present the ROC curve of the final model and summarize the AU-ROC using the median and over the 10-folds (38).

The area under the precision-recall-curve (AU-PRC)

In an imbalanced dataset where the incidence of the outcome is low, the AU-ROC does not provide insight into the balance between the sensitivity and the PPV (39-41). Therefore, besides the AU-ROC, we also obtain AU-PRC.

The AU-PRC shows the trade-off between the precision and the recall of the MPM using different probability thresholds (42). The terms "recall" and "precision", correspond respectively to the sensitivity and PPV. The closer the AU-PRC is to one, the better the MPM is.

Calibration

Calibration is the degree of agreement between the predicted and observed early-mortality rates across the full probability range. For obtaining the calibration-intercept and calibration-slope according to the Cox approach (43), one fits a logistic regression model for early-mortality using the log-odds of the predictions as the only predictor.

If the predicted probabilities were perfectly calibrated, then the coefficients of the linear predictor of this logistic regression model would be 0 for the calibration-intercept and 1 for the calibration-slope.

Good calibration is observed if the 95% confidence interval (CI) for the calibration-intercept includes zero, and the 95% CI of the calibration-slope includes one. The calibration graph of the final model (from the AIC variable selection strategy) is also presented, based on the average predictions per patient of the final model on the five imputation datasets.

Accuracy of predictions

We measured the accuracy of the predictions with the Brier-score, which is the mean squared error of the predicted early-mortality (18). A Brier-score ranges from zero to one, the lower the better.

For better insight, the Brier-score is converted into the Brier-Skill Score (BSS). The BSS measures the proportional improvement of the predictions over a non-informative reference model that simply predicts the prior probability of the early-mortality for all patients. The maximum value for BSS is 1, which indicates a perfect deterministic prediction, i.e., the model was able to exactly predict the observed outcomes (17). A BSS of zero means that there is no improvement compared to the predictions of the reference model.

Forest plot

We also show the incremental improvement in the AU-ROC of the main prediction strategy via a forest plot. The forest plot shows the improvement in the AU-ROC where the predictors are shown with a decreasing contribution to the AU-ROC. The variable list builds downwards, adding to the existing variables; for example, the AU-ROC shown in the fifth row is for the model that includes NYHA, access route, serum creatinine, LVEF, and body surface area. For this plot, the five imputed datasets were stacked to form one dataset (where each observation gets a weight of 0.2), and then we split a sample for training (75%) and testing (25%), where each patient's observations are all either in the training or test set. The estimated early-mortality risks derived from these models were evaluated in the test set.

Nomogram

Using the approach in (44), we constructed a nomogram to present the final model. A nomogram is a mathematical method or model that is used to predict certain endpoints, such as early-mortality (44), based on several predictors. It is a graphical calculating device that allows clinicians and other users to easily calculate the linear predictor (LP) and the risk of early-mortality after TAVI.

By indicating a predictor's value by a user, the nomogram shows the points associated with this value. The total number of points can then be used to calculate the LP and the associated risk of early-mortality. We also provided a computer-based dynamic nomogram graphical interface for the final model. For this, we used the approach and the materials (R-code) from (45).

E-supplementary text 2. Details about the performed sensitivity analyses and their results

1. Methods used for Sensitivity analysis

A. Model development with variables selection based on AIC and applying restricted cubic splines (RCS)

During model development, the continuous (numeric) variables (such as age, BAS, BMI, and LVEF) were kept the same. However, in sensitivity analysis, when the relationship between the continuous variables and the outcome seemed not monotonous, we used the restricted cubic spline (RCS) transformation to assess the effect on variable selection.

The RCS was used in step two; all the other steps were kept the same. In step two, we fit the logistic regression model using the RCS for the numeric variables with three or more knots, as necessary, to allow for a non-linear relationship between the variable and the log odds of the outcome.

B. Model development with variables selection based on AIC and applying grouping and ungrouping of categorical variables

Unlike the main analysis, we also used categorical variables as they were originally defined in the NHR registration, without regrouping them.

C. Model development with variable selection based on AIC with applying various majority approaches

During model development, we checked the effect of also including variables selected by the AIC in at least three or at least four out of the five imputation datasets (20).

D. Model development with variable selection based on the least absolute shrinkage and selection operator (LASSO)

To assess the effect of using an alternative approach for variable selection other than the AIC, we performed a sensitivity analysis in which we applied the same model development strategy but with the LASSO instead of the AIC. We determined lambda (a hyperparameter that controls the penalty given to model complexity in terms of the number of included predictors) based on 10 cross-validation. The penalty results in shrinking the model coefficients towards 0 and some variables will have their coefficients shrunk to exactly 0 meaning they are not selected in the model selected final predictor variables.

E. Dealing with missing values during model validation

During model validation (internal and temporal-validation) we imputed the training and test sets separately. However, we also performed a sensitivity analysis in which the training and test sets are jointly imputed before being separated. Moreover, we measured the effect of including a variable that has a very large proportion of missing values, on the predictive performance measures.

2. Results of Sensitivity analysis

After applying the transformation with RCS on the continuous variables, the predictive performance of these models observed to be inferior to the performance of the final model. In E-supplementary table 2, the categories of the discretized continuous variables are shown with their (monotonously) increasing mortality risk. In the prediction model we used the continuous to avoid information loss.

Fitting the models with the ungrouped categorical variables observed to have lower predictive performance.

Allowing variables that are selected in at least three or four out of five imputation datasets, resulted in models having 14 and 12 variables, respectively. The AU-ROC of these two models were 0.67 and 0.66, respectively. All the predictive performance of these models in cross-validation is given in E-supplementary table 6.

Performing LASSO for variable selection (when including variables that are selected in at least five, four, or three out of five imputation datasets) resulted in models with 17, 22 and 24 variables, respectively. The median AU-ROC in cross-validation of the model when including variables emerging in all five imputation datasets was 0.68 (IQR; 0.65-0.74). E-supplementary table 7A and 7B show the predictor variables (and their coefficients) and details about the predictive performance of these models.

Imputing the whole dataset before dividing to training and testing sets showed no differences from imputing each partition separately.

In the last sensitivity analysis, the variable frailty-status (with 86% missing values) was imputed and included as a continuous variable in the development of a prediction model. The AIC has selected this variable as it appeared in all five imputed datasets. The median AU-ROC in cross-validation was 0.71 (IQR; 0.69-0.80). E-supplementary tables 8A, 8B and 8C provide details about the selected variables and performance of this sensitivity analysis.

E-supplementary tables

E-supplement table 1. Basic characteristics of the total (N=9144) TAVI-population, stratified by the status of the early 30-day mortality after TAVI (Survivors = 8776, vs. non-Survivors = 368)

Continuous variable predictors (unit)	Overall Mean (SD)	Survivors Mean (SD)	Non- survivors Mean (SD)	Test p-value *	Univariate analysis before imputation Odds ratio (95% CI), P-value	Univariate analysis after imputation Odds ratio (95% CI), P-value
Age (years)	79.80 (6.90)	79.78 (6.88)	80.24 (7.32)	0.215	1.01 (0.99 - 1.03), 0.22	1.01 (0.99 - 1.03), 0.22
Body Surface Area (m^2)	1.89 (0.22)	1.89 (0.22)	1.83 (0.22)	<0.001	0.27 (0.16 - 0.44), < 0.001	0.26 (0.16 – 0.43), <0.001
Body Mass Index (kg/m^2)	27.29 (6.00)	27.32 (6.01)	26.47 (5.57)	0.009	0.97 (0.94 - 0.99), 0.003	0.97 (0.94 - 0.99), 0.003
Serum creatinine (µmol/L)	106.82 (65.57)	106.41 (65.39)	116.66 (69.05)	0.004	1.001 (1.00 - 1.002), 0.004	1.002 (1.00 - 1.003), 0.002
Left Ventricular Ejection Fraction (%)	50.08 (10.52)	50.20 (10.46)	47.15 (11.48)	<0.001	0.98 (0.97 - 0.98), < 0.001	0.98 (0.97 - 0.99), <0.001
Systolic Pulmonary Arterial Pressure (mm Hg)	30.71 (10.74)	30.62 (10.65)	32.92 (12.73)	0.001	1.02 (1.01 - 1.03), 0.001	1.02 (1.01 - 1.03), 0.01
Categorical variable predictors	**Overall (%) † N=9144**	**Survivors (%) ± N=8776**	**Non- survivors (%) ϕ N=368**	**Test p-value ***	**Univariate analysis before imputation Odds ratio (95% CI), P-value**	**Univariate analysis after imputation Odds ratio (95% CI), P-value**
Male (%)	4514 (49.4)	4336 (49.4)	178 (48.4)	0.703	1.0 (ref)	1.0 (ref)
Female (%)	4630 (50.6)	4440 (50.6)	190 (51.6)	0.736	1.04 (0.85 - 1.28), 0.69	1.04 (0.85 - 1.28), 0.69
Chronic lung disease (%)	1970 (21.6)	1875 (21.4)	95 (26.2)	0.035	1.30 (1.02 - 1.65), 0.03	1.30 (1.02 - 1.64), 0.04
Extra-cardiac arteriopathy (%)	2020 (22.2)	1922 (22.0)	98 (27.1)	0.025	1.32 (1.04 - 1.67), 0.02	1.30 (1.03 - 1.65), 0.03
Neurological dysfunction (%)	319 (4.1)	308 (4.1)	11 (3.8)	0.898	0.92 (0.47 - 1.61), 0.78	0.90 (0.52 - 1.56), 0.70
Previous cardiac surgery (%)	1909 (21.5)	1835 (21.5)	74 (21.6)	1.000	1.01 (0.77 - 1.30), 0.96	1.02 (0.78 - 1.32), 0.90
Critical preoperative state (%)	59 (0.7)	44 (0.5)	15 (4.2)	<0.001	8.70 (4.64 - 15.43), < 0.001	9.00 (4.90 - 15.30), < 0.001
Recent myocardial infarction (%)	168 (1.9)	154 (1.8)	14 (3.9)	0.006	2.27 (1.25 - 3.83), 0.004	2.25 (1.29 - 3.92), 0.004
Dialysis (%)	110 (1.2)	100 (1.2)	10 (2.8)	0.011	2.47 (1.20 - 4.53), 0.007	2.35 (1.22- 4.54), 0.01
Poor mobility (%)	565 (8.9)	539 (8.8)	26 (12.3)	0.097	1.46 (0.94 - 2.18), 0.08	1.35 (0.87 - 2.09), 0.17
CCS class IV angina (%)	193 (2.7)	185 (2.7)	8 (3.1)	0.816	1.17 (0.52 - 2.25), 0.67	1.14 (0.46 - 2.87), 0.76
Previous CVA (%)	1009 (11.2)	969 (11.2)	40 (11.1)	1.000	0.99 (0.70 - 1.37), 0.96	0.98 (0.70 - 1.37), 0.92
Previous aortic valve surgery (%)	432 (5.0)	406 (4.8)	26 (7.6)	0.031	1.61 (1.04 - 2.38), 0.02	1.61 (1.05 - 2.47), 0.03
Previous permanent pacemaker (%)	731 (8.7)	704 (8.7)	27 (8.3)	0.913	0.96 (0.63 - 1.40), 0.83	0.98 (0.66 - 1.46), 0.93
Anaesthesia (%)	5457 (63.0)	5201 (62.5)	256 (74.9)	<0.001	1.79 (1.40 - 2.30), < 0.001	1.78 (1.39 - 2.27), < 0.001
Balloon pre-TAVI (%)	4090 (50.5)	3938 (50.4)	152 (51.0)	0.896	1.02 (0.81 - 1.29), 0.85	0.99 (0.75 - 1.30), 0.93
PABV (%)	1127 (14.4)	1087 (14.5)	40 (13.9)	0.855	0.95 (0.67 - 1.33), 0.79	0.97 (0.70 - 1.34), 0.85
Procedure weight (2 operations) (%)	98 (1.1)	95 (1.1)	3 (0.9)	0.876	0.78 (0.19 - 2.09), 0.68	0.72 (0.23 - 2.30), 0.58
Unstable angina pectoris (%)	17 (0.2)	17 (0.2)	0 (0.0)	0.840	0 (N.A.), 0.96	0 (N.A.), 0.96
Thoracic aortic surgery (%)	8 (0.1)	8 (0.1)	0 (0.0)	1.000	0 (N.A.), 0.97	0 (N.A.), 0.98
Post-myocardial infarction VSR (%)	2 (0.0)	0 (0.0)	2 (0.6)	<0.001	0 (N.A.), 0.95	0 (N.A.), 0.94
Procedure urgency				<0.001		
Elective (%)	8097 (90.7)	7819 (91.1)	278 (80.6)		1.0 (ref)	1.0 (ref)
Urgent (%)	805 (9.0)	774 (8.7)	61 (17.7)		2.31 (1.72 – 3.05), < 0.001	2.43 (1.80 – 3.28), <0.001
Emergency (%)	26 (0.3)	20 (0.2)	6 (0.9)		8.44 (6.07 – 19.99), < 0.001	8.18 (3.27 – 20.42), <0.001
Procedure urgency continuous variable					2.44 (1.88 – 3.12), <0.001	2.57 (1.97 – 3.35), <0.001
Diabetes Mellitus (DM) status				<0.001		
No DM (ref) (%)	6502 (72.6)	6249 (72.6)	253 (71.5)		1.0 (ref)	1.0 (ref)
DM without medication treatment (%)	394 (4.4)	362 (4.2)	32 (9.0)		2.18 (1.46 - 3.15), < 0.001	2.19 (1.49 – 3.23), <0.001
DM on medication treatment (%)	2065 (23.0)	1996 (23.2)	69 (19.5)		0.85 (0.65 - 1.11), 0.25	0.84 (0.64 – 1.10), 0.21
Functional NYHA class				<0.001		
NYHA class I (%)	896 (11.1)	871 (11.3)	25 (7.8)		1.0 (ref)	1.0 (ref)
NYHA class II (%)	2115 (26.3)	2062 (26.7)	53 (16.5)		0.90 (0.56 - 1.47), 0.65	0.81 (0.51 – 1.27), 0.36
NYHA class I and II *(combined)* (%)	3011 (37.4)	2933 (37.9)	78 (24.2)		1.0 (ref)	1.0 (ref)
NYHA class III (%)	4472 (55.5)	4275 (55.3)	197 (61.2)		1.61 (1.07 - 2.51), 0.03	1.49 (0.99 – 2.26), 0.06
NYHA class IV (%)	569 (7.1)	522 (6.8)	47 (14.6)		3.14 (1.93 - 5.23), < 0.001	2.95 (1.85 – 4.72), <0.001
Functional NYHA class continuous variable					1.82 (1.51 – 2.19), <0.001	1.81 (1.52 – 2.17), <0.001
Access route				<0.001		
Access route Transfemoral (ref) (%)	7075 (79.5)	6843 (80.0)	232 (65.7)		1.0 (ref)	1.0 (ref)
Access route subclavian artery (%)	485 (5.4)	462 (5.4)	23 (6.5)		1.47 (0.92 - 2.23), 0.09	1.44 (0.93 – 2.23), 0.10
Access route Transapical (%)	650 (7.3)	601 (7.0)	49 (13.9)		2.40 (1.73 - 3.28), < 0.001	2.44 (1.78 – 3.34), < 0.001
Access route direct aortic (%)	694 (7.8)	645 (7.5)	49 (13.9)		2.24 (1.61 - 3.05), < 0.001	2.23 (1.61 – 3.09), <0.001
Frailty status category¤				0.369		
Not fragile, category 1 (0-3) (ref) (%)	691 (53.4)	675 (53.8)	16 (40.0)		1.0 (ref)	1.0 (ref)
Mild fragile, category 2 (4-5) (%)	367 (28.3)	352 (28.0)	15 (37.5)		1.80 (0.87 - 3.70), 0.11	1.79 (0.88 – 3.68), 0.11
Moderate fragile, category 3 (6-8) (%)	201 (15.5)	193 (15.4)	8 (20.0)		1.75 (0.70 - 4.04), 0.21	1.74 (0.74 – 4.15), 0.20
Severe fragile, category 4 (9-14) (%)	36 (2.8)	35 (2.8)	1 (2.5)		1.21 (0.07 - 6.17), 0.86	1.21 (0.16 – 9.37), 0.86

* Test p-value: continuous predictors were tested using Students' t-test. While categorical predictors were compared using the chi-squared test or Fisher exact test as appropriate
† Percentage defined as = number of all cases from the variable predictor/total number of all patients (9144) x 100
± Percentage defined as = number of survivors with present variable predictor/total number of survivors patients (8776) x 100
ϕ Percentage defined as = number of fatalities with present variable predictor/total number of fatalities (368) x 100
¤ Numbers presented here are only for the patients with measured frailty score status (before imputation of the missing data).

Abbreviations: BMI = Body Mass Index; BSA = Body Surface Area; Balloon Pre-TAVI = Balloon aortic valvuloplasty prior to date of TAVI; CCS class = Canadian Cardiovascular Society grading of angina pectoris; CI =confidence interval; CVA = cerebrovascular accident; DM = Diabetes mellitus; N.A. = not applicable; NYHA = New York Heart Association functional Classification; PABV = Percutaneous Aortic Balloon Valvuloplasty (TAVI post-dilation), Ref = reference; VSR = poet myocardial infarction ventricular septal rupture.

E-supplementary table 2. Missing values and percentages of the variable predictors before imputation of the NHR TAVI dataset of 9144 patients

	Variable predictors	Missing#	%
1.	Frailty status	7849	85.8
2.	Poor mobility	2750	30.0
3.	systolic Pulmonary Arterial Pressure	2479	27.1
4.	Canadian Cardiovascular Society grading of angina pectoris	1963	21.5
5.	Percutaneous Aortic Balloon Valvuloplasty (TAVI post-dilation) (PABV)	1336	14.6
6.	Neurological dysfunction	1306	14.3
7.	Functional New York Heart Association functional Classification	1092	11.9
8.	Balloon aortic valvuloplasty prior to date of TAVI	1040	11.4
9.	Previous permanent pacemaker	698	7.6
10.	Endocarditis $	654	7.2
11.	Anaesthesia	480	5.3
12.	Previous aortic valve surgery	423	4.6
13.	Procedure weight (2 operations)	419	4.6
14.	post-myocardial infarction ventricular septal rupture	352	3.9
15.	Thoracic aortic surgery $	335	3.7
16.	Body surface area (m^2)	284	3.1
17.	Previous cardiac surgery	248	2.7
18.	Access route	240	2.6
19.	Procedure acuteness	216	2.4
20.	Dialysis	210	2.3
21.	Unstable angina $	197	2.2
22.	Left Ventricular Ejection Fraction;	195	2.1
23.	Diabetes Mellitus	183	2.0
24.	Recent myocardial infarction;	106	1.2
25.	Previous cerebrovascular accident	97	1.1
26.	Critical preoperative state	94	1.0
27.	Extra-cardiac arteriopathy	51	0.6
28.	Serum creatinine (µmol/L)	35	0.4
29.	Chronic lung disease	35	0.4

$ Not imputed, as all the registered cases were survival, and no incidence of mortality occurred among non-survivals, so technically the imputation of such variables are not possible.
# After imputation all 27 variables were 0 (%) missing values

E-supplementary table 3. Results of the cross-validated predictive performance of model revision ACC-TAVI on TAVI-NHR cohort 2013-2018

Performance measure model revision ACC-TAVI	Value of the validation measure	95% CI
AU-ROC	0.64	0.61-0.73
AU-PRC	0.09	0.064-0.16
Brier score	0.038	0.035-0.048
Brier-skill score	0.001	-0.012-0.026
Calibration-intercept#	-0.012	-0.09-0.28
Calibration-slope#	1.001	0.953-1.105

Calibration-intercepts and -slopes for each model were estimated assuming the slope(s) and intercept(s) equal to one and zero respectively. A satisfactory calibration considered if the 95%CI for the calibration-intercept and-slope included the zero and one, respectively.
Abbreviations: AU-ROC = area under the receiver operating characteristic curve = concordance (c) statistic; AU-PRC = area under precision-recall curve; CI =confidence interval.

E-supplementary table 4. Results of the external validation and the performance of the updated version of the Model revision IRRMA on TAVI-NHR cohort 2013-2018

Performance measure of the external validation IRRMA	Value of the validation measure	95% CI
AU-ROC	0.59	0.54-0.62
AU-PRC	0.09	0.06-0.11
Brier score	0.038	0.033-0.044
Brier-skill score	0.002	-0.01-0.016
Calibration-intercept#	-0.019	-0.09-0.106
Calibration-slope#	1.002	0.961-1.101
Performance measure of the updated version IRRMA		**95% CI**
AU-ROC	0.60	0.56-0.63
AU-PRC	0.07	0.05-0.09
Brier score	0.034	0.031-0.041
Brier-skill score	0.001	-0.007-0.019
Calibration-intercept#	-0.019	-0.09-0.106
Calibration-slope#	1.002	0.961-1.101

Calibration-intercepts and -slopes for each model were estimated assuming the slope(s) and intercept(s) equal to one and zero respectively. A satisfactory calibration considered if the 95%CI for the calibration-intercept and-slope included the zero and one, respectively.
Abbreviations: AU-ROC = area under the receiver operating characteristic curve = concordance *(c) statistic*; AU-PRC = area under precision-recall curve; CI =confidence interval.

E-supplement table 5. Observed early-mortality rates for the continuous variables divided in 3, or 4 subgroups.

Continuous variable predictors*	Total number of cases in each category (N)	Number of non-survivors (n)	Mortality risk % (n/N x 100)
Age (years)			
Age <75	1845	64	3.5
Age 75-80	2654	97	3.7
Age >80	5155	207	4.0
Body Surface Area			
BSA <1.6	734	50	6.8
BSA 1.6 - 2.0	5554	220	4.0
BSA >2.0	2616	81	3.1
Left Ventricular Ejection Fraction			
LVEF >50	5743	34	0.6
LVEF 30 - 50	2696	147	5.5
LVEF <30	553	172	31.1
Serum creatinine (μmol/L)			
Less than 200 (μmol/L)	8878	340	3.83
Between 200 – 400 (μmol/L)	267	17	6.37
More than 400 (μmol/L)	92	5	5.43
Systolic Pulmonary Arterial Pressure			
sPAP > 60 mmHg	151	9	6.0
sPAP 40-60 mmHg	805	43	5.3
sPAP 25-40 mmHg	1307	42	3.2
sPAP ≤25 mmHg	4434	151	3.4

* Values and frequencies presented in this table were calculated before imputation of the missing data

Abbreviations: BSA = Body Surface Area; LVEF = Left Ventricular Ejection Fraction; sPAP = systolic Pulmonary Arterial Pressure.

E-supplementary table 6. Performance measures of the internal validation of the prediction strategy in 10-fold cross-validation, for including predictors appearing in at least 3 or 4 of the 5 imputation datasets

Performance measure	Value of the validation measure	IQR
Performance measure based on majority of 5 voting out 5		
Number of selected variables = 9[$]		
AU-ROC	0.68	0.66-0.72
AU-PRC	0.11	0.08-0.17
Brier score	0.039	0.035-0.049
Brier-skill score	0.001	-0.016-0.045
Calibration-intercept[#]	-0.03	-0.15-0.28
Calibration-slope[#]	1.00	0.95-1.15
Performance measure based on majority of 4 voting out 5		
Number of selected variables = 12[¥]		
AU-ROC	0.67	0.63-0.79
AU-PRC	0.10	0.06-0.26
Brier score	0.039	0.035-0.052
Brier-skill score	-0.004	-0.031-0.074
Calibration-intercept[#]	-0.04	-0.17-0.29
Calibration-slope[#]	0.99	0.95-1.16
Performance measure based on voting selection of 3 out 5		
Number of selected variables = 14[¤]		
AU-ROC	0.66	0.62-0.74
AU-PRC	0.10	0.07-0.16
Brier score	0.038	0.035-0.046
Brier-skill score	-0.002	-0.015-0.029
Calibration-intercept[#]	-0.04	-0.15-0.24
Calibration-slope[#]	0.99	0.95-1.11

[#] Calibration-intercepts and -slopes for each model were estimated assuming the slope(s) and intercept(s) equal to one and zero respectively. A satisfactory calibration considered if the 95%CI for the calibration-intercept and-slope included the zero and one, respectively.

[$] Variables appeared in 5 imputation datasets: age (years), left ventricular ejection fraction, body surface area, chronic lung disease, critical preoperative state, NYHA class: (class II, class III, and class IV), procedure acuteness: (urgent, and emergency), TAVI access route: (subclavian artery, transapical, and direct aortic), and diabetes mellitus status: (diabetes without treatment and Diabetes on treatment).

[¥] Variables appeared in 4 imputation datasets: the above 9 variables, serum creatinine, anaesthesia, and systolic pulmonary arterial pressure.

[¤] Variables appeared in 3 imputation datasets: the above 12 variables, recent MI and poor mobility.

Abbreviations: AU-ROC = area under the receiver operating characteristic curve = concordance (c) statistic; AU-PRC = area under precision-recall curve; IQR = Interquartile range

E-supplementary table 7A. Predictor variables for the TAVI logistic regression model predicting the 30-day early-mortality, variable selection based on LASSO

Variable predictors	Coefficients
Intercept	-2.66
Age (years)	0.02
Body surface area (BSA) (m^2)	-1.09
Serum creatinine (µmol/L)	0.00
Left Ventricular Ejection Fraction (LVEF) (%)	-0.01
sPAP (mm Hg)	0.01
Female (Yes)	-0.07
Diabetes mellitus status	
Diabetes without treatment medication (Yes)	0.72
Diabetes on treatment medication (Yes)	-0.12
Chronic lung disease (Yes)	0.17
Previous cardiac surgery (Yes)	-0.10
Critical preoperative state (Yes)	1.49
Unstable angina pectoris (Yes)	-1.19
Recent myocardial infarction (Yes)	0.39
Thoracic aortic surgery (Yes)	-0.28
Dialysis (Yes)	0.28
Poor mobility (Yes)	0.21
NYHA	
NYHA class II (Yes)	-0.20
NYHA class III (Yes)	0.20
NYHA class IV (Yes)	0.45
CCS class IV angina (Yes)	-0.20
Procedure urgency	
Procedure urgency Urgent (Yes)	0.53
Procedure urgency Emergency (Yes)	1.08
Procedure weight	
Procedure weight (1 operation) (ref)	
Procedure weight (2 operation) (Yes)	-0.34
Previous aortic valve surgery (Yes)	0.28
Previous permanent pacemaker (Yes)	-0.12
Anaesthesia (Yes)	0.24
TAVI access route	
TAVI Access route, Subclavian artery (Yes)	0.23
TAVI Access route, Transapical (Yes)	0.65
TAVI Access route, Direct aortic (Yes)	0.53
Neurological dysfunction (Yes)	-0.14

Abbreviations: CCS class = Canadian Cardiovascular Society grading of angina pectoris; NYHA = New York Heart Association functional Classification; sPAP = systolic Pulmonary Arterial Pressure

E-supplementary table 7B. Internal validation of the prediction strategy based on LASSO

Performance measure	Value of the validation measure	IQR
Performance measure based on voting selection of 5 out 5		
Number of selected variables = 17[$]		
AU-ROC	0.68	0.65-0.74
AU-PRC	0.11	0.08-0.18
Brier score	0.038	0.034-0.05
Brier-skill score	0.006	-0.009-0.039
Calibration-intercept[#]	-0.01	-0.13-0.31
Calibration-slope[#]	1.00	0.94-1.11
Performance measure based on majority of 4 voting out 5		
Number of selected variables = 22[¥]		
AU-ROC	068	0.61-0.84
AU-PRC	0.16	0.09-0.28
Brier score	0.059	0.038-0.083
Brier-skill score	-0.012	-0.023-0.076
Calibration-intercept[#]	0.32	0.01-1.27
Calibration-slope[#]	0.88	0.79-1.09
Performance measure based on majority of 3 voting out 5		
Number of selected variables = 24[¤]		
AU-ROC	0.67	0.64-0.76
AU-PRC	0.11	0.08-0.18
Brier score	0.039	0.037-0.05
Brier-skill score	-0.003	-0.02-0.03
Calibration-intercept[#]	-0.02	-0.16-0.49
Calibration-slope[#]	1.00	0.96-1.13

[#] Calibration-intercepts and -slopes for each model were estimated assuming the slope(s) and intercept(s) equal to one and zero respectively. A satisfactory calibration considered if the 95%CI for the calibration-intercept and-slope included the zero and one, respectively.

[$] Variables appeared in 5 imputation datasets: age (years), body surface area, serum creatinine, left ventricular ejection fraction, female gender, diabetes mellitus status: (diabetes without treatment and Diabetes on treatment), chronic lung disease, critical preoperative state, recent myocardial infarction, poor mobility, NYHA class: (class II, class III, and class IV), procedure acuteness: (urgent, and emergency), procedure weight, previous aortic valve surgery, anaesthesia, TAVI access route: (subclavian artery, transapical, and direct aortic), and previous permanent pacemaker.

[¥] Variables appeared in 4 imputation datasets: the above 17 variables, systolic pulmonary arterial pressure, previous cardiac surgery, unstable angina pectoris, thoracic aortic surgery, CCS class IV angina,

[¤] Variables appeared in 3 imputation datasets: the above 22 variables, dialysis and neurological dysfunction.

Abbreviations: AU-ROC = area under the receiver operating characteristic curve = concordance (c) statistic; AU-PRC = area under precision-recall curve; IQR = Interquartile range

E-supplementary table 8A. Predictor variables and their pooled coefficients, odds ratios, standard errors (SE), and p value for the model after including the variable frailty-status in the five imputed datasets

Variable predictor	Coefficients (95% CI)	SE	OR (95% CI) P-value
Intercept	-2.64 (-4.69 - -0.59)	0.55	1.04 (0.07 - 0.01) 0.012
Age (years)	0.02 (0.00 - 0.03)	1.03	0.01 (1.02 - 1.00) 0.073
Body surface area (BSA) (m^2)	-1.17 (-1.74 - -0.60)	0.55	0.29 (0.31 - 0.18) 0.000
Left Ventricular Ejection Fraction (LVEF) (%)	-0.02 (-0.03 - -0.01)	0.99	0.01 (0.98 - 0.97) 0.002
Critical preoperative state (Yes)	1.95 (1.25 - 2.65)	14.12	0.36 (7.02 - 3.49) 0.000
Procedure acuteness			
Procedure acuteness, Urgent (Yes)	0.48 (0.15 - 0.82)	2.27	0.17 (1.62 - 1.16) 0.005
Procedure acuteness, Emergency (Yes)	1.58 (-0.13 - 3.29)	26.92	0.79 (4.86 - 0.88) 0.068
Frailty status score	0.19 (0.07 - 0.30)	1.36	0.05 (1.21 - 1.07) 0.007
Diabetes mellitus status			
Diabetes without treatment medication (Yes)	1.01 (0.56 - 1.46)	4.33	0.23 (2.75 - 1.75) 0.000
Diabetes on treatment medication (Yes)	-0.23 (-0.55 - 0.09)	1.09	0.16 (0.79 - 0.58) 0.154
TAVI access route			
TAVI Access route, Subclavian artery (Yes)	0.47 (-0.03 - 0.97)	2.65	0.25 (1.60 - 0.97) 0.066
TAVI Access route, Transapical (Yes)	0.88 (0.49 - 1.28)	3.59	0.20 (2.42 - 1.63) 0.000
TAVI Access route, Direct aortic (Yes)	0.76 (0.40 - 1.12)	3.07	0.18 (2.14 - 1.49) 0.000

Abbreviations: CI =confidence interval; OR = odds ratios, SE = standard errors.

E-supplementary table 8B. Predictor variables and their pooled coefficients, odds ratios, standard errors (SE), and p value for the model after including the variable frailty-status in the five imputed datasets from the dataset that include the patients with complete frailty-status score (n = 1295)

Variable predictor	Coefficients (95% CI)	SE	OR (95% CI) P-value
Intercept	-0.76 (-2.10 - 0.58)	1.78	0.67 (0.47 - 0.12) 0.260
Left Ventricular Ejection Fraction (LVEF) (%)	-0.02 (-0.03 - -0.01)	0.99	0.01 (0.98 - 0.97) 0.001
Critical preoperative state (Yes)	2.24 (1.56 - 2.92)	18.46	0.34 (9.38 - 4.76) 0.000
Recent myocardial infarction (Yes)	0.28 (-0.39 - 0.95)	2.58	0.34 (1.32 - 0.68) 0.410
CCS grade IV (Yes)	-0.09 (-1.19 - 1.00)	2.72	0.50 (0.91 - 0.31) 0.855
Frailty status score	0.20 (0.09 - 0.31)	1.37	0.05 (1.23 - 1.10) 0.003
Body surface area (BSA) (m^2)	-1.34 (-1.88 - -0.80)	0.45	0.28 (0.26 - 0.15) 0.000
Diabetes mellitus status			
Diabetes without treatment medication (Yes)	0.96 (0.50 - 1.43)	4.17	0.23 (2.62 - 1.65) 0.000
Diabetes on treatment medication (Yes)	-0.20 (-0.51 - 0.11)	1.12	0.16 (0.82 - 0.60) 0.202

Abbreviations: CCS class = Canadian Cardiovascular Society grading of angina pectoris; CI =confidence interval; OR = odds ratios, SE = standard errors.

E-supplementary table 8C. Results of the sensitivity analysis in which we included the variable frailty-status. The performance measures of the internal validation of this prediction strategy in 10-fold cross-validation of these two models

Performance measure on the 5 imputed datasets including frailty score	Value of the validation measure	IQR
Number of selected variables = 8		
AU-ROC	0.71	0.69-0.80
AU-PRC	0.11	0.08-0.17
Brier score	0.039	9.034-0.046
Brier-skill score	-0.002	-0.019-0.035
Calibration-intercept[#]	-0.02	-0.14-0.26
Calibration-slope[#]	0.99	0.96-1.09
Performance measure on the datasets with measurement frailty score n = 1295		**IQR**
Number of selected variables = 7		
AU-ROC	0.77	0.66-0.91
AU-PRC	0.32	0.22-0.58
Brier score	0.033	0.021-0.064
Brier-skill score	0.13	0.07-0.30
Calibration-intercept#	0.04	-0.21-0.10
Calibration-slope#	1.00	0.81-1.44

On each of the inner-folds, we repeated the prediction strategy (including generating five multiple imputation sets, fitting generalized logistic regression models, variable selection by stepwise Akaike information criterion (AIC), applying the majority voting, then pooling the estimate coefficients for the final model). The final model from each fold is tested on each of the outer-folds. The predictive performance was assessed in terms of discrimination, AU-PRC, Brier score, Brier skill score, calibration-intercept and calibration-slope.

[#] Calibration-intercepts and -slopes for each model were estimated assuming the slope(s) and intercept(s) equal to one and zero respectively. A satisfactory calibration considered if the 95%CI for the calibration-intercept and-slope included the zero and one, respectively.

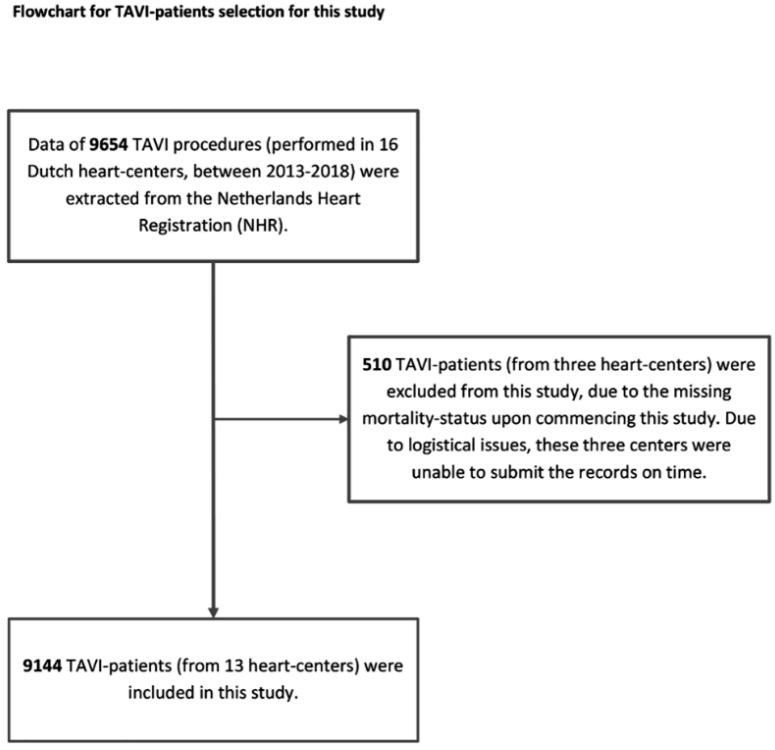
Abbreviations: AU-ROC = area under the receiver operating characteristic curve = concordance *(c) statistic*; AU-PRC = area under precision-recall curve; IQR=Interquartile range.

E-supplementary table 9. Research Reporting Guideline checklist TRIPOD: Prediction model development and validation

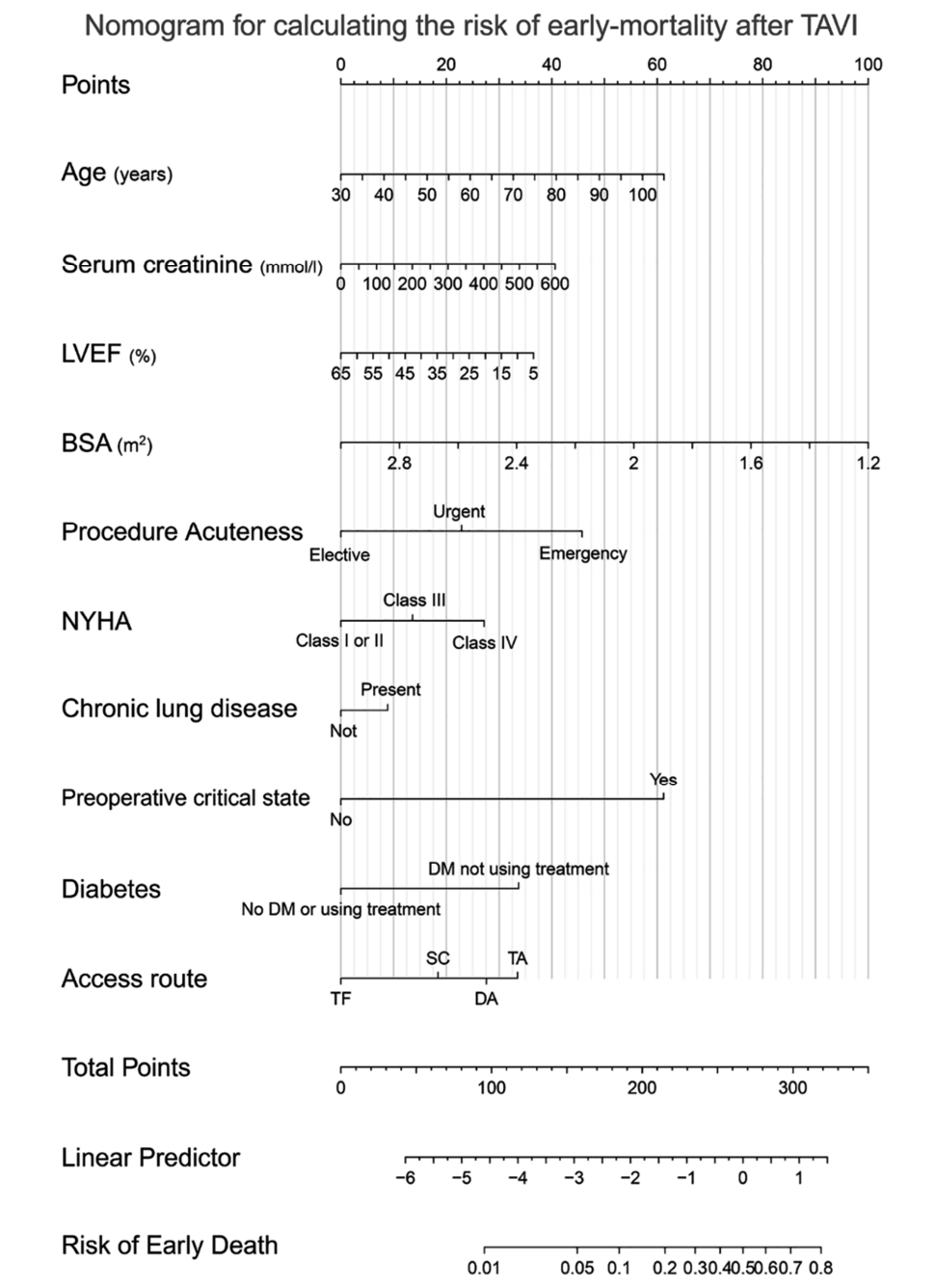
Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1*
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and **rationale** for developing or validating the multivariable prediction model, including references to existing models.	3
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	3
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	✓
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	✓
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	✓
	5b	Describe eligibility criteria for participants.	✓
	5c	Give details of treatments received, if relevant.	NA
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	✓
	6b	Report any actions to blind assessment of the outcome to be predicted.	✓
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	✓
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	✓
Sample size	8	Explain how the study size was arrived at.	✓
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	✓
Statistical analysis methods	10c	For validation, describe how the predictions were calculated.	✓
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	✓
	10e	Describe any model updating (e.g., recalibration) arising from the validation, if done.	✓
Risk groups	11	Provide details on how risk groups were created, if done.	✓
Development vs. validation	12	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	✓
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	✓
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	✓
	13c	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	✓
Model performance	16	Report performance measures (with CIs) for the prediction model.	✓
Model-updating	17	If done, report the results from any model updating (i.e., model specification, model performance).	✓
Discussion			
Limitations	18	Discuss any limitations of the study (such as non-representative sample, few events per predictor, missing data).	✓
Interpretation	19a	For validation, discuss the results with reference to performance in the development data, and any other validation data.	✓
	19b	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	✓
Implications	20	Discuss the potential clinical use of the model and implications for future research.	✓
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	✓
Funding	22	Give the source of funding and the role of the funders for the present study.	✓

E-supplementary figures

E-supplementary figure 1. Flow chart for TAVI-patients' selection for this study

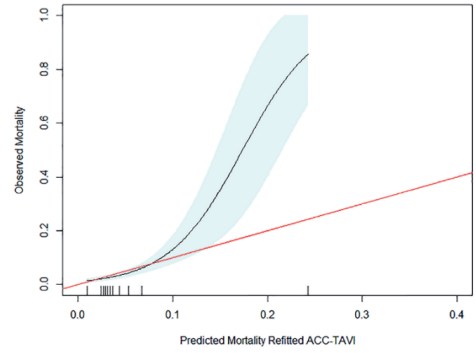


E-supplementary figure 2. Nomogram for early-mortality risk calculation after TAVI based on the NHR cohort 2013-2018

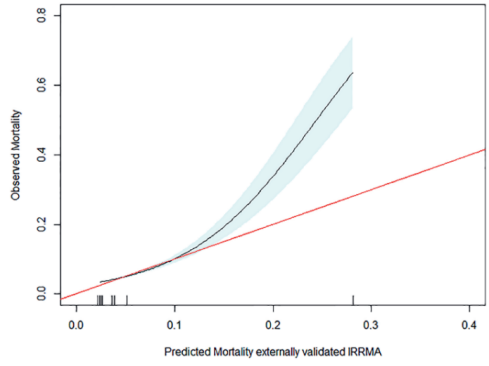


Abbreviations: LVEF = Left Ventricular Ejection Fraction; NYHA = New York Heart Association functional Classification; BSA = body surface area.
In this nomogram, a user can indicate a predictor's value on the graph and the points associated with this value. The total number of points can then be used to calculate the linear predictor (LP) and the associated risk of early-mortality.

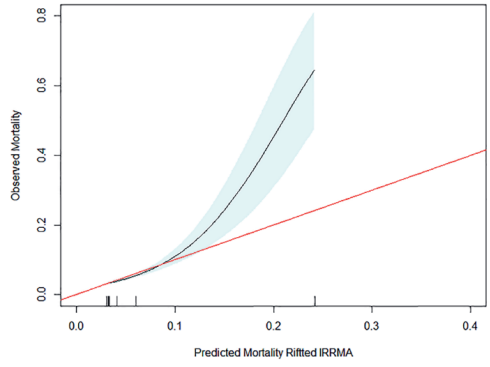
E-supplementary figure 3. Calibration plot of the updated and refitted ACC-TAVI (model revision) on TAVI-NHR cohort 2013-2018



E-supplementary figure 4. Calibration plot of the external validation the model IRRMA on TAVI-NHR cohort 2013-2018



E-supplementary figure 5. Calibration plot of the updated and refitted IRRMA (model revision) on TAVI-NHR cohort 2013-2018



References:

1) Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, et al. EuroSCORE II. Eur J Cardiothorac Surg. 2012;41(4):734-44; discussion 44-5.
2) Michel P, Roques F, Nashef SA, Euro SPG. Logistic or additive EuroSCORE for high-risk patients? Eur J Cardiothorac Surg. 2003;23(5):684-7; discussion 7.
3) 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 Suppl):S23-42.
4) Martin GP, Sperrin M, Ludman PF, de Belder MA, Gale CP, Toff WD, et al. Inadequacy of existing clinical prediction models for predicting mortality after transcatheter aortic valve implantation. Am Heart J. 2017;184:97-105.
5) Al-Farra H, Abu-Hanna A, de Mol B, Ter Burg WJ, Houterman S, Henriques JPS, et al. External validation of existing prediction models of 30-day mortality after Transcatheter Aortic Valve Implantation (TAVI) in the Netherlands Heart Registration. Int J Cardiol. 2020.
6) 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 Assessment (IRRMA) study. Int J Cardiol. 2016;215:227-31.
7) 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):46-52.
8) Iung B, Laouenan 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.
9) Al-Farra H, de Mol B, Ravelli ACJ, Ter Burg W, 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). Int J Cardiol Heart Vasc. 2021;32:100716.
10) Netherlands Heart Registration: NHR; 2019 (Cited 2020 June 15). Available from: https://nederlandsehartregistratie.nl/wp-content/uploads/2020/01/NHR-Rapportage-2019-per-spread-230120.pdf.
11) Van Buuren SG-O, K. . Mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011;45(3):1-67.
12) Sauerbrei W. The Use of Resampling Methods to Simplify Regression Models in Medical Statistics. Journal of the Royal Statistical Society: Series C (Applied Statistics). 1999;48(3):313-29.
13) Barnard J, Rubin DB. Small-Sample Degrees of Freedom with Multiple Imputation. Biometrika. 1999;86(4):948-55.
14) Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361-87.
15) Ozenne B, Subtil F, Maucort-Boulch D. The precision--recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. J Clin Epidemiol. 2015;68(8):855-9.
16) Boyd K, Eng KH, Page CD, editors. Area under the Precision-Recall Curve: Point Estimates and Confidence Intervals2013; Berlin, Heidelberg: Springer Berlin Heidelberg.
17) Brier G. Verification of Forecasts Expressed in Terms of Probability. Monthly Weather Review. 1950;78:1-3.
18) Cox DR. Two further applications of a model for binary regression. Oxford University Press on behalf of Biometrika Trust. 1958;45:562-5 (4 pages).
19) Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol. 2008;26(8):1364-70.
20) Rubin DB. Multiple Imputation for Nonresponse in Surveys. . New York: John Wiley and Sons; 1987.
21) R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing,Vienna, Austria. 2019.
22) Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.
23) Jalali A, Alvarez-Iglesias A, Roshan D, Newell J. Visualising statistical models using dynamic nomograms. PLoS One. 2019;14(11):e0225253.
24) Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.
25) Steyerberg E. Clinical Prediction Models, A Practical Approach to Development, Validation, and Updating. New York NY: Springer Science & Business Media, LLC; 2009.
26) Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014;35(29):1925-31.
27) 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.
28) Martin GP, Mamas MA, Peek N, Buchan I, Sperrin M. A multiple-model generalisation of updating clinical prediction models. Stat Med. 2018;37(8):1343-58.

29) Shimura T, Yamamoto M, Kano S, Kagase A, Kodama A, Koyama Y, et al. Impact of the Clinical Frailty Scale on Outcomes After Transcatheter Aortic Valve Replacement. Circulation. 2017;135(21):2013-24.
30) Rogers T, Alraies MC, Moussa Pacha H, Bond E, Buchanan KD, Steinvil A, et al. Clinical Frailty as an Outcome Predictor After Transcatheter Aortic Valve Implantation. Am J Cardiol. 2018;121(7):850-5.
31) Kiani S, Stebbins A, Thourani VH, Forcillo J, Vemulapalli S, Kosinski AS, et al. The Effect and Relationship of Frailty Indices on Survival After Transcatheter Aortic Valve Replacement. JACC Cardiovasc Interv. 2020;13(2):219-31.
32) Martin GP, Sperrin M, Ludman PF, deBelder MA, Gunning M, Townend J, et al. Do frailty measures improve prediction of mortality and morbidity following transcatheter aortic valve implantation? An analysis of the UK TAVI registry. BMJ Open. 2018;8(6):e022543.
33) Schoenenberger AW, Moser A, Bertschi D, Wenaweser P, Windecker S, Carrel T, et al. Improvement of Risk Prediction After Transcatheter Aortic Valve Replacement by Combining Frailty With Conventional Risk Scores. JACC Cardiovasc Interv. 2018;11(4):395-403.
34) Green P, Woglom AE, Genereux P, Maurer MS, Kirtane AJ, Hawkey M, et al. Gait speed and dependence in activities of daily living in older adults with severe aortic stenosis. Clin Cardiol. 2012;35(5):307-14.
35) Finn M, Green P. Transcatheter aortic valve implantation in the elderly: who to refer? Prog Cardiovasc Dis. 2014;57(2):215-25.
36) Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. J Clin Epidemiol. 2019;110:63-73.
37) Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29-36.
38) Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 2015;10(3):e0118432.
39) Ozenne B, Subtil F, Maucort-Boulch D. The precision--recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. J Clin Epidemiol. 2015;68(8):855-9.
40) Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd international conference on Machine learning; Pittsburgh, Pennsylvania, USA. 1143874:
41) ACM; 2006. p. 233-40.
42) Boyd K, Eng KH, Page CD, editors. Area under the Precision-Recall Curve: Point Estimates and Confidence Intervals2013; Berlin, Heidelberg: Springer Berlin Heidelberg.
43) Murphy AH. A New Vector Partition of the Probability Score. . J Appl Meteorol, National Center for Atmospheric Research, Boulder, Colo. 1973:595-600.
44) Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol. 2008;26(8):1364-70.
45) Jalali A, Alvarez-Iglesias A, Roshan D, Newell J. Visualising statistical models using dynamic nomograms. PLoS One. 2019;14(11):e0225253.

Chapter 5:

Incidence and trends of major adverse cardiac events and mortality after Transcatheter Aortic Valve Implantation (TAVI): analysis by age and operative risk groups

Submitted for publication

Hatem Al-Farra, Anita C.J. Ravelli, Ameen Abu-Hanna, Saskia Houterman, Bas A.J.M. de Mol, José P.S. Henriques; on behalf of the NHR THI Registration Committee[#]

Abstract

Background

Despite its minimal invasiveness, transcatheter aortic valve implantation (TAVI) is associated with major adverse cardiac events (MACE) and mortality. We describe the incidences and trends of TAVI-related MACE and mortality among different age and operative-risk groups.

Methods

We performed a retrospective analysis of TAVI-patients from the Netherlands Heart Registration (NHR) between 2013-2018. MACE was defined as the occurrence of major vascular bleeding (MVB), permanent pacemaker implantation (PPI), and stroke. Mortality at 30-days (early) and 1-year were reported. We calculated incidences and trends of adverse outcomes and compared these to recently reported incidences (NHR annual-report 2021). Age groups were divided into three fixed groups (<75, 75-80, and >80 years) and age tertiles. Risk-groups consisted of EuroSCORE-II tertiles.

Results

This cohort consists of 9144 TAVI-patients, with 60% being ≥80 years old. The number of TAVI-procedures significantly increased over years from 786 to 2289 interventions. MACE incidences were: PPI (10.8%), MVB (3.2%), and stroke (1.7%). Early-mortality was 4.0% and 1-year mortality was 11.6%. Mortality significantly decreased over time. MACE incidences remained stable. The incidences of mortality, MVB, and stroke were the highest in patients >80 years, and in the higher EuroSCORE-II tertile. However, PPI occurred more often in the 75-80 age group and the lower EuroSCORE-II tertile.

Conclusion

In the Netherlands, although TAVI-procedures have increased over time, the mortality incidence significantly decreased, while the MACE incidence remained stable. Most MACEs occur in older and high-operative-risk patients. However, younger and lower operative-risk patients were more at risk for PPI.

Introduction

Transcatheter aortic valve implementation (TAVI) has become the standard treatment for elderly-patients with aortic stenosis and increased surgical-risk (1, 2). Over the last few years, the number of TAVI-procedures has increased as the indication for TAVI expanded from only higher to intermediate and intermediate-low operative-risk. The expanded use of TAVI might be associated with more adverse cardiac events. These adverse outcomes can result in prolonged hospital-stay and higher costs. Major adverse cardiac events (MACE) after TAVI have been associated with mortality and with a negative impact on the quality of life (3). According to the contemporary Valve Academic Research Consortium-2 criteria, TAVI-related MACE includes the need for permanent pacemaker implantation (PPI), minor vascular bleeding, major vascular bleeding (MVB), stroke, renal failure, aortic regurgitation and mortality (3-5). Previous data have shown that the incidence of PPI varies between 20-40% (6-10). The reported incidence of MVB varies from 4% to 6.3% (11-15). TAVI-related stroke varies between 0.6-6.9% (2, 16-20). Despite its low incidence, stroke is a devastating complication and is associated with poor prognosis and perhaps most important in these patients a low quality of life. Most TAVI-related MACE studies describe MACE and mortality rates in the total TAVI-population. Specific studies that focused on detailed incidences among different age groups or different perioperative-risk groups are scarce. Outcome data that is more age specific would provide more insight into MACEs and mortality in various subgroups of patients. In addition, it would support better individual patient-tailored information and decision-making. Each of these adverse outcomes may also have a different impact on the individual patient, especially after extending TAVI to younger patients and patients at lower operative-risk. This study has two aims. The first aim is to describe incidences and trends of TAVI-related MACEs (major vascular bleeding, pacemaker implantation, and stroke) and mortality in the Netherlands between 2013-2018, and to compare these results with the incidences presented by the recent NHR annual report (the year 2019-2020). The second aim is to specifically investigate the incidences of these TAVI-related MACEs among the various age groups, and perioperative-risk groups of EuroSCORE-II.

Methods

Study design and population

This was a retrospective study in a national cohort using prospectively entered data from the Netherlands Heart Registration (NHR) (21). In the Netherlands, 16 heart centers perform TAVI procedures. Data entry to the NHR is obligatory. These data include demographics, clinical characteristics, intervention risk factors, procedural details, mortality status, complications, and follow-up data after hospital-discharge. For this study, we obtained anonymized data on each performed TAVI-procedure from the NHR in the period between January 1, 2013, and December 31, 2018. For each patient to be included in the current study, mortality status was required. Data from three of the 16 centers were excluded from this study, because of failure to report mortality outcomes. In addition, we compared the outcomes of our analysis with the recently published findings (the years 2019 and 2020) presented in the NHR annual-report 2021.

Measured outcomes

TAVI-related MACEs were defined according to the Valve Academic Research Consortium (5). We focused on the incidence and trends of the following adverse outcomes that were available in the NHR. These were: 1) need for PPI within the first 30-days, 2) occurrence of MVB within the first 30-days, 3) occurrence of stroke within the first 72-hours, 4) early (30-day) mortality, and 5) 1-year mortality.

Statistical analysis

First, the incidences of the MACEs and mortality were calculated. Then the incidences were analyzed in three consecutive two calendar-years (2013-2014, 2015-2016, and 2017-2018). After that, the trends over the years were calculated. We tested the trends of the TAVI-related MACEs and mortality using the Cochran Armitage test for trends (22). In addition, we compared the incidence data of these MACEs and mortality with the data that

was recently published in the NHR annual-report 2021 (which represents the incidences up to the year 2020) to compliment the trends tested in this study. Subsequently, we studied the incidences of MACE and mortality over years in the different age groups. For the age-group analysis, two different age groupings were used. Firstly, dividing patients into three fixed age groups: patients <75 years, between 75-80 years, and >80 years. Secondly, all patients were divided into three age tertiles. In addition to the age group analysis, we also used age as a continuous variable and plotted the incidences of each of the adverse outcomes in a graph with age on the horizontal axis. Moreover, the incidences of adverse outcomes in age tertiles were also plotted. The last analysis was performed in the various perioperative risk-groups, which were divided into three EuroSCORE-II tertiles.

For the baseline characteristics analysis, we used variables about patient demographics, past medical history, and procedural details. Continuous variables are summarized as mean (with standard deviation) for the normally distributed predictors, or median (interquartile range) for the non-parametric predictors; and they were compared using the Students' t-test or the Mann-Whitney test, as appropriate. Categorical variables were summarized as frequencies and percentages and were compared using the chi-squared test or Fisher exact test, as appropriate. A p-value <0.05 of a 2-tailed test was considered significant for all analyses. All statistical analyses were performed in R statistical environment version 3.5.1 (23).

Results

The total NHR-TAVI cohort between 2013-2018 consisted of 9654 patients. After excluding patients from three centers with unknown mortality outcomes (N=510), the study population, with known mortality outcomes, comprised 9144 TAVI-patients. See the flowchart for the patients' selection in e-supplementary figure 1. E-supplementary table 1 shows the baseline characteristics of the study population. The number of TAVI-procedures increased significantly over the years from 786 in 2013 to 2289 in 2018 (figure 1 A and B). Most of the TAVI patients (60%) were >80 years old. However, a non-significant increase in the percentage of patients <80 years was observed over time, from 38% in 2013 to 42% in 2018, (details shown in e-supplementary tables 2b and 4). The incidences of TAVI-related MACEs were as follows: the need for PPI within 30-days was (10.8%), the 30-days MVB (3.2%), and stroke within 72-hours (1.7%) (Table 1). MACE rates remained steady over years. The incidence of early-mortality was 4.0%, which significantly decreased over the years from 7.8% to 2.8%. The 1-year mortality also significantly declined from 17.3% to 8.3% over the years (details appear in table 1, figure 2, e-supplementary table 2a and 2b, and e-supplementary figure 3).

Figure 1. A. Total numbers and percentages of <u>all</u> TAVI procedures performed in the Netherlands in each calendar year (N= 9654); and B. The total numbers and percentages of <u>only</u> the performed TAVI procedures with registered mortality outcomes in each calendar year (N= 9144)

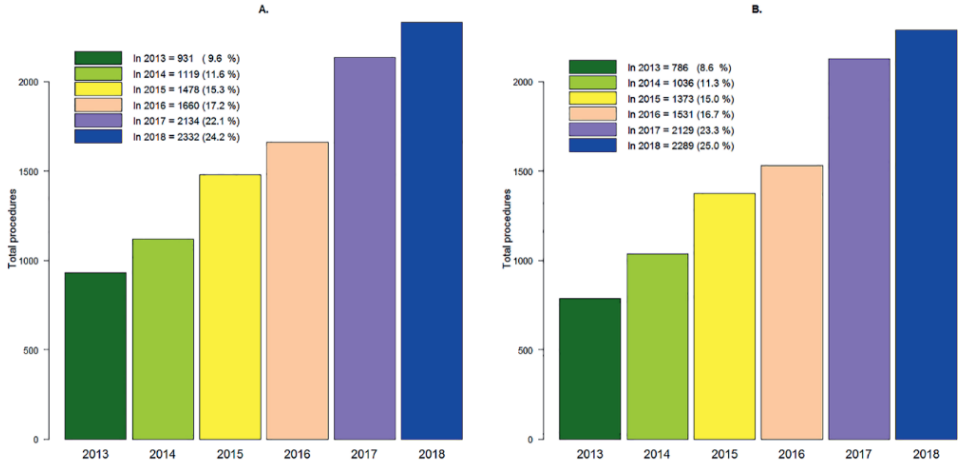
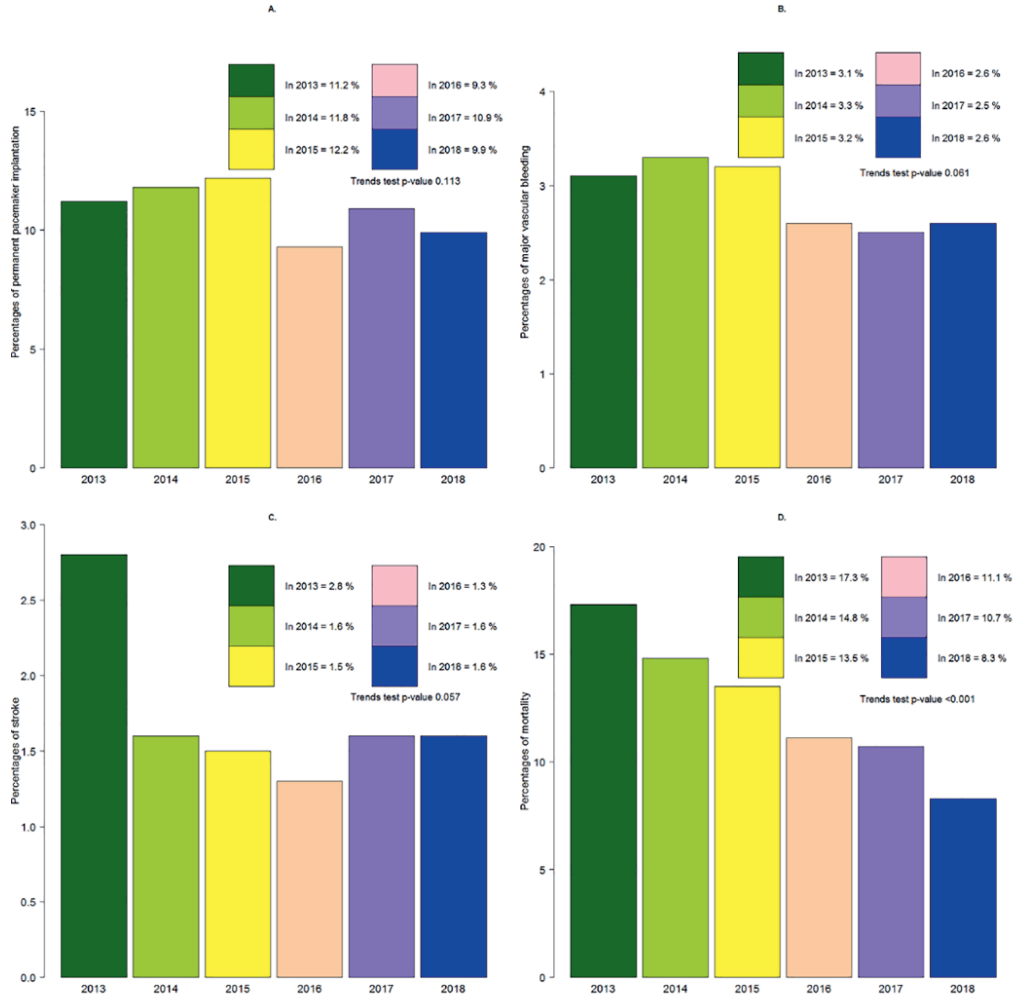


Table 1. The incidences and trends of major adverse cardiac events (MACE) and mortality after TAVI procedures and the trends over the years (2013-2018) for the 9144 TAVI-patients from the Netherlands Heart Registration registry (NHR)

Major Adverse events after TAVI	Events and total cases 2013-2018	Incidence %	Trends over the years						Trend test p-value
			2013-2014	%	2015-2016	%	2017-2018	%	
Permanent pacemaker implantation in the first 30-days	980/9075*	10.8	210/1813	11.6	310/2877	10.8	460/4385	10.5	0.113
Major vascular bleeding in the first 30-days	255/8084*	3.2	58/1601	3.6	84/2541	3.3	113/3942	2.9	0.061
Stroke in first 72 hours	149/8574*	1.7	39/1695	2.3	41/2605	1.6	69/4274	1.6	0.057
30-day mortality	368/9144	4.0	122/1822	6.7	112/2904	3.9	134/4418	3.0	<0.001
One-year mortality	1062/9144	11.6	289/1822	15.9	355/2904	12.2	418/4418	9.5	<0.001

* Missing values in the corresponding outcomes from the total the 9144 included patients as follows: Permanent pacemaker implantation N= 69; Major vascular bleeding N= 1060; and Stroke N= 570 (see E-Supplementary fig. 1 Flowchart for patients included in the study)

Figure 2. Annual (2013 up to 2018) incidences of A. Permanent pacemaker implantation, B. Major vascular bleeding, C. Stroke, and D. One-year mortality



When these yearly incidences were compared to the recent NHR-published incidences, we observed that the incidence of PPI (10.8%) in our study cohort (2013-2018) was comparable to PPI in 2019 and 2020, 9.9% and 11.5%, respectively. Moreover, MVB was 3.2 % in our study cohort and comparable to 2.9% in 2019 and 2020.

The incidence of stroke was 1.7% in our study cohort. However, stroke incidence was slightly higher in 2019 and 2020, at 1.9% and 2.5%, respectively. The early-mortality incidence in our study cohort was 4.0%, which declined to 2.1% and 2.7% in 2019 and 2020.

Also, 1-year mortality (11.6%) declined in 2019 to 10.3% (there were no published details about the 1-year mortality in 2020 at the time of this study) (21).

Details appear in e-supplementary table 2a and e-supplementary figure 2. Analysis by three age groups (<75, 75-80, >80 years) showed that most MACE and mortality occur more frequently in the oldest age group (>80 years) (table 2, figure 3 A, and e-supplementary tables 2b, 2c and 3).

Table 2. The incidence and totals of the major adverse cardiac events (MACE) in different age groups of the 9144 TAVI patients. The patients are sub-grouped according to the following age groups <75, 75-80, and >80 year

	Age groups of the TAVI patients								
Major adverse cardiac events (MACE)	**<75 years**	**%**	**75-80 years**	**%**	**>80 years**	**%**	**Totals**	%	**P-value***
Permanent pacemaker implantation (PPI)	173		286		521		980		0.24
Number of patients in this group	1749	19.3	2480	27.3	4846	53.4	9075	100	
Incidence %	173/1749	9.9	286/2480	11.5	521/4846	10.8	980/9075	10.8	
P-value*		0.17		0.17		0.88			
Major vascular bleeding (MVB)	38		71		146		255		0.81
Number of patients in this group	1515	18.7	2203	27.3	4366	54.0	8084		
Incidence %	38/1515	2.5	71/2203	3.2	146/4366	3.3	255/8084	3.2	
P-value*		0.11		0.83		0.29			
Stroke	25		41		83		149		0.73
Number of patients in this group	1611	18.8	2351	27.4	4612	53.8	8574	100	
Incidence %	25/1611	1.6	41/2351	1.7	83/4612	1.8	149/8574	1.7	
P-value*		0.53		0.98		0.64			
30-days mortality	64		97		207		368		0.50
Number of patients in this group	1759	19.2	2499	27.3	4886	53.4	9144		
Incidence %	64/1759	3.6	97/2499	3.9	207/4886	4.2	368/9144	4.0	
P-value*		0.36		0.67		0.27			
One-year mortality	207		267		588		1062		0.38
Number of patients in this group	1759	19.2	2499	27.3	4886	53.4	9144	100	
Incidence %	207/1759	11.8	267/2499	10.7	588/4886	12.0	1062/9144	11.6	
P-value*		0.82		0.09		0.18			

*P-values: chi-sq testing the number of cases in the category with the outcome of the corresponding outcome

The incidences of MACE and mortality were also analyzed in age tertiles. The age tertiles ranged from younger (34-77) to medium (78-82) and high age (83-101 years). Most of the adverse outcomes, especially mortality, were observed in the older patients' group ≥83 years (e-supplementary table 3). However, strikingly, in the younger age tertile, mortality was higher than in the medium age tertile.

The graph in (figure 3 B) presents the incidences of each of the adverse outcomes in the age tertiles. Lastly, the patients were divided based on the perioperative-risk EuroSCORE-II into tertile groups. The first group has a low operative-risk score (between 0.5-2.7), the second group: has a moderate risk (>2.7-5.2), and the third: has a higher risk (≥5.2) (table 3).

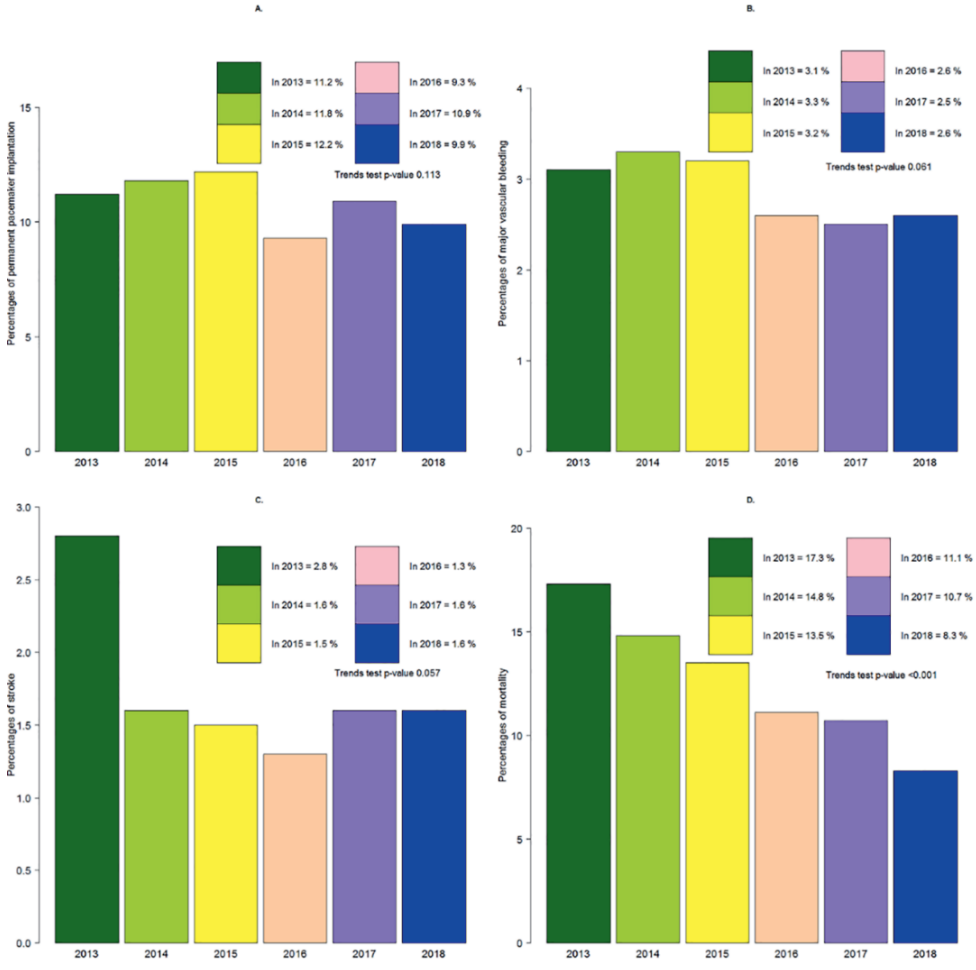
The incidence of the 30-days and 1-year mortality was higher among the patients with higher EuroSCORE-II than the other two groups (table 3). However, the incidences of PPI, MVB, and stroke did not follow this pattern. The incidences of PPI were higher in the lowest-risk group compared to the high-risk group (11.1% versus 10.2%). While in the moderate-risk groups, stroke (2.1%) was non-significantly higher compared to the high-risk group (1.6%).

Table 3. Incidences of TAVI-related major adverse cardiac events (MACE) in the 9144 patients divided into three operative risk groups (33.3 % tertiles) EuroSCORE-II estimated

	Three groups of TAVI-patients divided with 33.33% tertiles of the EuroSCORE-II								
	T1		**T2**		**T3**		**Total**	**%**	**p-value***
	Total	**%**	**Total**	**%**	**Total**	**%**			
Number of patients (n)	3048	33.3	3048	33.3	3048	33.3	9144	100	
EuroSCORE-2 (range)	0.5-2.7		>2.7-5.2		>5.2-68		0.5-68		
EuroSCORE-2 (mean)	1.9		3.8		10.1		5.3		
EuroSCORE-2 (median)	1.9		3.7		8.1		3.7		
Major adverse cardiac events (MACE)									
Permanent pacemaker implantation (PPI)	337		332		311		980		0.49
Incidence **%**	337/3048	11.1	332/3048	10.9	311/3048	10.2	980/9144	10.7	
Deceased (%) from this group in 1 year	26/337	7.7	37/332	11.1	46/311	14.8	109/980	11.1	
Major vascular bleeding (MVB)	75		92		88		255		0.40
Incidence **%**	75/3048	2.5	92/3048	3.0	88/3048	2.9	255/9144	2.8	
Deceased (%) in 1 year	18/75	24.0	30/92	32.6	37/88	42.0	85/255	33.3	
Stroke	38		63		48		149		< 0.05
Incidence **%**	38/3048	1.2	63/3048	2.1	48/3048	1.6	149/9144	1.6	
Deceased (%) from this group in 1 year	7/38	18.4	18/63	28.6	20/48	41.7	45/149	30.2	
30-day mortality	74		124		170		368		< 0.05
Incidence **%**	74/3048	2.4	124/3048	4.1	170/3048	5.6	368/9144	4.0	
1-year mortality	237		344		481		1062		< 0.05
Incidence **%**	237/3048	7.8	344/3048	11.3	481/3048	15.8	1062/9144	11.6	

*P-values: chi-sq testing the number of cases in the category with the outcome of the corresponding outcome
Abbreviations: MACE = major adverse cardiac events

Figure 3. A. Incidences (%) of each of the TAVI-related major adverse cardiac events and mortality per age (in years) for the 9144 TAVI-patients. Age is presented on the horizontal axis as a continuous variable. B. The incidence of TAVI-related major adverse cardiac events in three age groups of 33.3% tertiles of the 9144 TAVI patients



Discussion

In our national cohort of 9144 TAVI-patients, we noticed a strong increase in TAVI-procedures and a significant reduction in mortality over the years, between 2013 and 2018. The MACEs incidences were: PPI (10.8%), MVB (3.2%), and stroke (1.7%). The MACEs incidences remained stable over the 2013-2018 period.

Our results on TAVI-related adverse outcomes rates support other published studies. Whereas mortality rates were comparable to previously published mortality rates (3.8% and 10%) (14, 24, 25), we found a lower PPI rate in our cohort when compared with previously data.

In our cohort, the PPI rate was 10.8%, while previous data ranged from 15% to 40% (6-10, 26). In addition, the MVB rate (3.2%) was lower in our cohort compared with previous data ranging from 4% to 9.3% (2, 11-15, 27). The stroke rate in our cohort was 1.7%, which is comparable to other reports (2, 16-20).

When our results were compared to the recently NHR-reported incidence rates of mortality and MACE (2019 and 2020), the incidence of mortality continued to decline, and PPI and MVB continued stable, while stoke increased.

Our study adds to the existing literature by offering insights into the incidence of TAVI-related MACE and mortality in the different age groups and perioperative-risk groups. Little data exist on adverse outcomes in various age groups or perioperative-risk groups. Some studies reported the incidence among genders (28), or in patients >85 years versus younger patients. Habertheuer et. al., reported more neurologic events among patients >85.8 years, predominantly female (68.6%) (18).

A recent study by Dąbrowski et. al. reported a significantly higher operative-risk EuroSCORE-II of 7.2%, and a higher rate of 30-days and 1-year mortality (5% and 9.4%, respectively) in patients ≥85 years (29). However, they reported a similar rate of MACEs in patients >85 years, except for MVB, which was significantly higher in elderly patients with 18.5% and 8.5%, respectively. Other recent data from a large TAVI-population showed that 3.9% of the patients that underwent transfemoral TAVI have developed MVB, were the majority of them were females (30).

Our study has the inherent limitations of any retrospective registry analysis. Unlike randomized controlled studies, the cases in our registry-based non-prospective cohort were not strictly monitored, except for mortality, especially after hospital discharge. Another limitation of our study is related to the types of used TAVI-prostheses. Since we included TAVI patients starting in 2013, different TAVI-prostheses types (self-expanding and balloon-expanding) and different generations (old and new) were used. However, this heterogeneous cohort of patients might reflect the current clinical practice. Unfortunately, our national NHR dataset does not capture details on prostheses type.

Our study has also important strengths. We present one of the largest cohorts to date investigating the incidence and trends of TAVI-related MACE and mortality with in-depth analyses among different age and operative-risk groups. The study uses a national cohort with excellent completeness of registration.

The occurrence of TAVI-related adverse outcomes mandates ongoing critical evaluation of risk-factors associated with MACE. Such continuous research utilizing data about the used TAVI-prosthesis devices contributes to quality improvement and gains more insight into the success of the centers with low MACE rates.

This type of research allows physicians to evaluate the outcome and address them to effectively improve clinical outcomes.

To improve clinical outcomes, better identification of risk factors needed to be added in large registries. It will allow the development of MACE-related risk prediction models, which are highly needed for individualized and patient-tailored information and treatment.

The current NHR-TAVI variables set has limited capability for the development of such prediction models. As such, for better prediction of PPI, we would suggest adding relevant echocardiographic characteristics (left ventricular end-diastolic and end-systolic diameters, interventricular septum thickness, aortic valve annulus diameter) and electrocardiographic characteristics (PR-interval, degree atrioventricular block, QRS-duration and axis, and the presence of right or left bundle branch block).

Conclusion

Our study reports on MACE rates and mortality over the years in the Netherlands. The number of performed TAVI-procedures has increased, while the mortality rate has decreased over the years. Most of the adverse outcomes have occurred in elderly patients and patients with higher operative-risk. More attention is warranted to the younger and low-risk patients, which may be at higher risk for PPI.

Contributors

HF, AR, and JH contributed to the conception and design of the study. HA performed the statistical analysis using existing registry data. AR and AA-H also contributed to the analysis. HF wrote the first draft. HF, AR, and JH contributed to writing the first manuscript. HF, AA-H, JH, BM, AR, and SH contributed to the critical revision of the article. The final version was approved by all the authors. The first author HF and principal investigators AA-H, JH, and BM, take primary responsibility for the paper and the coordination of the study.

Funding

The authors have not declared a specific grant for this research from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

Hatem Al-Farra, Anita CJ Ravelli, Ameen Abu-Hanna, Saskia Houterman, Bas AJM de Mol, and José PS Henriques declare no conflict of interest.

Ethics approval

The study was approved by the institutional review board of the Catharina Hospital (Approval number: W19.-270). The NHR registry commission approved using the NHR registry data for the research question. The used anonymized data conformed to the Declaration of Helsinki principles. No further ethical approval was needed.

Addendum

The following physicians are members of the NHR THI Registration Committee. They represent the hospitals that have provided the data for this study. Contact with NHR THI Registration Committee can be via the e-mail: info@nederlandsehartregistratie.nl

[#] The NHR THI Registration Committee:

M.M. Vis, Amsterdam University Medical Centres.
P. den Heijer, Amphia Hospital.
L. Timmers, St. Antonius Hospital.
W.A.L. Tonino, Catharina Hospital.
C.E. Schotborgh, Haga Hospital.
V. Roolvink, Isala.
F. Porta, Leeuwarden Medical Centre.
M.G. Stoel, Medisch Spectrum Twente.
S. Kats, Maastricht University Medical Centre.
G. Amoroso, Onze Lieve Vrouwe Gasthuis.
H.W. van der Werf, University Medical Centre Groningen.
M. Voskuil, University Medical Centre Utrecht.
N.M.D.A. van Mieghem, Erasmus University Medical Centre.

Appendix. E-components

E-supplementary table 1. Baseline characteristics of the TAVI study population stratified by the outcome of Permanent pacemaker implantation within the first 30 days after the TAVI procedure. Baseline characteristics stratified by mortality status can be reviewed in our pervious published study in (31)

Variable	Total Population (%)	No permanent Pacemaker implantation (%)	Permanent Pacemaker Implantation (%)	P-value
	9075	8095 (89.2)	980 (10.8)	
Continuous variables				
Age (years) (mean (SD))	79.79 (6.90)	79.75 (6.94)	80.09 (6.51)	0.155
SBA (mean (SD))	1.89 (0.22)	1.89 (0.22)	1.91 (0.22)	0.009
Body Mass Index BMI (kg/m^2) (mean (SD))	27.29 (6.01)	27.25 (6.12)	27.62 (4.98)	0.074
LVEF (mean (SD))	50.05 (10.54)	49.99 (10.53)	50.56 (10.59)	0.116
Serum creatinine (µmol/L) (mean (SD))	106.78 (65.74)	106.01 (63.90)	113.19 (79.03)	0.001
eGFR (mean (SD))	59.97 (29.37)	60.02 (26.50)	59.51 (46.77)	0.608
Pulmonary artery systolic pressure (mean (SD))	30.76 (10.77)	30.73 (10.74)	31.02 (11.01)	0.498
EuroSCORE-II (mean (SD))	5.66 (5.57)	5.70 (5.66)	5.30 (4.68)	0.135
Outcomes				
Early (30-days) mortality (%)	2366 (26.1)	2096 (26.0)	270 (27.6)	0.299
Stroke within 72 hours (%)	165 (2.0)	149 (2.0)	16 (1.8)	0.824
Vascular bleeding within 30 days (%)				0.624
No	7244 (89.7)	6481 (89.7)	763 (90.2)	
Major vascular bleeding	255 (3.2)	226 (3.1)	29 (3.4)	
Minor vascular bleeding	574 (7.1)	520 (7.2)	54 (6.4)	
Binary categorical variables				
Female Gender (Yes)	4594 (50.6)	4141 (51.2)	453 (46.2)	0.004
Chronic lung disease (Yes)	1963 (21.8)	1742 (21.6)	221 (22.6)	0.523
Extra-cardiac arteriopathy (Yes)	2015 (22.4)	1805 (22.5)	210 (21.5)	0.538
Neurological dysfunction (Yes)	318 (4.1)	283 (4.1)	35 (4.1)	1.000
Previous cardiac surgery (Yes)	1895 (21.5)	1720 (21.9)	175 (18.5)	0.021
Critical preoperative state (Yes)	59 (0.7)	55 (0.7)	4 (0.4)	0.436
Unstable angina (Yes)	17 (0.2)	16 (0.2)	1 (0.1)	0.787
Recent MI (Yes)	168 (1.9)	152 (1.9)	16 (1.7)	0.693
Dialysis (Yes)	110 (1.2)	92 (1.2)	18 (1.9)	0.088
Poor mobility (Yes)	563 (9.0)	498 (8.8)	65 (10.3)	0.265
CCS class IV angina (Yes)	193 (2.7)	174 (2.8)	19 (2.4)	0.690
Procedure weight (2 operations) (Yes)	97 (1.1)	91 (1.2)	6 (0.6)	0.183
Previous CVA (Yes)	1009 (11.2)	905 (11.2)	104 (10.6)	0.599
Previous aortic valve surgery (Yes)	429 (5.0)	408 (5.3)	21 (2.2)	<0.001
Anaesthesia (Yes)	5451 (63.0)	4901 (63.4)	550 (59.3)	0.016
Balloon pre-TAVI (Yes)	4086 (50.5)	3678 (50.8)	408 (47.6)	0.083
Non-binary categorical variables				
Functional NYHA class				0.103
NYHA class I (Yes)	887 (11.1)	811 (11.4)	76 (8.7)	
NYHA class II (Yes)	2097 (26.3)	1862 (26.3)	235 (26.8)	
NYHA class III (Yes)	4421 (55.5)	3917 (55.2)	504 (57.4)	
NYHA class IV (Yes)	564 (7.1)	501 (7.1)	63 (7.2)	
Diabetes Mellitus (DM)				0.283
No DM	6434 (72.5)	5748 (72.7)	686 (70.6)	
DM is not on medication	388 (4.4)	338 (4.3)	50 (5.1)	
DM on medication	2056 (23.2)	1821 (23.0)	235 (24.2)	
Access route				0.001
Access route Transfemoral (Yes)	7068 (79.5)	6267 (78.9)	801 (84.2)	
Access route subclavian artery (Yes)	485 (5.5)	436 (5.5)	49 (5.2)	
Access route Transapical (Yes)	649 (7.3)	600 (7.6)	49 (5.2)	
Access route direct aortic (Yes)	694 (7.8)	642 (8.1)	52 (5.5)	
Procedure acuity				0.400
Elective	8019 (90.7)	7160 (90.8)	859 (89.7)	
Urgent	798 (9.0)	701 (8.9)	97 (10.1)	
Emergency	26 (0.3)	24 (0.3)	2 (0.2)	

Values are mean (or median where appropriate) ± standard deviation (S.D.), or number (n) and percentage (%).

Abbreviations: BMI = Body mass index; Balloon pre-TAVI = Balloon aortic valvuloplasty prior to the date of TAVI; CCS class = Canadian Cardiovascular Society grading of angina pectoris; CKD = Chronic Kidney Disease; CVA = cerebrovascular accident; DM = Diabetes mellitus; eGFR = estimated Glomerular Filtration Rate; LVEF = Left Ventricular Ejection Fraction; MI = myocardial infarction; NYHA = New York Heart Association functional Classification; PABV = Percutaneous Aortic Balloon Valvuloplasty (TAVI post-dilation), Post-MI VSR = poet myocardial infarction ventricular septal rupture; SBA = surface body area; sPAP = systolic Pulmonary Arterial Pressure.

E-supplementary table 2a. Incidences (%) of the MACE and mortality in each calendar-year as calculated in this study, and as reported in the NHR annual report of 2021

Adverse outcomes after TAVI	Incidences (%) as calculated in this study							Incidence (%) from the NHR annual report 2021 (1)	
	2013	2014	2015	2016	2017	2018	2013-2018	2019	2020
Permanent pacemaker implantation first 30-days	11.2	11.8	12.2	9.3	10.9	9.9	10.8	9.9	11.5
Major vascular bleeding in the first 30-days	3.1	3.3	3.2	2.6	2.5	2.6	3.2	2.9	2.9
Stroke in first 72 hours	2.8	1.6	1.5	1.3	1.6	1.6	1.7	1.9	2.5
30-day mortality	7.8	5.7	3.9	3.8	3.3	2.8	4.0	2.1	2.7
One-year mortality	17.3	14.8	13.5	11.1	10.7	8.3	11.6	10.3	N.A.

N.A. = not available

E-supplementary table 2b. Numbers of patients in each year and each age group, numbers and percentages of mortality that occurred on 30-days or one year after TAVI in each calendar year from 2013 up to 2018. Note this data is about the patients with known mortality outcomes (N= 9144)

	Number of patients in each age group			Total patients (= TAVI procedures)	Mortality after TAVI	
Calendar year	Patients below 75 years	Patients between 75-80 years	Patients more than 80 years		30-days early mortality (%)	One-year mortality (%)
2013	146	205	435	786	61 (7.8)	136 (17.3)
2014	186	260	590	1036	61 (5.7)	153 (14.8)
2015	255	382	736	1373	54 (3.9)	185 (13.5)
2016	293	435	803	1531	58 (3.8)	170 (11.1)
2017	405	569	1155	2129	71 (3.3)	227 (10.7)
2018	474	648	1167	2289	63 (2.8)	191 (8.3)
Totals	**1759**	**2499**	**4886**	**9144**	**368 (4.0)**	**1062 (11.6)**

E-supplementary table 2c. Total number of mortality that occurred in each sequential year following the TAVI procedure, in the whole cohort, and each age group. Note that there are no enough available data about long-term follow-up after TAVI in our dataset with regards to mortality. For example, for patients who operated in 2014, we have data about the next 5 years, but for patients from 2017 we have only 2-3 years of data, and so on

Long-term mortality	Patients below 75 years	Patients between 75-80 years	Patients more than 80 years	Totals number of mortality
Mortality occurred only in the:				
1st year after TAVI	207	267	588	1062
2nd year after TAVI	90	133	275	498
3rd year after TAVI	81	103	229	413
4th year after TAVI	42	58	156	256
5th year after TAVI	20	38	72	130
6th year after TAVI	4	3	6	13

E-supplementary table 2d. Cumulative numbers of mortality the occurred in the years following the TAVI procedure

Long-term mortality after TAVI	Cumulative numbers of mortality	% of mortality after TAVI
Mortality after TAVI occurred within a complete period of	n	%
One year	1062	44.8
Two years	1560	65.8
Three years	1973	83.2
Four years	2229	94.0
Five years	2359	99.5
Six years	2372	100.0

E-supplementary table 3. Incidence and totals of the TAVI-related major adverse cardiac events (MACE) in three age groups of 33.3% tertiles of the age of the 9144 TAVI patients

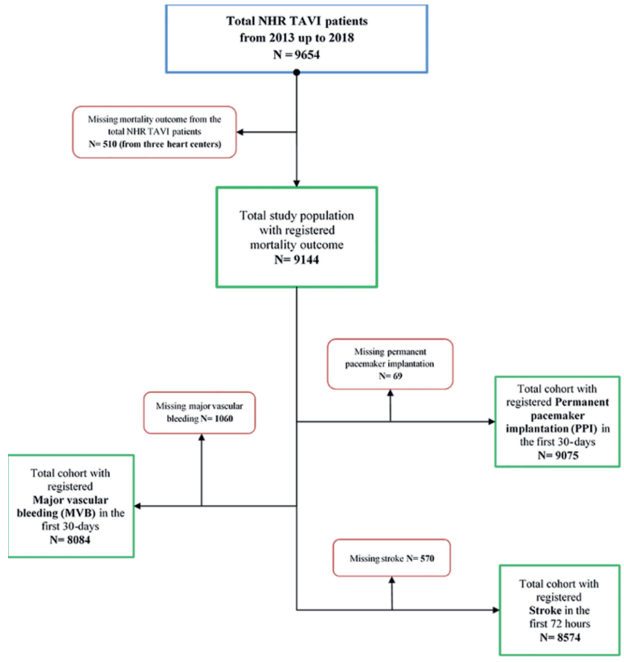
	Three age groups divided by tertiles classes of TAVI patients								
	T 1 (young ages)		T 2 (mid. ages)		T 3 (older ages)				
	Total	%	Total	%	Total	%	Total	%	P-value
Age range for each of the tertiles (years)	34 – 77	-	78 – 82	-	83 – 101	-	-	-	-
Median Age in each of the tertiles (years)	74	-	81	-	86	-	-	-	-
Major adverse cardiac events									
Permanent pacemaker implantation	313		329		338		980		0.61
Number of patients (n)	3025	33.3	3025	33.3	3025	33.3			
Incidence %	313/3025	10.3	329/3025	10.9	338/3025	11.2	980/9075	10.8	
P-value*		n.s.		n.s.		n.s.			
Major vascular bleeding	75		85		95		255		0.24
Number of patients (n)	2695	33.3	2695	33.3	2694	33.3	8084	100	
Incidence %	75/2695	2.8	85/2695	3.2	95/2694	3.5	255/8084	3.2	
P-value*		n.s.		n.s.		n.s.			
Stroke	46		52		51		149		0.70
Number of patients (n)	2858	33.3	2858	33.3	2858		8574		
Incidence %	46/2858	1.6	52/2858	1.8	51/2858	1.8	149/8574	1.7	
P-value*		n.s.		n.s.		n.s.			
30-day mortality	121		103		144		368		0.03
Number of patients (n)	3048	33.3	3048	33.3	3048	33.3	9,144	100	
Incidence %	121/3048	4.0	103/3048	3.4	144/3048	4.7	368/9144	4.0	
P-value*		n.s.		0.02		0.02			
1-year mortality	353		321		388		1062		0.04
Number of patients (n)	3048	33.3	3048	33.3	3048	33.3	9144		
Incidence %	353/3048	11.6	321/3048	10.5	388/3048	12.7	1062/9144	11.6	
P-value*		n.s.		0.02		0.02			

P-value: chi-sq testing the number of cases in the category with the outcome of the corresponding outcome

E-supplementary table 4. Sub-analysis for incidences of TAVI related major adverse cardiac events in different age groups over the years. Note that the number of TAVI procedures increased over time from 20% in 2013-2014 to 48% in 2017-2018. Also, note that 60% of the patients were 80 years and older. The percentage of 80+ patients slightly declined over time from 62% to 58%

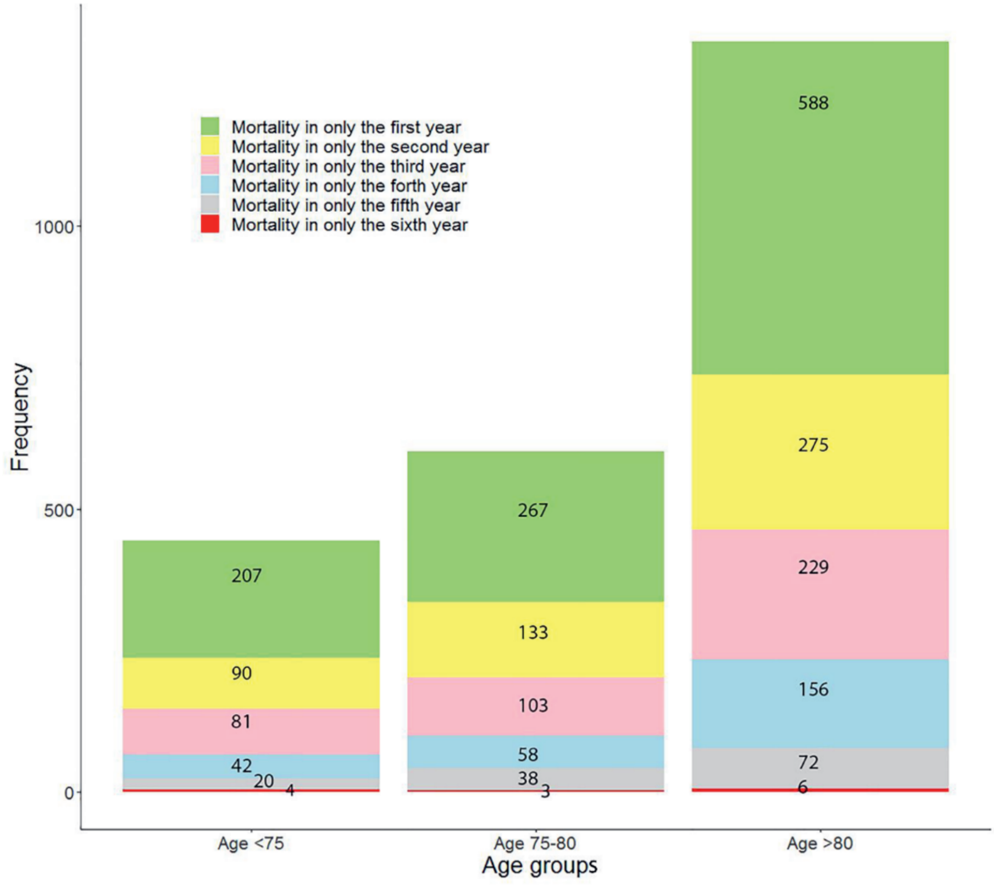
Years		Overall	2013-2014	2015-2016	2017-2018
Age					
30-54	**Total**	46/9144	17/46	14/46	15/46
	%	0.5	37.0	30.4	32.6
55-64	**Total**	211/9144	37/211	74/211	100/211
	%	2.3	17.5	35.1	47.4
65-74	**Total**	1502/9144	278/1502	460/1502	764/1502
	%	16.4	18.5	30.6	50.9
75-79	**Total**	1930/9144	359/1930	617/1930	955/1930
	%	21.1	18.6	32.0	49.4
<80	**Total**	3689	691	1165	1833
	%	40.3	38	40	42
≥80	**Total**	5455/9144	1131/5455	1739/5455	2585/5455
	% row	59.7	20.7	31.9	47.4
	Total		1131/1822	1739/2904	2585/4418
	% column		62	60	58
Total patients	**Total**	9144	1822/9144	2904/9144	4418/9144
%	**%**	100	19.9	31.8	48.3

E-supplementary figure 1. Flowchart for TAVI-patients included in the study



E-supplementary figure 2. The annual incidences of major adverse cardiac events (MACE) and (early and one-year) mortality as reported by the NHR (21)

E-supplementary figure 3. Long-term mortality after TAVI in each sequential year per age group. Note that there are not enough available data about long-term after TAVI in our dataset with regards to mortality. For example, for patients who operated in 2014, we have data about the next 5 years, but for patients from 2017 we have only 2-3 years of data, and so on

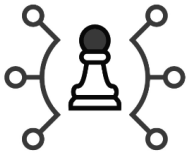


	Age group <75		Age group 75-80		Age group >80	
	N=1759		N=2499		N=4886	
Mortality only in the	Numbers	%	Numbers	%	Numbers	%
First year	207	46.6	267	44.4	588	44.3
Second year	90	20.3	133	22.1	275	20.7
Third year	81	18.2	103	17.1	229	17.3
Forth year	42	9.5	58	9.6	156	11.8
Fifth year	20	4.5	38	6.3	72	5.4
Sixth year	4	0.9	3	0.5	6	0.5
Total (N=2372)	444	100%	602	100%	1326	100%
Mortality %	444/1759	25.2	602/2499	24.1	1326/4886	27.1

References:

1) Leon MB, Smith CR, 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.
2) Smith CR, Leon MB, Mack MJ, 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.
3) Mollmann H, Kim WK, Kempfert J, Walther T, Hamm C. Complications of transcatheter aortic valve implantation (TAVI): how to avoid and treat them. Heart. 2015;101(11):900-8.
4) Kappetein AP, Head SJ, Genereux P, Piazza N, van Mieghem NM, Blackstone EH, et al. Updated standardized endpoint definitions for transcatheter aortic valve implantation: the Valve Academic Research Consortium-2 consensus document. J Am Coll Cardiol. 2012;60(15):1438-54.
5) Kappetein AP, Head SJ, Genereux P, Piazza N, van Mieghem NM, Blackstone EH, et al. Updated standardized endpoint definitions for transcatheter aortic valve implantation: the Valve Academic Research Consortium-2 consensus document. J Thorac Cardiovasc Surg. 2013;145(1):6-23.
6) Mangieri A, Montalto C, Pagnesi M, Lanzillo G, Demir O, Testa L, et al. TAVI and Post Procedural Cardiac Conduction Abnormalities. Front Cardiovasc Med. 2018;5:85.
7) Poels TT, Engels EB, Kats S, Veenstra L, van Ommen V, Vernooy K, et al. Occurrence and Persistency of Conduction Disturbances during Transcatheter Aortic Valve Implantation. Medicina (Kaunas). 2021;57(7).
8) Chamandi C, Barbanti M, Munoz-Garcia A, Latib A, Nombela-Franco L, Gutierrez-Ibanez E, et al. Long-Term Outcomes in Patients With New-Onset Persistent Left Bundle Branch Block Following TAVR. JACC Cardiovasc Interv. 2019;12(12):1175-84.
9) Nuis RJ, Van Mieghem NM, Schultz CJ, Tzikas A, Van der Boon RM, Maugenest AM, et al. Timing and potential mechanisms of new conduction abnormalities during the implantation of the Medtronic CoreValve System in patients with aortic stenosis. Eur Heart J. 2011;32(16):2067-74.
10) Nai Fovino L, Cipriani A, Fabris T, Massussi M, Scotti A, Lorenzoni G, et al. Anatomical Predictors of Pacemaker Dependency After Transcatheter Aortic Valve Replacement. Circ Arrhythm Electrophysiol. 2021;14(1):e009028.
11) van Kesteren F, van Mourik MS, Vendrik J, Wiegerinck EMA, Henriques JPS, Koch KT, et al. Incidence, Predictors, and Impact of Vascular Complications After Transfemoral Transcatheter Aortic Valve Implantation With the SAPIEN 3 Prosthesis. Am J Cardiol. 2018;121(10):1231-8.
12) Sedaghat A, Neumann N, Schahab N, Sinning JM, Hammerstingl C, Pingel S, et al. Routine Endovascular Treatment With a Stent Graft for Access-Site and Access-Related Vascular Injury in Transfemoral Transcatheter Aortic Valve Implantation. Circ Cardiovasc Interv. 2016;9(8).
13) Ullery BW, Jin R, Kirker EB, Hayes G, Siwek L, Brevig J, et al. Trends in vascular complications and associated treatment strategies following transfemoral transcatheter aortic valve replacement. J Vasc Surg. 2020;72(4):1313-24 e5.
14) Batchelor W, Patel K, Hurt J, Totten J, Burroughs P, Smith G, et al. Incidence, Prognosis and Predictors of Major Vascular Complications and Percutaneous Closure Device Failure Following Contemporary Percutaneous Transfemoral Transcatheter Aortic Valve Replacement. Cardiovasc Revasc Med. 2020;21(9):1065-73.
15) Genereux P, Webb JG, Svensson LG, Kodali SK, Satler LF, Fearon WF, et al. Vascular complications after transcatheter aortic valve replacement: insights from the PARTNER (Placement of AoRTic TraNscathetER Valve) trial. J Am Coll Cardiol. 2012;60(12):1043-52.
16) Krasopoulos G, Falconieri F, Benedetto U, Newton J, Sayeed R, Kharbanda R, et al. European real world trans-catheter aortic valve implantation: systematic review and meta-analysis of European national registries. J Cardiothorac Surg. 2016;11(1):159.
17) Muralidharan A, Thiagarajan K, Van Ham R, Gleason TG, Mulukutla S, Schindler JT, et al. Meta-Analysis of Perioperative Stroke and Mortality in Transcatheter Aortic Valve Implantation. Am J Cardiol. 2016;118(7):1031-45.
18) Habertheuer A, Gleason TG, Kilic A, Schindler J, Kliner D, Bianco V, et al. Impact of Perioperative Stroke on Midterm Outcomes After Transcatheter Aortic Valve Replacement. Ann Thorac Surg. 2020;110(4):1294-301.
19) Mack M, Hamandi M. Why Surgical Risk Algorithms Are Not Predictive of Transcatheter Aortic Valve Replacement Outcomes! Circ Cardiovasc Interv. 2019;12(1):e007560.
20) 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(18):1706-15.
21) Netherlands Heart Registration: NHR; 2019 (Cited 2020 June 15). Available from: https://nederlandsehartregistratie.nl/wp-content/uploads/2020/01/NHR-Rapportage-2019-per-spread-230120.pdf.
22) Signorell Aema. DescTools: Tools for Descriptive Statistics. 2022.
23) R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing,Vienna, Austria. 2019.
24) Mollmann H, Husser O, Blumenstein J, Liebetrau C, Dorr O, Kim WK, et al. Lower mortality in an all-comers aortic stenosis population treated with TAVI in comparison to SAVR. Clin Res Cardiol. 2019.
25) Hoffmann R, Almutairi B, Herpertz R, Lotfipour S, Stohr R, Aktug O, et al. Two-year mortality after transcatheter aortic valve implantation versus medical therapy for high-surgical risk or inoperable aortic stenosis patients. J Heart Valve Dis. 2013;22(1):71-8.

26) Hengstenberg C, Chandrasekhar J, Sartori S, Lefevre T, Mikhail G, Meneveau N, et al. Impact of pre-existing or new-onset atrial fibrillation on 30-day clinical outcomes following transcatheter aortic valve replacement: Results from the BRAVO 3 randomized trial. Catheter Cardiovasc Interv. 2017;90(6):1027-37.
27) Baron SJ, Arnold SV, Wang K, Magnuson EA, Chinnakondepali K, Makkar R, et al. Health Status Benefits of Transcatheter vs Surgical Aortic Valve Replacement in Patients With Severe Aortic Stenosis at Intermediate Surgical Risk: Results From the PARTNER 2 Randomized Clinical Trial. JAMA Cardiol. 2017;2(8):837-45.
28) Sannino A, Szerlip M, Harrington K, Schiattarella GG, Grayburn PA. Comparison of Baseline Characteristics and Outcomes in Men Versus Women With Aortic Stenosis Undergoing Transcatheter Aortic Valve Implantation. Am J Cardiol. 2018;121(7):844-9.
29) Dabrowski M, Pylko A, Chmielak Z, Kwiecinski J, Kukula K, Wysocki K, et al. Comparison of transcatheter aortic valve implantation outcomes in patients younger than 85 years and those aged 85 years or older: a single-center study. Pol Arch Intern Med. 2021;131(2):145-51.
30) Sherwood MW, Xiang K, Matsouaka R, Li Z, Vemulapalli S, Vora AN, et al. Incidence, Temporal Trends, and Associated Outcomes of Vascular and Bleeding Complications in Patients Undergoing Transfemoral Transcatheter Aortic Valve Replacement: Insights From the Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapies Registry. Circ Cardiovasc Interv. 2020;13(1):e008227.
31) Al-Farra H, Abu-Hanna A, de Mol B, Ter Burg WJ, Houterman S, Henriques JPS, et al. External validation of existing prediction models of 30-day mortality after Transcatheter Aortic Valve Implantation (TAVI) in the Netherlands Heart Registration. Int J Cardiol. 2020;317:25-32.



STRATEGY

Chapter 6:

Development of a strategic information management plan for a heart center using the Practical Guideline

Submitted for publication

Hatem Al-Farra, Willem Jan P.P. ter Burg, Ameen Abu-Hanna, José P.S. Henriques, Anita C.J. Ravelli, Bas A.J.M. de Mol

Abstract

Background

Strategic information management plans (SIM-plan) guide healthcare facilities with planning and managing hospital information systems (HIS). There is paucity in the literature about the experience of healthcare facilities with developing SIM-plans. We share our experience in developing a SIM-plan at a Heart Center in a large hospital. We outline how we developed the SIM-plan and present the key results of the process.

Methods

The study was conducted at a Heart Center in an academic hospital in the Netherlands. All steps of the SIM-plan development process were carried out based on a published Practical Guideline (an empirical approach to developing SIM-plans). It implies starting with organization analysis, identification of (business and IT) goals, evaluating the current HIS situation, defining the future HIS situation, outlining a roadmap and a migration-path, and finally SIM-plan approval and deployment.

Results

The guideline-based SIM-plan development took six months. It required reviewing 21 internal documents, conducting 50 semi-structured interviews, and participating in 14 meetings and six roundtables. These activities helped define 15 business goals and 6 IT goals; describe and assess the current HIS situation; and plan the future HIS situation. Understanding hospital-wide IT governance arrangements helped align business- and IT goals across multiple organizational levels. The main outcome of the development process is the SIM-plan document, which is used to develop the projects' portfolio.

Conclusion

Using the Practical Guideline allowed the development of a SIM-plan for the Heart Center. The guideline-based development process enabled goals identification and assessing the HIS situation, which facilitated planning the future HIS situation and development of the project portfolio. The study contributes by confirming that business goals and IT goals alignment are important issues in healthcare facilities for information management and that such healthcare facilities should make deliberate efforts to realize them. This study corroborates the usefulness of the Practical Guideline for developing SIM-plans in a specific clinical department.

Introduction

The mission of healthcare facilities is to provide high-quality patient care. Healthcare facilities make deliberate efforts to realize this mission by carful strategic planning of their business goals and resources. Nowadays, hospital information systems (HIS) are considered an essential resource supporting strategic business goals.

Therefore, the management of HIS is necessary to realize the stated mission. Planning of information management (IM) in healthcare facilities has three scopes or levels: strategic IM (SIM), tactical IM (TIM), and operational IM (OIM) (1). The overall SIM scope is necessary to inform and guide the other two scopes of IM planning.

Planning the SIM (SIM-planning) results in generating a strategic information management plan (SIM-plan) document. This document comprises the strategic, business-goals, and IT goals; description and assessment of the current HIS situation; and description of the new HIS situation (2). This document delineates how the long-term goals of a particular healthcare facility shall be supported and implemented by information systems.

The SIM-plan document, in turn, provides the basis to develop the strategic project portfolio, which can be translated as moderate or short-term tactical projects in the TIM scope. The OIM planning scope is intended to operate and implement tactical projects (1, 3).

SIM-planning plays an important role in managing the HIS and prioritizing IT projects according to the strategic business goals and the facility mission. It, therefore, enables hospitals to align IT goals with business goals (4-8). For healthcare facilities, hence, both strategic business planning and SIM-planning (as part of it) are important. Both involve defining all actionable steps required to achieve the mission and vision.

A systematic literature review (9) showed that there are various approaches proposed to develop SIM-plans in healthcare organizations. These approaches range from theoretical, in the sense that they were written from the authors' experience without illustrating a subject organization, to empirical, in the sense that they were written through the authors' experience with a subject organization. Whereas most of the reported approaches were theoretical, two were empirical approaches (4, 6).

Brigl et al. (2005) suggested an approach, called the Practical Guideline, which harmonizes advice on developing hospital SIM-plans (6). As such, the Practical Guideline has the promise to be a useful practical guide for developing SIM-plans. The Practical Guideline is used in various hospitals, particularly in West Europe. Although the Practical Guideline is an empirical approach described by the authors themselves as practical, it is still based on a narrow basis of evidence (4, 6, 8, 10-13).

Moreover, there is a limited number of studies reporting or sharing the experience of developing SIM-plans using this approach (11, 14-22). This limits our understanding of the steps of the development process and their end products; and the outcomes from developing SIM-plans using the Practical Guideline in various healthcare settings, specifically, a medical department in a large hospital. The paper of Rosenberger and Kaiser is a qualitative study performed in a private healthcare organization. They reported a dynamic framework for IM for a specific healthcare facility (4).

Unlike the Practical Guideline, the framework of Rosenberger and Kaiser was not structured as a procedural guideline, but as a high-level guideline. Our experience was that the framework of Rosenberger and Kaiser was rather difficult to deploy. Therefore, the Practical Guideline from Brigel et al. was selected by the project team to be used to develop a SIM-plan for our heart center.

This study provides an overview of the process of developing a SIM-plan at a heart center in a large academic hospital. Specifically, we aim to share our experience and describe the developmental process steps (including their end products) applying the Practical Guideline to a Heart Center in a large academic medical hospital.

This study has two important characteristics. First, the study is conducted in a heart center containing more than one specialty. Second, it provides the challenge to align IT goals to the business goals of a clinical department (Heart Center) and, simultaneously, to the hospital-wide strategies.

Our work broadens the knowledge about using Practical Guideline to develop SIM-plans in various healthcare settings.

Methods

Study settings

The setting for developing the SIM-plan is a Heart Center, which is a large clinical department in a large tertiary academic medical center, (further referred to as Hospital), in the Netherlands. The Heart Center has its board of directors. It has three main sub-departments (cardiology, interventional cardiology, and cardiothoracic surgery). It is a large important division of the nine divisions of the Hospital.

The Hospital has a federal IT governance arrangement, where all divisions have limited autonomy in IT plans. Decentral IT plans should be centrally coordinated and aligned. However, the Heart Center has high demands on IT infrastructure and facilities.

Ethics statement

The study was approved by the review board of the medical information department of the University of Amsterdam. Informed consent has been obtained from each of the participants in the interviews.

Approach

This study was performed in and for a single center (Heart Center), and conducted by one observer (the SIM-plan's project leader). The SIM-plan was developed by applying all the steps of the empirical approach to the Practical Guideline. The developmental process steps are as follows: 1) project initiation; 2) analysis of the organization and goals identification; 3) capturing and assessing current HIS situation; 4) defining future HIS situation and the migration path; and finally, 5) SIM-plan approval (details on the steps appear in table 1).

Each step has its own aims, contents, and specific recommendations (activities and tools) (1, 6). This involved reviewing internal documents, holding meetings and discussion sessions, giving presentations, engaging stakeholders, applying the critical success factor for the strategic business alignment, assessing HIS using the quality assessment criteria, applying a problem-oriented approach to develop the strategic project portfolios, and modeling HIS (23), (details on the steps appear in table 1).

Furthermore, additional data collection activities were applied. These activities were conducting interviews with stakeholders, and visiting sub-departments (9). These activities are not included in the Practical Guideline.

Semi-structured interviews were chosen to allow focused, conversational, two-way communication. Most of the questions were designed beforehand, while some questions were raised during the interviews, allowing for discussing details when required.

Throughout the developmental process, data triangulation was used to assure the consistency of the collected data. During the process, stakeholders were continuously kept engaged. This is to ensure ongoing evaluation of the IT needs and IT issues and to keep them informed about SIM-plan development. We report on our experience by stating the lessons learned from the whole developmental process, and by stating our observations about using the Practical Guideline in one individual department in a large hospital.

Table 1. Details on the steps undertaken during the development of the SIM-plan

Steps	Aspect	Details
Project-Initiation	Aim	• Project set-up
	Contents	• General outline and planning of the project • Team and roles assignment
	Activities and tools	• Periodic meetings
Organization analysis and goals identification	Aim	• To get an overview of the organization's strategic planning and business model
	Contents	• Collect the business goals and IT goals • To have a preliminary overview of the IT needs (although this will be studied in detail in the next step) • To process and achieve, as far as possible, business goals and IT goals alignments.
	Activities and tools	• Internal documents review • Meetings and discussion sessions with stakeholders • Semi-structured interviews with stakeholders • Applying the critical success factor approach to process the alignment of IT goals to the business goals.

Capturing and assessing the current HIS situation	Aim	• To identify the current HIS situation • To identify the future of HIS situation
	Contents	• Identifying the all details about IT weaknesses, IT needs, and IT opportunities of the current HIS situation
	Activities and tools	• Utilizing the data collected from the first steps. • Conducting semi-structured interviews with different stakeholders and users • Site visits to the sub-departments of the Heart Center • Assessment of the HIS was conducted using the Requirements Index for Information Processing in Hospitals • HIS modeling: the current HIS architecture was modeled, using the 3LGM^2 tool • Strengths, weaknesses, opportunities, and threats (SWOT) analysis was completed utilizing all collected data.
Planning the future HIS situation and the migration path	Aim	• To identify the future HIS situation in detail.
	Contents	• Translating the IT goals into the intended HIS architecture. • Gap analysis between current and future HIS situation
	Activities and tools	• The future HIS situation was modeled using the 3LGM^2 tool
Writing the SIM-plan document	Aim	• Finalize the SIM-plan document
	Contents	• Compile and arrange the collected data and end-products in the document
	Activities and tools	• Writing activity • Discussing the final findings with the project owner and project team
SIM-plan approval	Aim	• Get the SIM-plan approved for the deployment
	Contents	• Submitting the plan to the head of the Heart Center for approval
	Activities and tools	• Submission and presentation to the stakeholders
Developing the strategic project-portfolio	Aim	• To develop the strategic project portfolio
	Contents	• Develop a strategic project portfolio that will contain all the future IT projects in the next five years.
	Activities and tools	• Applying the problem-oriented approach

Results

In this section, the results of this study are presented by first describing each individual undertaken developmental process (the end products of each step) (Tables 1 and 2 are showing the aim, contents, used activities, and tools; and the end products of each step).

After that, the SIM-plan key results and outcomes in the Heart Center are highlighted. Subsequently, our experience is presented by asserting the lessons learned and our observations on using the Practical Guideline.

Table 2. Details of the main end products from each step

Steps	End products of steps
Project-Initiation	• Project outlines and time frame for conducting the whole SIM-plan development process
Organization analysis and goals identification	• Overview of the organization (hospital and department) settings. Specifically, we figured out that the Heart Center has business and financial autonomy (with central business alignment). • Understanding the hospital and department IT governance model: which is a federal model • Compiling two lists of business goals (n=50) and IT goals (n=29) • The interviews with the sub-departments directors allowed a good (preliminary) overview to be gained of the IT needs and current situation and future goals view. • Multi-level (Hospital and Heart Center) business goals and IT goals alignments.
Capturing and assessing the current HIS situation	• A complete overview of the current HIS situation was obtained • The architecture of the current HIS showed a fragmented HIS with central EPR and many application components; and many standalone systems that are not directly to each other. • Lists for all HIS strengths and weaknesses were made.
Planning the future HIS situation and migration path	• Model of the future HIS situation • The migration path (roadmap) for implementing the SIM-plan
Writing SIM-plan document	• SIM-plan document
SIM-plan approval	• SIM-plan approved
Developing the strategic project-portfolio	• The strategic project-portfolio document/chapter

Details and results of the steps of the development process

Project Initiation

The SIM-plan project started in December 2016, commissioned by the managers of the cardiothoracic surgery and interventional cardiology sub-departments (project owners). The planned project's duration was six months. The project leader (the SIM-plan developer) was a physician and medical informatician with theoretical schooling in SIM-planning. The project team consisted of the project leader, three academicians with SIM-plan and/or medical informatics backgrounds, and data managers from the Heart Center.

The general outline of the project was initially planned and discussed with the project team. The developmental process, activities, data collection, data analysis, challenges, progress, and outcomes were discussed in the monthly meetings.

Project Initiation

In this step, more than 20 internal documents from the Heart Center and the Hospital were reviewed. This entailed strategic plans, IT strategies, annual plans, and annual reports for the last five years. In these documents, various business goals (about improving the quality of care, patient safety, communication with partners, etc.) were collected.

In addition, some related IT goals were collected. Next, a series of meetings and discussion sessions with stakeholders were arranged and attended. These activities provided a preliminary understanding of the organization's settings and IT needs.

It was clear that the Heart Center has business and financial autonomy (with central business alignment). However, it was found that IT decision-making is centrally governed by the Hospital, with the possibility of decentral IT planning, which still should be centrally aligned, i.e., a federal IT governance model. Moreover, eight semi-structured interviews with the Heart Centre's executives were conducted. All interviews were recorded with permission.

Participants were interviewed about business goals, IT goals, and the current and future HIS situation. During the interviews, the collected goals and issues were verified with the respective managers.

Subsequently, two preliminary lists of 50 business goals and 29 IT goals were compiled. These lists showed variations in focus among the sub-department managers, probably due to their roles and specific IT needs. The collected business goals and IT goals were clustered, by induction, into nine business goals and six IT goal areas, respectively (Tables 3 and 4).

In addition, a modified Delphi method was applied to obtain the eventual nine and six goals areas. A goal area represents a general strategic aim, which was named according to the relation between the underlining goals. For example, the business goal area "Quality and safety monitoring" includes all goals aimed to improve the quality of care and patients' safety. Each goal area has a list of related goals, and each goal has its specific objectives.

Table 3. The business goals areas of the Heart Center SIM-plan (2018-2022)

Goal area	The business goals areas
I	Quality and safety monitoring
II	Value-Based healthcare competition
III	Staff engagement
IV	Inter-institutional healthcare data exchange
V	Patient care processes
VI	Patient cantered care
VII	Research mission
VIII	Academic mission
IX	Efficient financial system

Table 4. The IT goals areas of the Heart Center SIM-plan (2018-2022)

Goal area	The IT goals areas
I	Accessibility and availability of the information
II	Communications and IT infrastructure
III	Use and privacy of information
IV	Academic and research missions
V	Business processes and services
VI	Users' and patients' requirements

For the strategic business alignment, the critical success factor approach was applied (24). Which is a top-down approach that first identifies factors critical to the hospital's success or failure.

After applying this approach, some goals from the definitive business goals and IT goals lists were excluded, because they were not aligned with the Hospital goals. This helped ensure multiple levels (Hospital and Heart Center) of business and IT goal alignments. These lists were further discussed with the project team.

A list of top-priority business goals was made by the project owners, based on their experience with the most prevalent issues (table 5). The end-products of this step were 2 lists for business goals (15 goals and 38 objectives) and IT goals (6 goals and 26 objectives). All IT goals were aligned with the business strategies of both the Heart Center and the Hospital.

Table 5. The top priority business goals for the Heart Center (2018-2022)

No.	The top priority business goals
1.	Maximizing staff engagement
2.	Geographical expansion of the organization
3.	Providing a transparent quality monitoring
4.	Striving to provide and achieve a high level of value-based competition in healthcare
5.	Providing IT solutions to facilitate patient care
6.	Improve communications with partners
7.	Effective participation in the research mission
8.	To have an effective and efficient financial system

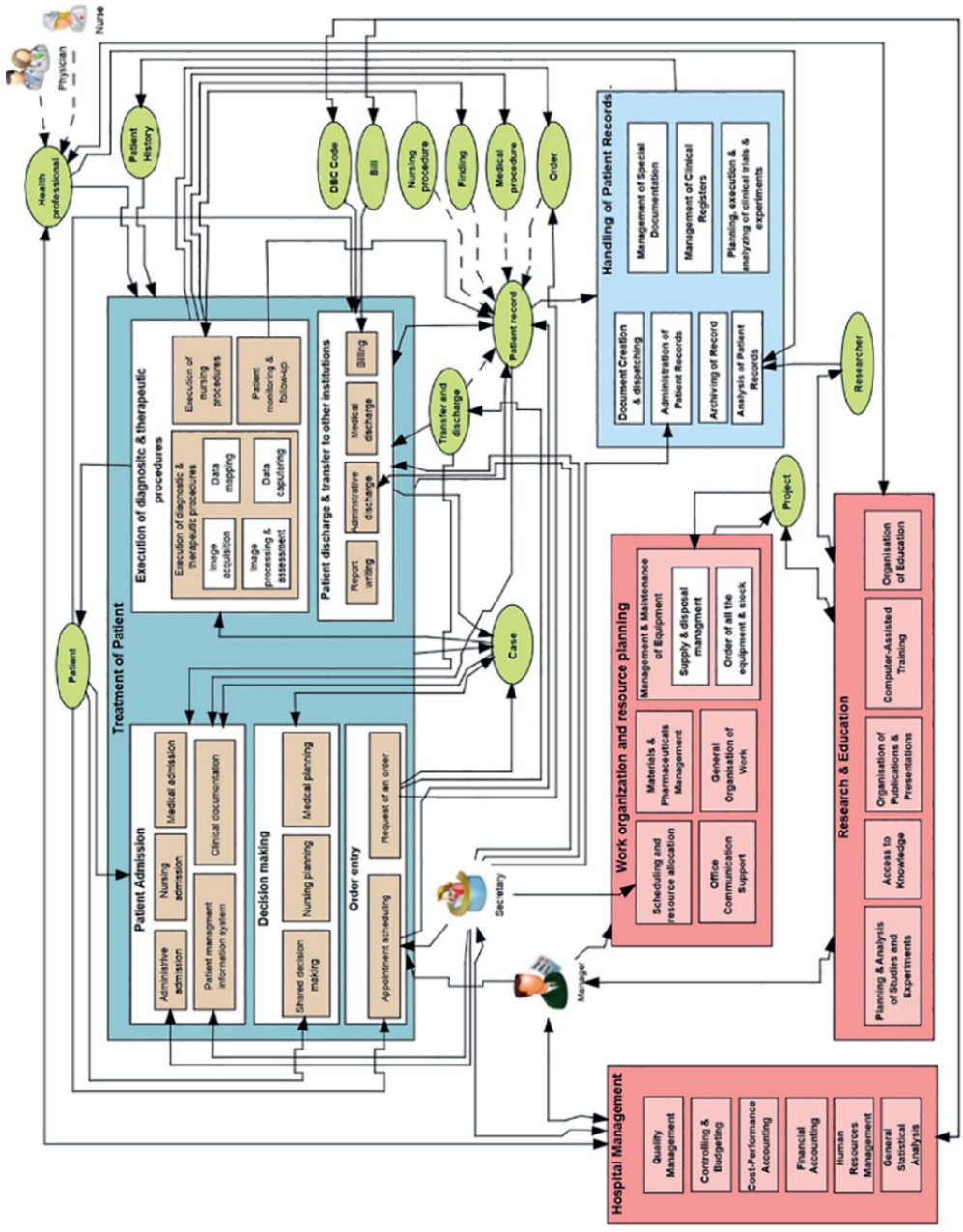
Capturing and assessing the current HIS situation

The next step is capturing the current HIS situation and assessing it. The collected data from the first step helped to identify the IT weaknesses, IT needs, IT opportunities, and current and future HIS situation from the top executives' viewpoint. To comprehensively capture and assess the HIS situation from all perspectives, 42 semi-structured interviews were conducted with stakeholders (physicians, nursing managers of each sub-department, staff nurses, IT managers, data managers, admins, finance staff, researchers, data quality managers, supporting staff). Participants were interviewed about IT needs IT issues, and current and future HIS situations. Stakeholders were interviewed about the current and future HIS situation (for their respective sub-departments).

Additionally, the sub-departments and clinical wards were visited (eight visits) to observe business workflow and get a closer insight into IT needs. To guide the current HIS situation assessment, the "Requirements Index for Information Processing in Hospitals" was used (25). This is a catalog for all hospital functions. To assess the current HIS situation, the quality assessment criteria were applied (1).

The collected data were further used to perform a strengths, weaknesses, opportunities, and threats (SWOT) analysis. Furthermore, to completely understand and assess the current HIS situation, the HIS architecture was modeled, using the 3LGM^2 tool (a meta-modeling tool for modeling Health Information Systems) (1, 23). The current HIS architecture was characterized by the presence of a central electronic patient record (EPR) and many (best-of-breed) application components (AC) (Fig 1).

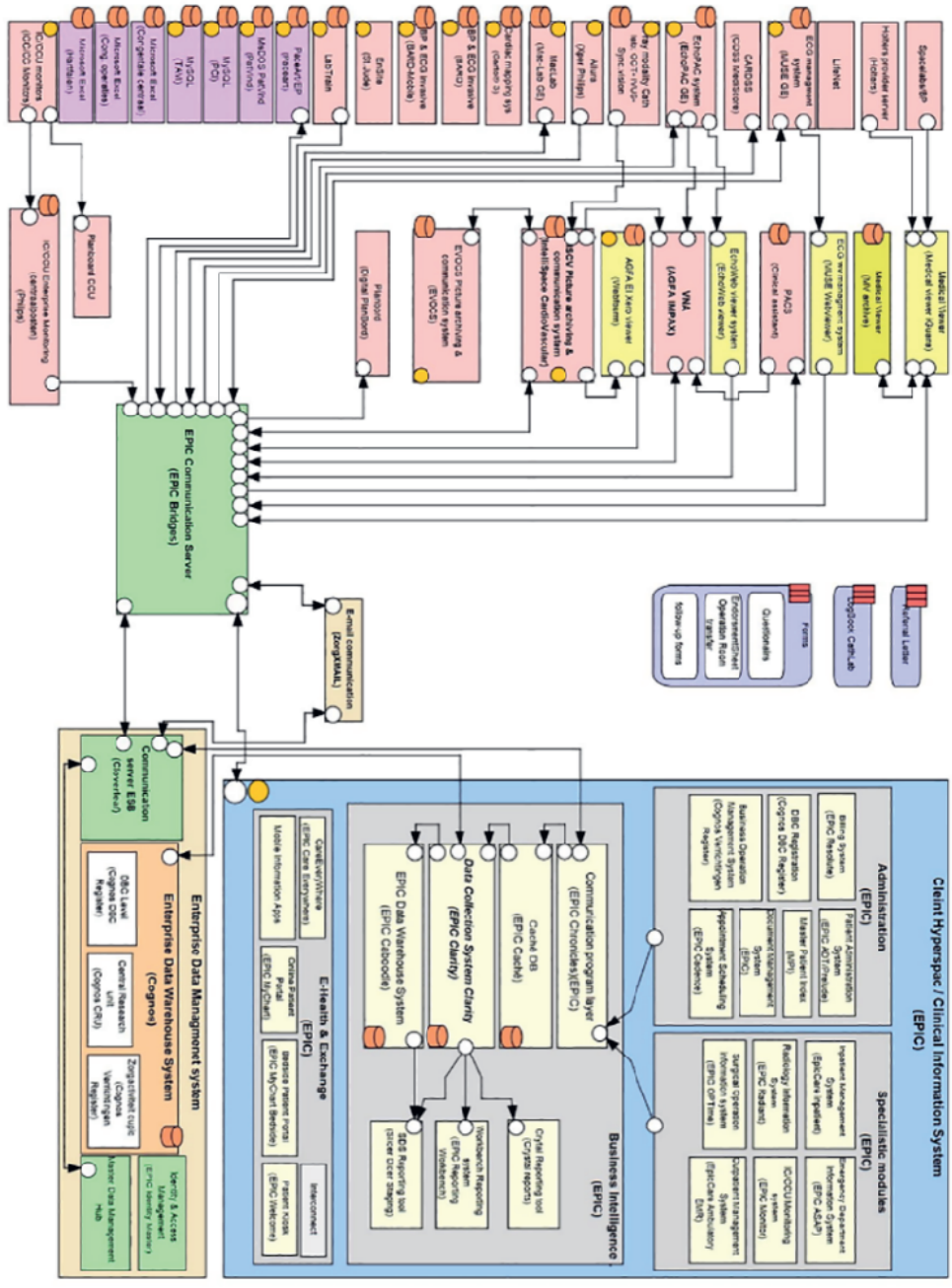
Figure 1. The identified current domain layer using the 3LGM[2]



The domain layer (first layer) defines the heart center by its enterprise functions and the relations to each other by system users and entity types, who can update or interpret those functions. The domain layer presented here of the current HIS architecture represents all the identified hospital (enterprise) functions together with the system users and entities.

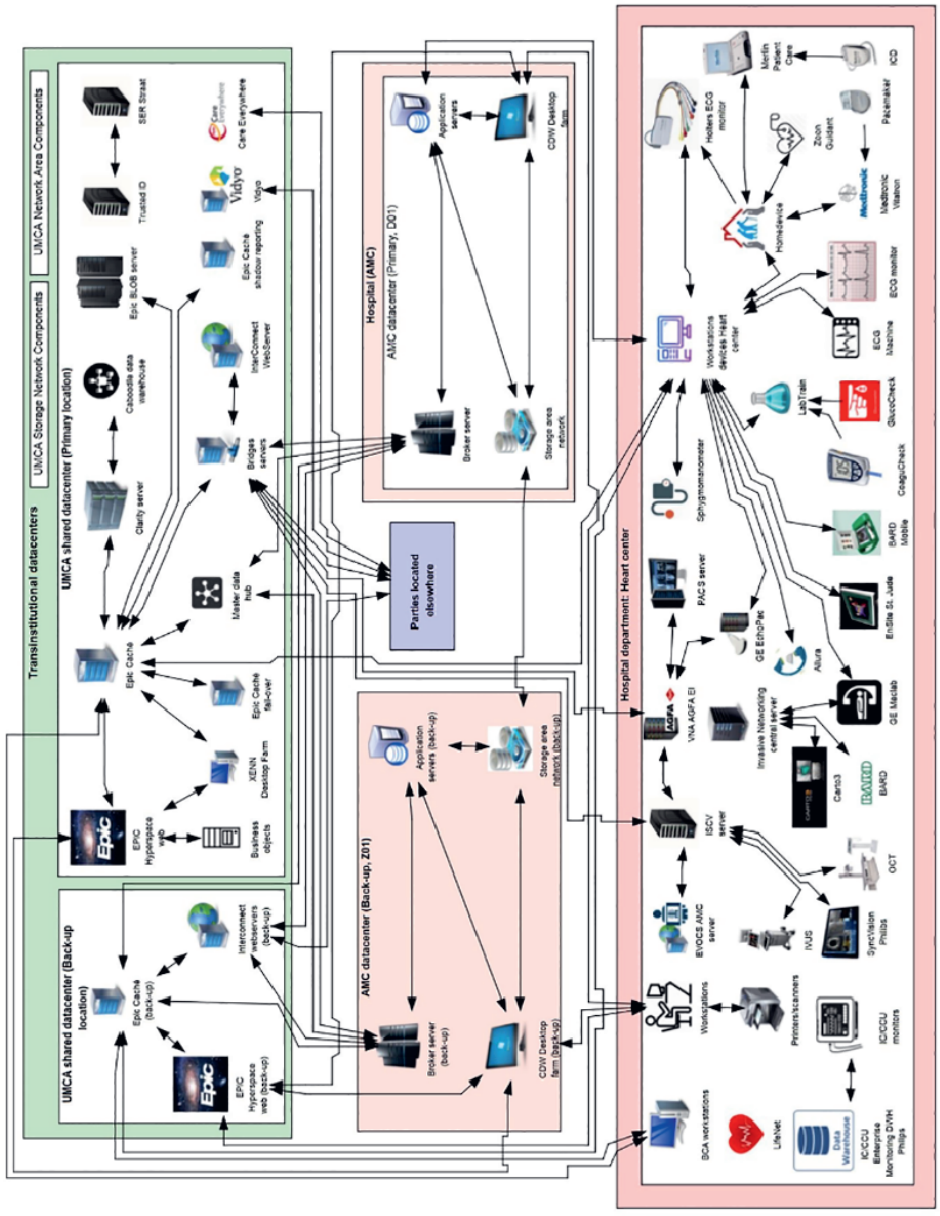
HIS architecture was fragmented and did not offer the required support for some hospital functions (Fig 2-3).

Figure 2. The current identified logical layer using the 3LGM[2]



The logical tool layer (second layer), is the application components, which support the hospital functions. The working staff or the system users (actors) are executing the business processes (functions) by updating the enterprise entities, using the applications and functionalities of the heart center and the Hospital. These business processes (functions) together with the system users and entities are the integral constituents of the Domain layer.

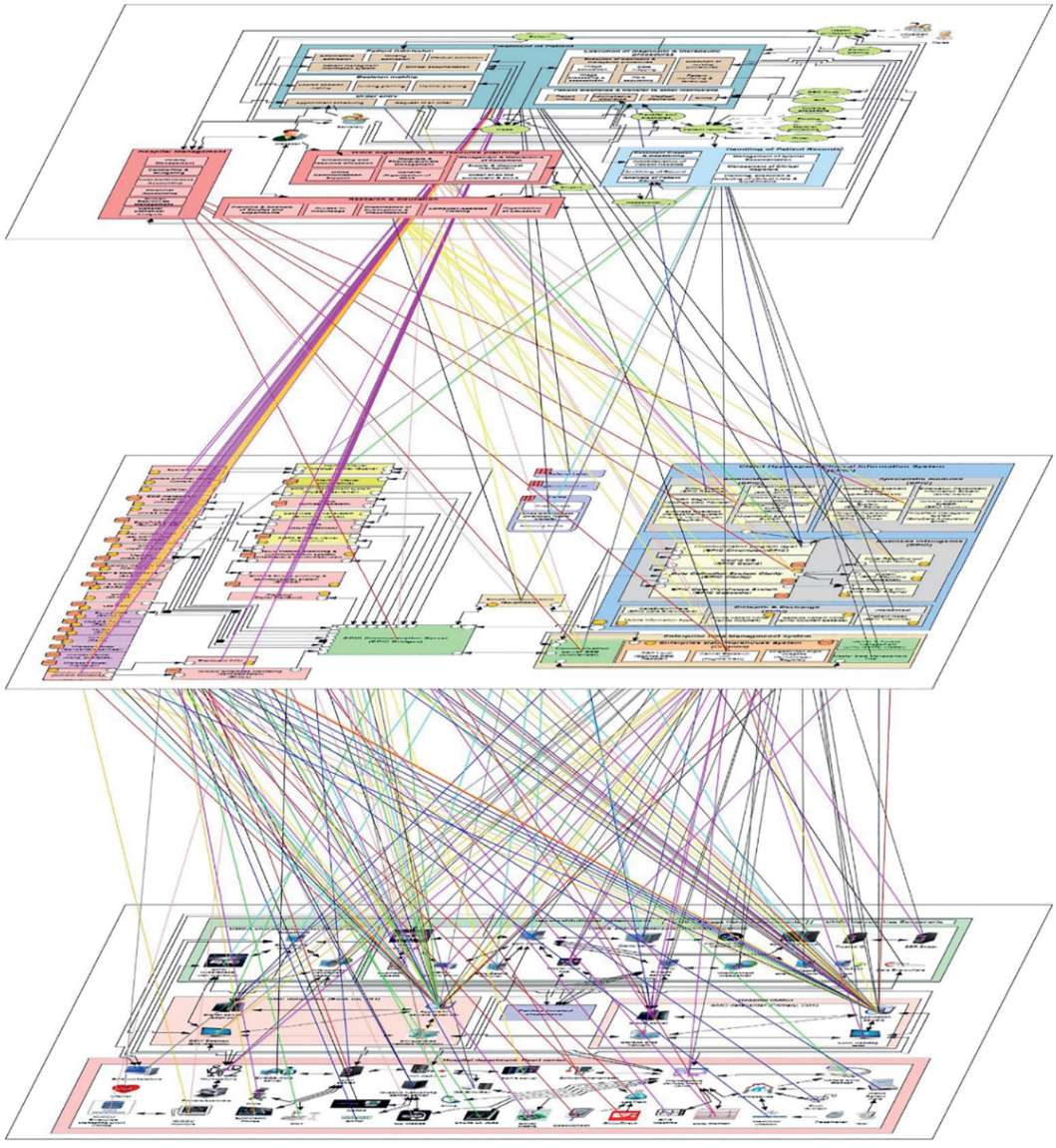
Figure 3. The current identified physical layer using the 3LGM²



The physical tool layer (third layer) contains a set of physical data processing systems, on which the application components are running and supported.

The Heart Center uses many standalone ACs and physical data processing tools (e.g., devices) to support the various hospital functions (e.g., patient care, research, and education). Examples of the used ACs and devices are database management systems, vital signs monitors, pacemakers' programming machines, imaging modalities, and different portable devices (Fig 4).

Figure 4. The currently identified interrelation of the three layers using the 3LGM² tool



There was no direct data integration between most of the ACs and the EPR. The generated data are kept in different storage places (Fig 2-4). The Heart Center had adopted various databases to collect data from the various ACs, to support the quality of care and research (Fig 2-4). These situations complicate data retrieval and reuse.

The Heart Center has significant IT needs especially for improving systems integration and interoperability. A commonly reported IT issue during the interviews – also observed during HIS-modelling - was data redundancy and the resulting financial loads from dealing with this issue in each sub-department.

The end products of these steps were a complete overview of the current HIS situation and lists of strengths and weaknesses (problems list, and IT needs).

Planning the future HIS situation and the migration path

The next steps of the developmental process were planning the future HIS situation and making the migration path. Planning future HIS situation was performed by translating the IT goals into the intended HIS architecture. The initial consensus between sub-departments was not easily achievable. However, suggesting alternative solutions that could simultaneously realize many IT goals had enhanced consensus between sub-departments (e.g., instituting a dedicated data warehouse).

Furthermore, the goal and problem lists were used to describe the difference (the gap) between the current and the future HIS situation. The gap was wide in some aspects, most notably in: HIS future planning (Hospital-wide vs. Heart Center), systems integration (interoperability), providing the required support for some hospital functions, data registration and extraction, communication with partners, tools to improve quality management and control, utilizing decision support systems (such as risk scores and mortality prediction models), and employing advanced research and analytics tools. After completing planning, the future HIS situation, the migration path (roadmap) for implementing the SIM-plan was formulated.

Writing the SIM-plan document

After completing all the steps, the collected data and end products were allocated to the relevant chapters in the SIM-plan document. The document has several chapters and is arranged according to the sequence of the developmental process steps.

Moreover, each sub-department of the Heart Center has one chapter (mini-SIM-plan), to describe its own IT needs and IT plans. The SIM-plan document was discussed with the project owners and project team. All necessary revisions and changes were made.

SIM-plan approval

The last step was to submit the SIM-plan document to the head of the Heart Center for approval. Subsequently, a final oral presentation about the SIM-plan document and the generated strategic project portfolio was given to the stakeholders.

Developing the strategic project portfolio

The SIM-plan document formed the basis for developing the strategic project portfolio. To generate the project's portfolio the "problem-oriented approach" was applied. This approach considers denoting change actions against the weak points of the HIS architecture, then matching them to the related IT goals, and translating this into feasible IT projects.

The project portfolio has a concrete list of 25 projects. These projects will implement the goals and objectives of the SIM-plan (table 6). The list was then prioritized (three urgent projects) based on the top priority goals.

Table 6. The strategic project portfolio of the Heart Center (2018-2022), (6 out of the 25 projects)

	The strategic project portfolio
1.	Improving data registration by implementing a cardiology-specific module in the electronic patient record
2.	Instituting a dedicated Data warehouse for the Heart Center
3.	Improving the quality of care
4.	Investing in areas of excellence (e.g., Transcatheter Aortic Valve Implantation "TAVI")
5.	Providing advanced tools to support the research mission
6.	Providing quality monitoring tools

SIM-planning outcomes

SIM-planning helped to change problem-solving and IT planning styles for many stakeholders. Stakeholders adapted strategic IT planning styles, instead of providing ad hoc solutions based on former knowledge. A raised stakeholders' awareness about the SIM-planning has been noticeably observed. This helped to minimize the individual (uncoordinated) IT projects. SIM-planning helped to coordinate the ongoing IT projects. In terms of the usefulness of the developed SIM-plan, and after SIM-plan presentation, six IT projects from the project portfolio were launched. These projects were about data-warehousing (e-supplementary figure 1), improving data registration, data integration, communications with partners, and developing mortality prediction models.

Important lessons learned

Before and during the development process, a couple of important lessons have been learned, which are of interest to other similar efforts. The lessons are elaborated point by point in this section as follows:

The SIM-planning was started at sub-department levels (bottom-up approach). Later, it appeared that some new IT goals were not aligned with central strategies (of the Heart Center and the Hospital as well). Hence, it is better to do top-down SIM-planning by identifying (and understanding) the higher-level goals, and then proceeding to lower levels. Understanding the hospital-wide IT governance arrangements helped in producing multiple levels of aligned IT goals (minimize conflicts).

In the beginning, several stakeholders requested individual standalone applications, infrastructures, or solutions to improve their departmental workflow and cooperation. These solutions would further fragment the HIS architecture into an undesired situation. Still, it is difficult to create acceptance and consensus amongst stakeholders. However, providing an alternative standardized solution that might help all the sub-departments simultaneously enabled alleviate this situation. Moreover, engaging stakeholders had indeed reinforced the SIM-planning. Performing a SWOT analysis helped uncover important details that enriched the SIM-plan with relevant potentially realizable plans.

The sequential order of the Practical Guideline steps formed useful guidance during the process. The findings of each step formed substrates to the succeeding steps. It was still required to use additional data collection activities to enhance the developmental process (see below).

Observations about using the Practical Guideline

The Practical Guideline was proposed to develop SIM-plans for hospitals, which implicitly implies the possibility of using the Practical Guideline to develop SIM-plans for clinical departments. However, developing SIM-plans at the department levels needs specific considerations. Developers should investigate the IT governance arrangements at all organizational levels. This is important to avoid conflicts later in the process. In our case, a careful investigation of all IT levels of governance arrangements was performed. However, the Practical Guideline did not specify recommendations to investigate these arrangements. Therefore, it would be better to consider adding such a recommendation when updating the Practical Guideline.

The Practical Guideline provided a relatively narrow spectrum of data collection activities. For example, to identify the goals, the Practical Guideline recommended initiating meetings and discussion sessions. In our study, starting the SIM-planning with these activities was not possible. This is because the SIM-plan developer was a new employee at the Heart Center. These activities might fit settings where the developer is a known (senior) staff member (e.g., chief information officer). Therefore, interviews with top executives were conducted as initial activities.

Throughout the developmental process, additional data collection activities were used, such as visiting departments and observing workflow (9), and drawing the enterprise core diagram (26). The Practical Guideline would better add a wider scope of data collection activities.

Discussion

This study showed that the SIM-plan developmental process had enabled a comprehensive exploring the HIS strengths and opportunities for the Heart Center. The developmental process had enabled identifying weaknesses including insufficient AC integration, suboptimal alignment between central and departments' IT plans, and lack of the breadth of the strategic goals and lack in IT related details to align with these goals. The

SWOT analysis has enhanced the developmental process and enriched the end products of each step. This study demonstrated the Practical Guideline's usefulness as an empirical approach to developing a SIM-plan in clinical departments. However, this study suggested improvements in the Practical Guideline in terms of improving the actionability of the steps (missing advice on interviews and visiting departments), and the coverage and concreteness of the Practical Guideline (missing guidance on analyzing IT governance arrangements). The main SIM-planning outcomes were developing the SIM-plan document and the strategic project portfolio, which are important to direct the succeeding phases "SIM-plan implementations phase".

Strengths and limitations

To the best of our knowledge, this is the first study that described the SIM-plan developmental process using the Practical Guideline in a clinical-department (Heart Center) within a large hospital. This study represents a case at a single center that has its characteristics, which might limit the generalizability of the results. However, such a study of a particular center with three major clinical departments in a large academic hospital reflects the complexity of various clinical departments, and yet such cases are uncommon in the SIM-plan literature. The developer's background (medical doctor and medical informatics) could be considered a strength of this study. The advantage of this is having both the knowledge of being an end-user healthcare professional, and being a SIM-planning developer. This would minimize the efforts that would be spent by the information officers to understand the hospital functions and entities. Recording interviews and data triangulation strengthened the methods. It is worth mentioning that these two activities improved the validity of the collected data and the analysis. Moreover, they revealed several issues that were not spotted during the interviews. A large number of interviews, as well as meetings, and discussion sessions, have also strengthened the methods.

The implication of the study

This study provides evidence-based practice for the usefulness of the Practical Guideline to SIM-planners, IT managers, managers commissioning SIM-plans development, and researchers assessing current SIM-plans and/or developing new SIM-planning approaches. Finally, the study can be used as an example for educational purposes and for training pertaining to applying evidence-based practices in hospitals and SIM-planning for clinical departments as well as hospitals.

Recommendations for further research

In future work, it is recommended to conduct more rigorous studies on SIM-plan development in diverse clinical settings in different countries, based on the Practical Guideline approach. Researchers might consider performing larger studies that include different facilities with various organizational settings, in the same study, and compare the results of developing SIM-plans between those facilities. However, this kind of study might be difficult to conduct, as each organization has its policies and structure, and the stakeholders might be reluctant to participate in such studies.

Conclusion

The systematic empirical use of the Practical Guideline enabled the development of a SIM-plan in a complex healthcare facility (heart center). The development process allowed business and IT goal identification and having a complete overview of the current HIS architecture and enlisting all its strong and weak points. This has aided the planning of the future HIS architecture and allowed the development of the project portfolio. The study showed that aligning IT goals with business goals is an indispensable task and an important issue for healthcare facilities to manage the HIS and realize their mission and values.

Funding sources

The authors received no financial support for the research, authorship, and/or publication of this article.

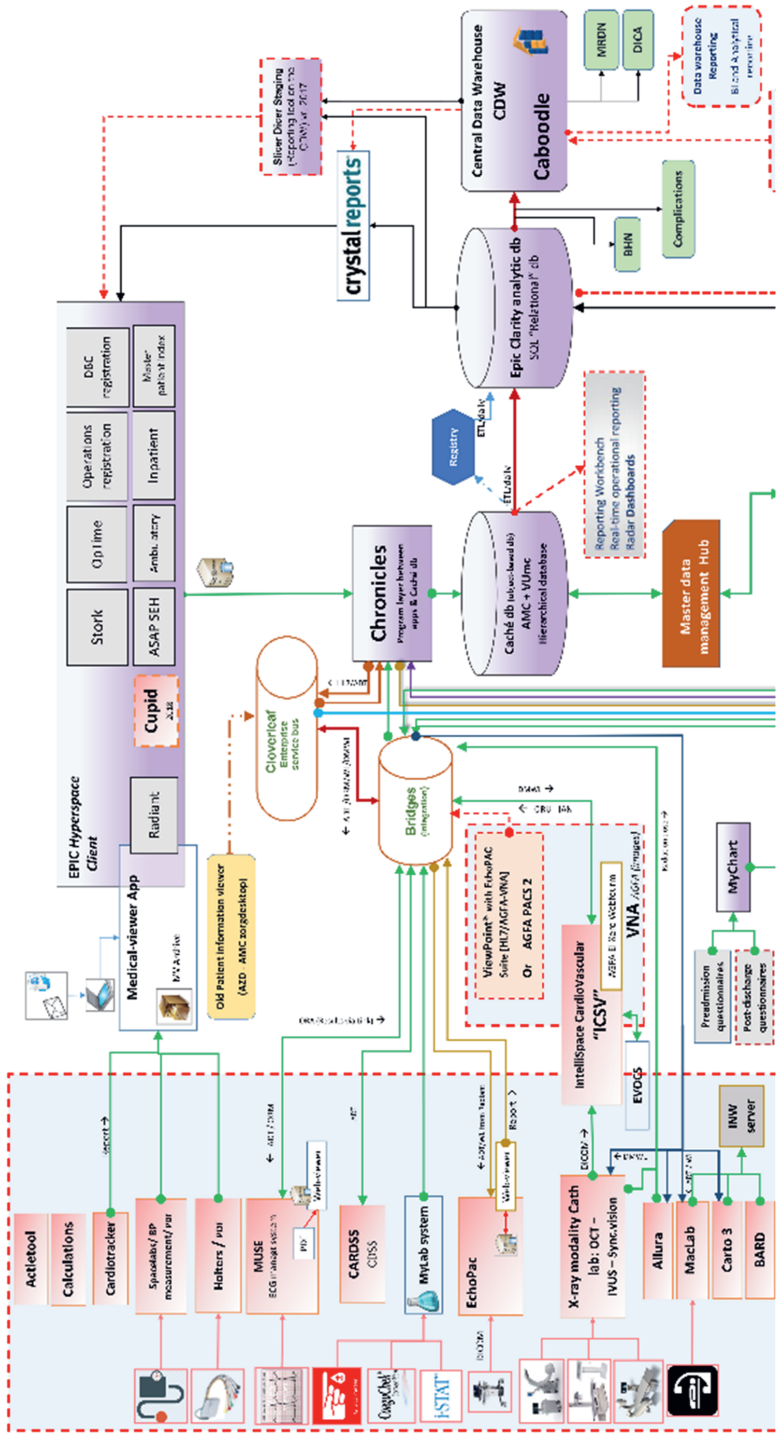
Declaration of Competing Interest

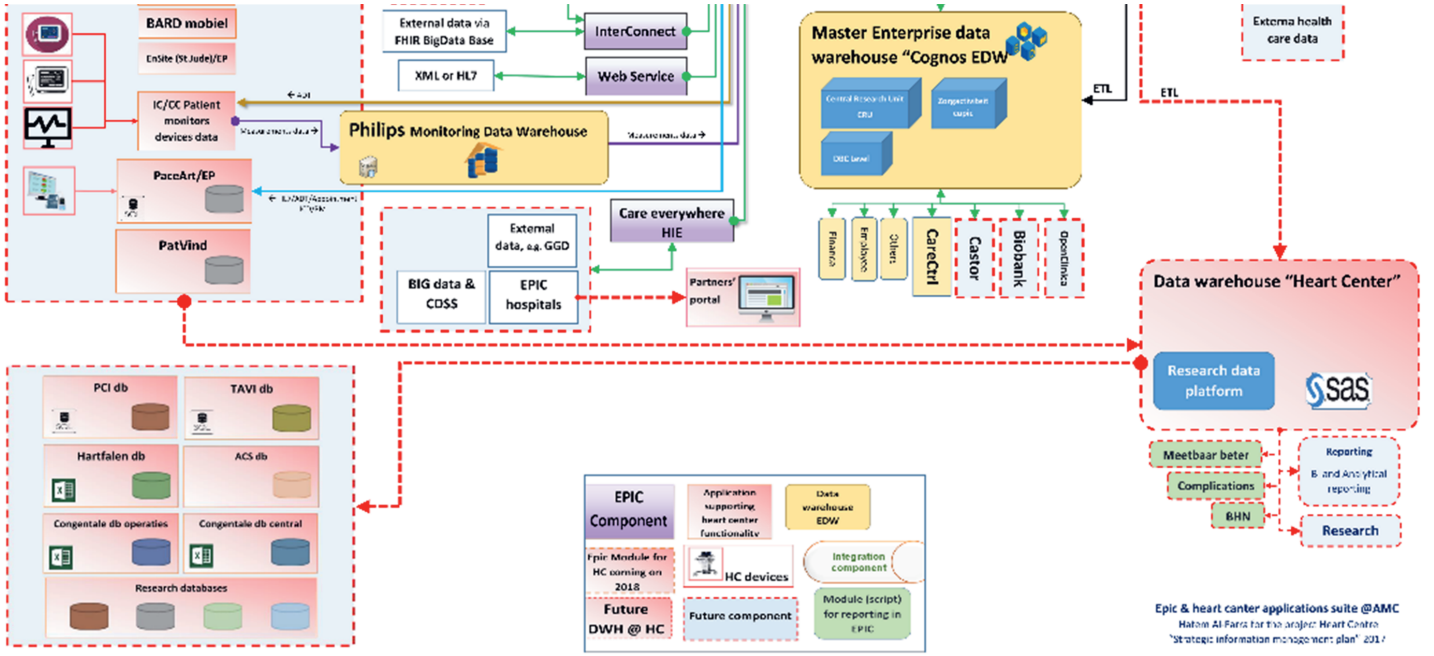
The authors declared no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

References:

1) Alfred Winter RH, Elske Ammenwerth, Birgit Brigl, Nils Hellrung, Franziska Jahn. Health Information Systems: Architectures and Strategies. 2nd ed. London: Springer Science & Business Media; 2011 oktober 2010. 340 p.
2) Winter A, Brigl B, Buchauer A, Dujat C, Graber S, Hasselbring W, et al. Purpose and structure of strategic plans for information management in hospitals. Stud Health Technol Inform. 2000;77:880-4.
3) John P. Glaser CS. The strategic application of information technology in healthcare organizations. 3rd edition ed: San Francisco: Jossey-Bass; 2011 Mar-2011.
4) Rosenberger HR, Kaiser KM. Strategic planning for healthcare management information systems. Healthcare Manage Rev. 1985;10(1):7-17.
5) Bush M, Lederer AL, Li X, Palmisano J, Rao S. The alignment of information systems with organizational objectives and strategies in healthcare. Int J Med Inform. 2009;78(7):446-56.
6) Brigl B, Ammenwerth E, Dujat C, Graber S, Grosse A, Haber A, et al. Preparing strategic information management plans for hospitals: a practical guideline SIM plans for hospitals: a guideline. Int J Med Inform. 2005;74(1):51-65.
7) Winter AF, Ammenwerth E, Bott OJ, Brigl B, Buchauer A, Graber S, et al. Strategic information management plans: the basis for systematic information management in hospitals. Int J Med Inform. 2001;64(2-3):99-109.
8) Gunasekaran S, Garets DE. Business value of IT: the strategic planning process. J Healthc Inf Manag. 2003;17(1):31-6.
9) Lee T, Ghapanchi AH, Talaei-Khoei A, Ray P. Strategic Information System Planning in Healthcare Organizations. J Organ End User Com. 2015;27(2):1-31.
10) Moriarty DD. Strategic information systems planning for health service providers. Healthcare Manage Rev. 1992;17(1):85-90.
11) Taylor TB. Information management in the emergency department. Emerg Med Clin North Am. 2004;22(1):241-57.
12) (1) 12. van Bemmel JH, editor An international perspective on information management and technology in healthcare. Proceedings of Conference on Clinical Information, London; 1993.
13) Lederer AL, Sethi V. Guidelines for strategic information planning. Journal of Business Strategy. 1991;12(6):38-43.
14) Killingsworth B, Newkirk HE, Seeman E. An integrative health information systems approach for facilitating strategic planning in hospitals. Healthcare Manage Rev. 2006;31(2):119-29.
15) Belkin M. The assessment of information systems effectiveness in private and hospital pathology: Business Information Technology, RMIT University; 2009.
16) Bricknall R, Darrell G, Nilsson H, Pessi K. Aligning IT Strategy with Business Strategy through the Balanced Scorecard in a Multinational Pharmaceutical Company. 2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07); 2007 40th Waikoloa, HI, 20072007. p. 235b-b.
17) Hubner-Bloder G, Ammenwerth E, Brigl B, Winter A. Specification of a reference model for the domain layer of a hospital information system. Stud Health Technol Inform. 2005;116:497-502.
18) Brigl B, Hubner-Bloder G, Wendt T, Haux R, Winter A. Architectural quality criteria for hospital information systems. AMIA Annu Symp Proc. 2005:81-5.
19) Iveroth E, Fryk P, Rapp B. Information technology strategy and alignment issues in healthcare organizations. Healthcare Manage Rev. 2013;38(3):188-200.
20) Lammintakanen J, Kivinen T, Saranto K, Kinnunen J. Strategic management of healthcare information systems: nurse managers' perceptions. Stud Health Technol Inform. 2009;146:86-90.
21) Mettler T, Fitterer R, Rohner P, Winter R. Does a hospital's IT architecture fit with its strategy? An approach to measure the alignment of health information technology. Health Systems. 2014;3(1):29-42.
22) Kitsios F, Kamariotou M, Manthou V, editors. Hospital Information Systems Planning: Strategic IT Alignment in Healthcare2019; Cham: Springer International Publishing.
23) Winter A, Brigl B, Funkat G, Haber A, Heller O, Wendt T. 3LGM(2)-Modelling to Support Management of Health Information Systems. Stud Health Technol Inform. 2005;116:491-6.
24) Tan JK. The critical success factor approach to strategic alignment: seeking a trail from a health organization's goals to its management information infrastructure. Health Serv Manage Res. 1999;12(4):246-57.
25) Haux R. Requirements Index for Information Processing in Hospitals: Department of Medical Informatics, Institute for Medical Biometry and Informatics, University of Heidelberg; 2001 (cited 2018 December). Version 1.0:(Available from: https://iig.umit.at/images/projects/documents/requirement%20index.pdf.
26) P. RJW. Enterprise Architecture as strategy- creating a foundation for business execution. Boston, Massachusetts. : Harvard Business school press; 2006.

Appendix. E-components figure 1




BARD mobiel
EnSite (St Jude)/EP
IC/CC Patient monitors devices data
PaceArt/EP
PatVind
External data via FHIR BigData Base
XML or HL7
InterConnect
Web Service
Philips Monitoring Data Warehouse
Care everywhere HIE
External data, e.g. GGD
BIG data & CDSS
EPIC hospitals
Partners' portal
Master Enterprise data warehouse "Cognos EDW"
Finance
Employee
Others
CareCtrl
Castor
Biobank
OpenClinica
ETL
ETL
Externa health care data
Data warehouse "Heart Center"
Research data platform
sas
Meetbaar beter
Complications
BHN
Reporting
& and Analytical reporting
Research
PCI db
TAVI db
Hartfalen db
ACS db
Congentale db operaties
Congentale db central
Research databases
EPIC Component
Application supporting heart center functionality
Data warehouse EDW
Epic Module for HC coming on 2018
HC devices
Integration component
Future DWH @ HC
Future component
Module (script) for reporting in EPIC
Epic & heart canter applications suite @AMC
Hatam Al Faraj for the project Heart Centre
'Strategic information management plan' 2017

Chapter 7:

Wrap-up, discussion, and future developments

Wrap-up, discussion, and future developments

The thesis describes how to improve the preoperative prediction of TAVI-related risk of 30-day mortality and morbidity. The main focus has been on the improvement of existing international prediction models used for TAVI by adjusting or developing models with data from The Netherlands Heart Registry (NHR).

The ultimate goal of this work is to improve patient safety and contribute to the quality of TAVI-care. Currently different international models are in use to predict TAVI-related mortality in many countries over the world, including The Netherlands. Examples are the EuroSCORE-I and –II (1, 2), STS-PRoM (3), OBSERVANT (4), German-AV (5), FRANCE-2 (6), and ACC-TAVI (7). There is no evidence about the validity and predictive performance of these models for TAVI patients in The Netherlands. It is important to examine their validity and usefulness for the Dutch TAVI population before recommending one of these models for daily clinical use in The Netherlands. Subsequently, we examined if updating the best-performing models indeed improves predictive performance. Although that was the case, the research also resulted in the development of a redesigned TAVI-specific mortality prediction model based on a recent Dutch TAVI cohort retrieved from NHR. To this end, we had to address several sub-questions, the answers of which are presented in the respective chapters.

The first research question was to which degree the current prediction models performed when applied to the Dutch population. Chapter 2, describes that in 6177 TAVI-procedures, retrieved from 2013-2018 in the TAVI-NHR registry, the observed early-mortality rate was 4.5% (n=280). The analysis showed that each of the investigated models when externally validated had a different ability to predict the early-mortality. In this analysis, the ACC-TAVI model scored best with an Area under the Receiver-Operating-Characteristic curve (AU-ROC) of 0.64 (95% CI 0.61-0.67). The model FRANCE-2 scored an AU-ROC of 0.63 (0.60-0.67). However, these AU-ROCs were inferior to the AU-ROC reported in the original development studies of these models (2-7).

The findings obtained from this analysis confirm the results of other external validation studies performed in other countries, were TAVI-models reported to perform poorly on external populations (8-12). In this study, and unlike other external-validation studies, the deployment of a comprehensive set of predictive performance measures allowed for obtaining a wide overview of model performance.

Due to the observed sub-optimal performance of the existing TAVI-models, their use in clinical practice, at least in The Netherlands, is indeed questionable. The analysis done in this study revealed the need to either update the best-performing existing models, or even to develop and validate a new TAVI-specific prediction model, which we did in the subsequent studies.

Chapter 3, it was investigated the impact could be of updating the two best-performing models of chapter 2 (ACC-TAVI and FRANCE-2) on our population. Both models were updated by applying the best update method selected from different available model update strategies (13) and subjecting the models to an existing closed-testing procedure (14) for the entire NHR TAVI-cohort.

For the model FRANCE-2, the selected update-method, using the closed-testing procedure, was model-intercept-update. The predicted-mortality of the updated-model was 4.8% (vs 7.4% predicted mortality for the original un-updated model on our cohort). When the updated-model was internally validated, the AU-ROC was 0.63 (95%CI 0.62-0.66). In the calibration-graph, the updated model showed an overestimation of the predicted early-mortality. Calibration-intercept and calibration-slope of the updated-model significantly deviated from their ideal values.

When this model was temporally validated, the AU-ROC was 0.61 (95%CI 0.53-0.67). The selected update-method for ACC-TAVI was model-revision, which resulted in a predicted-mortality of 4.5% (vs. 4.4% for the original un-updated model). This updated-model, ACC-TAVI, has an AU-ROC of 0.63 (0.63-0.66) in internal-validation. In the calibration-graph, the updated-model was shown to be only calibrated up to the risk of 20% of the predicted mortality, and afterward, it underestimated the predictions. Both the calibration-intercept and calibration-slope of this updated-model significantly deviated from their ideal values associated with an AU-ROC of 0.65 (95%CI 0.58-0.72) in temporal-validation.

It was concluded that updating the models (FRANCE-2 and ACC-TAVI) indeed reached their original performance in internal-validation (6, 7) for application to a new population. Additionally, the predictive performance of the

models was better than the performance reported in our first external-validation study (chapter 2) and other external-validation studies (8, 9).

It was inferred that the updated revised version of the ACC-TAVI could potentially be the best currently available prediction model for TAVI-related early-mortality in the Dutch TAVI population. It included more parameters available in the NHR registry, while the FRANCE-2 model needs one variable that is not routinely registered in the NHR registry. Nevertheless, the predictive ability of the updated-model ACC-TAVI model remained suboptimal with an AU-ROC of 0.65.

This justifies the recommendation that not only in The Netherlands but also in other countries, model updates for specific populations should be considered, or a newly developed and validated TAVI-specific prediction model should be attempted.

As a consequence of this recommendation, we aimed to develop and validate a novel prediction model for early-mortality after TAVI based on NHR, called the TAVI-NHR risk model. The results of the new TAVI-specific model are presented in chapter 4. Data from 9144 TAVI patients from the NHR were used to develop and internally validate TAVI-NHR. The data represents a cohort of patients who underwent a TAVI procedure for six years from January 1 2013 to December 31, 2018. The observed early-mortality rate in the cohort was 4.0% (N=368).

As mentioned, TAVI-NHR has ten variables, including age (in years), serum creatinine, left ventricular ejection fraction, body surface area, NYHA class, procedure-acuteness status, chronic lung disease, critical-preoperative state, diabetes-mellitus status, and TAVI access-routes. Body surface area and diabetes mellitus emerged as new predictors that were not used in the currently available TAVI-specific models. At internal-validation, the median AU-ROC of the TAVI-NHR model was 0.69 (IQR 0.646-0.75). There have been no signs of miscalibration observed. The AU-ROC in temporal-validation was 0.71 (95% CI 0.64-0.78), which is better than the updated ACC-TAVI model, which was suboptimal with an AU-ROC of 0.65.

In this study, the performance of the new model was compared to updated versions of the models ACC-TAVI (7) and IRRMA (9) on our population. It was observed that the TAVI-NHR model outperformed both updated models (p-value <0.05). The study concluded that the TAVI-NHR model is superior in early-mortality risk assessment to the currently available TAVI-specific models even after their adjustment to our population. It should be noted that TAVI-NHR included new variables, which seem to have yielded improved discrimination and good calibration. In terms of clinical relevance, the model's specific feature of good calibration may enable good pre-procedural risk-assessment and patient counseling. In addition, this work resulted in a computer-based dynamic nomogram with a user graphical interface tool, which may assist the heart team with the selection of TAVI-patients.

Beside parameters relevant to predicting mortality, the NHR registry contains data on TAVI-related minor and major adverse cardiac events (MACE). This data constitutes information about the quality of TAVI-care in The Netherlands as well as in the participating centers. As described in chapter 5, we used the same cohort of chapter 4 to describe the prevalence of MACE, including major vascular bleeding, the need for permanent pacemaker implantation, and stroke. It was investigated whether mortality and adverse events did occur at a larger rate than expected in particular subgroups.

The share of TAVI treatments increased over time and, in accordance with an indication for TAVI, the majority of patients were 80 years or older. During the observation period, early- and 1-year mortality rates were 4.0% and 11.6%, respectively, and 10.8% for permanent pacemaker implantation, 3.2% for major vascular bleeding, and 1.7% for stroke. Although the number of interventions significantly increased, mortality decreased significantly, but rates of MACE remained stable. This trend is comparable to observations in other countries (15-26).

The results observed in this cohort remained stable in the following years 2019 and 2020 according to the annual NHR reports with a trend towards a further decrease of adverse events, except for an increase in stroke rate (27). Further subgroup analysis showed that elder high-risk patients appeared most vulnerable to MACE and mortality, while younger patients faced an increased rate of permanent pacemaker implantation. These findings suggest that these specific sub-groups at risk may benefit from being treated in those centers which have the best record or in specialized centers.

In accordance with the ambition of the NHR that participants can compare their outcomes and practices with best-performing centers, each hospital is obliged to annually provide a standardized and complete as a possible data set of variables. We envisioned that apart from an administrative duty to comply with NHR requirements for data upload based on its hospital information system (HIS), hospitals should design a comprehensive system of continuous information gathering and analysis, which couples the various sources of information available in hospitals.

Chapter 6 described in detail the development of the strategic information management plan (SIM-plan) for the Heart Center at the Amsterdam University Medical Center. The produced document contains the strategic-goals, the business-goals, and the IT goals of the heart center, between the years 2017-2021. The document provides also a description of the new and the intended HIS situation for the heart center in relation to the entire hospital (28). Moreover, the document delineates how the long-term goals of the heart center can be supported and realized by the HIS. The empirical approach of Brigl et al., called the "Practical-Guideline" was leading in this process (29).

Implementing these activities helped to identify 15 business goals and six IT goals. Moreover, the activities required described and assessed the current HIS situation and identified the IT problems and related issues. The identified architecture of the current HIS has shown a fragmented HIS, with a central electronic patient chart system and many application components, and many standalone systems that are not directly connected to each other. The most frequently reported IT issues were data redundancy and the associated financial burden. In addition, hospital-wide IT governance often collided with practical initiatives and requirements needed to support innovation in the heart center. The evolution of a cardiac center-based information plan must be in alignment and compliant with hospital-wide developments and standards. Cardiac services are not stand-alone or isolated activities but heavily rely on multi-disciplinary collaboration, shared infrastructure, and standardized quality control, which by far extends the obligation to meet for example requirements demanded by the NHR.

The chapter concludes the usefulness of the Practical-Guideline to develop a SIM-plan in healthcare facilities is helpful and necessary. It is evident that also SIM-plans need a continuous update. This demands feedback from users and developers. Future SIM-plans will become more complex due to the availability and actual application of big-data imaging and machine-learning applications.

The main strength of the current work is that it is based on a recent, large, and complete TAVI cohort in The Netherlands. Data used in chapters 2-5, represent the data of all TAVI-procedures performed in the Dutch heart centers, between the beginning of 2013 to the end of 2018. Most of the studies that deal with the validation of prediction models relied on either evaluating the discriminative power (AU-ROC) or calibration graphs. However, in our studies, a comprehensive set of predictive performance measures were evaluated for both internal and external validation. It is an additional asset of our studies that we performed multiple sensitivity analyses for the external validation of existing models, and for the development of NHR-TAVI. With annual NHR reports available, it is to be awaited if and how NHR-TAVI will be used in practice.

Today, shared decision-making for TAVI is common and is therefore individualized and patient-tailored. In the vast majority of cases, outcomes for TAVI are good and are reliably predicted for the individual patient. Predicting is more difficult in extreme- and high-surgical-risk patients. Outcomes also rely on effective and pro-active management of complications in a hospital, which is more critical in high-risk patients. However, in the individual patient, the expected benefits usually outweigh a rather high perioperative risk and there is also a bias in favor of TAVI. Probably, the institutional experience with this class of patients has more impact on decision-making than the NHR-TAVI model. The quality of the NHR data is relevant to the strength of the model. Our research suggests that adding some variables can reinforce the model, but the quality of data may be as important. It is foreseeable that machine-learning technology will be applied as well as big data analysis in conjunction with imaging data and wearable recordings. Therefore, it seems important that at the institutional level experiments with datasets will be organized for exploring large-scale applications. This differs from the use for the sake of quality control and decision support. At the institutional level also patient interests in terms of risk perception and risk-taking can receive more attention.

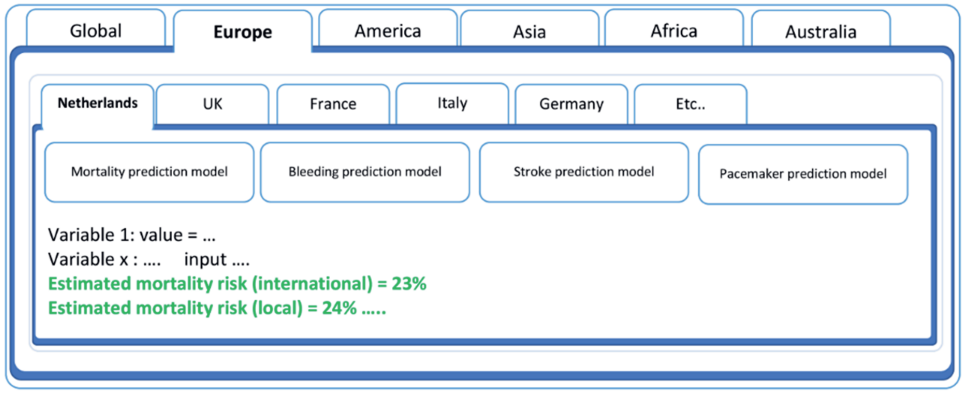
The analysis of the data did not explain in detail the observed trends of improved outcomes in chapter 5. Evidently, growing experience plays a role, but the funnel plots of the annual NHR reports also show differences between participating institutions. Evidently, patient selection and practice variation are present. In order to

investigate its impact, additional data should be made available such as type of device, number of procedures per operator, procedure times, etc., but also patient variables such as cessation of smoking and post-procedural exercise performance. However, currently, a better quality of data is probably easier to realize than including additional data, as for the latter evidence about their utility would be needed.

Another issue to be addressed for future work is the question of who is going to maintain the NHR-TAVI model. Is this an interest of a single research group, a multi-center effort or would it become a part of the NHR ambition to optimize the use of data? In this context, stakeholders should consider a kind of SIM-plan approach in the first place. The end-product could be a portal for standardized data input by participating centers, but it could also provide easy access to tools, models, and initiatives which are continuously updated thanks to the availability of fresh data and insights.

Next, one may question whether a portal initiative should be designed, and supported by single national or joint international registries in Europe in order to mirror the STS-ACC TVT Registry (Society of Thoracic Surgeons-American College of Cardiology Transcatheter Valve Therapy Registry) in the United States of America, see figure below. At the national level, NHR already maintains a complex network of stakeholders and is always available to support initiatives that are in accordance with its mission.

Figure 1: Suggestion for a global portal for predicting TAVI-related outcomes



Without exaggeration, one may conclude that the NHR and other local heart registries are treasures of institutional and national data on TAVI, which indeed are used by professionals to improve patient care and patient safety. However, the yield of these efforts can be increased, and challenges regarding the usability of the models will remain. There is no doubt that permanent maintenance is required as well as the longstanding commitment of professionals and patients.

References:

1) Roques F, Michel P, Goldstone AR, Nashef SA. The logistic EuroSCORE. Eur Heart J. 2003;24(9):881-2.
2) Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, et al. EuroSCORE II. Eur J Cardiothorac Surg. 2012;41(4):734-44; discussion 44-5.
3) 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 Suppl):S23-42.
4) 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.
5) 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(5):971-7.
6) Iung B, Laouenan 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.
7) 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):46-52.
8) Martin GP, Sperrin M, Ludman PF, de Belder MA, Gale CP, Toff WD, et al. Inadequacy of existing clinical prediction models for predicting mortality after transcatheter aortic valve implantation. Am Heart J. 2017;184:97-105.
9) 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 Assessment (IRRMA) study. Int J Cardiol. 2016;215:227-31.
10) Durand E, Borz B, Godin M, Tron C, Litzler PY, Bessou JP, 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.
11) Zbronski K, Huczek Z, Puchta D, Paczwa K, Kochman J, Wilimski R, et al. Outcome prediction following transcatheter aortic valve implantation: Multiple risk scores comparison. Cardiol J. 2016;23(2):169-77.
12) Wolff G, Shamekhi J, Al-Kassou B, Tabata N, Parco C, Klein K, et al. Risk modeling in transcatheter aortic valve replacement remains unsolved: an external validation study in 2946 German patients. Clin Res Cardiol. 2020.
13) Steyerberg E. Clinical Prediction Models, A Practical Approach to Development, Validation, and Updating. New York NY: Springer Science & Business Media, LLC; 2009.
14) Vergouwe Y, Nieboer D, Oostenbrink R, Debray TPA, Murray GD, Kattan MW, et al. A closed testing procedure to select an appropriate method for updating prediction models. Stat Med. 2017;36(28):4529-39.
15) Smith CR, Leon MB, Mack MJ, 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.
16) van Kesteren F, van Mourik MS, Vendrik J, Wiegerinck EMA, Henriques JPS, Koch KT, et al. Incidence, Predictors, and Impact of Vascular Complications After Transfemoral Transcatheter Aortic Valve Implantation With the SAPIEN 3 Prosthesis. Am J Cardiol. 2018;121(10):1231-8.
17) Sedaghat A, Neumann N, Schahab N, Sinning JM, Hammerstingl C, Pingel S, et al. Routine Endovascular Treatment With a Stent Graft for Access-Site and Access-Related Vascular Injury in Transfemoral Transcatheter Aortic Valve Implantation. Circ Cardiovasc Interv. 2016;9(8).
18) Ullery BW, Jin R, Kirker EB, Hayes G, Siwek L, Brevig J, et al. Trends in vascular complications and associated treatment strategies following transfemoral transcatheter aortic valve replacement. J Vasc Surg. 2020;72(4):1313-24 e5.
19) Batchelor W, Patel K, Hurt J, Totten J, Burroughs P, Smith G, et al. Incidence, Prognosis and Predictors of Major Vascular Complications and Percutaneous Closure Device Failure Following Contemporary Percutaneous Transfemoral Transcatheter Aortic Valve Replacement. Cardiovasc Revasc Med. 2020;21(9):1065-73.
20) Genereux P, Head SJ, Van Mieghem NM, Kodali S, Kirtane AJ, Xu K, et al. Clinical outcomes after transcatheter aortic valve replacement using valve academic research consortium definitions: a weighted meta-analysis of 3,519 patients from 16 studies. J Am Coll Cardiol. 2012;59(25):2317-26.
21) Genereux P, Webb JG, Svensson LG, Kodali SK, Satler LF, Fearon WF, et al. Vascular complications after transcatheter aortic valve replacement: insights from the PARTNER (Placement of AoRTic TraNscathetER Valve) trial. J Am Coll Cardiol. 2012;60(12):1043-52.
22) Krasopoulos G, Falconieri F, Benedetto U, Newton J, Sayeed R, Kharbanda R, et al. European real world trans-catheter aortic valve implantation: systematic review and meta-analysis of European national registries. J Cardiothorac Surg. 2016;11(1):159.
23) Muralidharan A, Thiagarajan K, Van Ham R, Gleason TG, Mulukutla S, Schindler JT, et al. Meta-Analysis of Perioperative Stroke and Mortality in Transcatheter Aortic Valve Implantation. Am J Cardiol. 2016;118(7):1031-45.
24) Habertheuer A, Gleason TG, Kilic A, Schindler J, Kliner D, Bianco V, et al. Impact of Perioperative Stroke on Midterm Outcomes After Transcatheter Aortic Valve Replacement. Ann Thorac Surg. 2020;110(4):1294-301.
25) Mack M, Hamandi M. Why Surgical Risk Algorithms Are Not Predictive of Transcatheter Aortic Valve Replacement Outcomes! Circ Cardiovasc Interv. 2019;12(1):e007560.

26) Popma JJ, Adams DH, Reardon MJ, Yakubov SJ, Kleiman NS, Heimansohn D, et al. Transcatheter aortic valve replacement using a self-expanding bioprosthesis in patients with severe aortic stenosis at extreme risk for surgery. J Am Coll Cardiol. 2014;63(19):1972-81.
27) Netherlands Heart Registration: NHR; 2019 (Cited 2020 June 15). Available from: https://nederlandsehartregistratie.nl/wp-content/uploads/2020/01/NHR-Rapportage-2019-per-spread-230120.pdf.
28) Winter A, Brigl B, Buchauer A, Dujat C, Graber S, Hasselbring W, et al. Purpose and structure of strategic plans for information management in hospitals. Stud Health Technol Inform. 2000;77:880-4.
29) Brigl B, Ammenwerth E, Dujat C, Graber S, Grosse A, Haber A, et al. Preparing strategic information management plans for hospitals: a practical guideline SIM plans for hospitals: a guideline. Int J Med Inform. 2005;74(1):51-65.

Summary

Summary

Chapter 1 describes the objectives and problems addressed in this thesis. Specifically, despite the availability of TAVI data in the National Heart Registration (NHR), no studies have been conducted on how existing models predict mortality and morbidity in terms of validity and reliable clinical use in The Netherlands. In addition, it is unknown whether adjustments and regular updates should be pursued for these models or whether a redesigned model based on NHR data yields better results for the Dutch population. Finally, it is unknown what IT challenges heart centers to face, and how establishing their strategic information plans are aided by a given guideline for creating them.

In chapter 2, the performance of existing prediction models of mortality including the ACC-TAVI model and the FRANCE-2 model was assessed with TAVI data from NHR. The ability to discriminate between survivors and non-survivors by means of the AU-ROC metric was between 0.58 and 0.64. It was inferred that externally validated TAVI models showed an inadequate and suboptimal predictive performance on the external Dutch population dataset.

Chapter 3 showed that an updated revised version of the ACC-TAVI should be the better currently available prediction model for TAVI-related early-mortality in the Dutch population. The predictive ability of the updated ACC-TAVI model was still suboptimal. It was recommended that other countries consider model updates in their populations. Moreover, the study recommended the development of a new validated TAVI-specific prediction model with the use of NHR data.

In chapter 4, data of 9144 TAVI patients from the NHR were used to develop and internally validate a new TAVI-specific model. The final model (TAVI-NHR) has ten variables, including age (in years), serum creatinine, left ventricular ejection fraction, body surface area, NYHA class, procedure-acuteness status, chronic lung disease, critical-preoperative state, diabetes-mellitus, and TAVI access-routes. Body surface area and diabetes mellitus emerged as new predictors that were not used in the currently available TAVI-specific models. The AU-ROC of the TAVI-NHR model was 0.69 (IQR 0.646-0.75). There have been no signs of miscalibration observed.

In chapter 5, it was additionally found that in spite of the significant expansion of the performed TAVI procedures over the years, the mortality rates significantly dropped. The majority of serious complications still occur in elderly patients and patients at higher surgical risk.

Chapter 6 advocates the establishment of a strategic information management plan (SIM-plan) for organizing the necessary data management infrastructure in a heart center. Implementing a SIM plan for a heart center aided the identification of 15 business goals and 6 IT goals. Moreover, the activities enabled describing and assessing the current HIS situation and identifying the IT problems and issues, which could form priorities to be addressed. Maintenance of SIM-plans by means of regular updates due to increasing demands and changes in the hospital information system (HIS) infrastructure is necessary.

Chapter 7 provides a wrap-up of the findings and suggestions for the future are given. The most relevant recommendation is to consider how to maintain the NHR-TAVI model and to organize optimal data collection to meet ambitions in modeling and applications of Artificial Intelligence.

Samenvatting

Samenvatting

Hoofdstuk 1 beschrijft de doelstellingen en problemen die in dit proefschrift aan de orde komen. Ondanks de beschikbaarheid van TAVI-gegevens in de nationale Nederlandse Hart Registratie (NHR), waren er geen studies uitgevoerd naar predictiemodellen voor Nederlandse TAVI-patiënten. De vraag is hoe valide zijn de bestaande internationale predictiemodellen voor de mortaliteit en de morbiditeit van TAVI-patiënten in Nederland. Daarnaast is het onduidelijk of deze bestaande internationale predictiemodellen kunnen worden aangepast (geüpdatet), of dat een nieuw model op basis van TAVI-gegevens van de NHR-registratie betere resultaten zou opleveren voor de Nederlandse bevolking. De gegevens uit de NHR-registratie zijn afkomstig uit de verschillende hartcentra in Nederland. Het is echter niet bekend met welke IT-uitdagingen deze hartcentra worden geconfronteerd. Daarom hebben we deze uitdagingen geïnventariseerd en een strategisch informatieplan (SIM) voor een hartcentrum opgesteld aan de hand van een bestaande SIM-richtlijn.

In hoofdstuk 2 worden de prestaties beschreven van de bestaande 30 dagen sterfte predictiemodellen voor TAVI-patiënten, waaronder het ACC-TAVI-model en het FRANCE-2-model, die extern gevalideerd zijn met behulp van Nederlandse TAVI-gegevens van de NHR. Het vermogen van het model om onderscheid te maken tussen overleven en overlijden wordt gemeten door middel van de oppervlakte onder de curve (AUC). De AUC van de bestaande modellen lag tussen de 0,58 en 0,64. Er werd geconcludeerd dat de extern gevalideerde TAVI-modellen ACC-TAVI en het FRANCE-2 een niet toereikende en suboptimale voorspellende prestatie vertoonden op basis van de externe Nederlandse bevolkingsdataset.

Hoofdstuk 3 laat zien dat de geüpdatete versie van de ACC-TAVI een beter predictiemodel is voor TAVI-gerelateerde vroege (30-dagen) sterfte in Nederland in vergelijking met het bestaande model. Het voorspellend vermogen van het geüpdatete ACC-TAVI-model was echter qua AUC nog steeds laag. Andere landen wordt geadviseerd om te overwegen het ACC-TAVI-model ook voor hun eigen populaties te updaten, voordat die gebruikt kan worden. Bovendien adviseerde de studie de ontwikkeling van een nieuw gevalideerd TAVI-specifiek predictiemodel met behulp van NHR TAVI-gegevens.

In hoofdstuk 4 worden gegevens van 9144 TAVI-patiënten uit de NHR-registratie gebruikt om een nieuw TAVI-specifiek model te ontwikkelen en intern te valideren. Het nieuwe predictiemodel (TAVI-NHR) heeft tien predictie variabelen. Deze zijn: leeftijd (in jaren), serum kreatinine gehalte, linkerventrikelejectiefractie, lichaamsoppervlak (Body Surface Area, BSA), NYHA-klasse, urgentie van de procedure, chronische longziekte, kritieke preoperatieve toestand, diabetes mellitus, en TAVI-toegangsweg. Lichaamsoppervlak en diabetes mellitus kwamen naar voren als nieuwe voorspellers die niet werden gebruikt in de huidige beschikbare internationale TAVI-specifieke modellen. De AUC van het nieuwe TAVI-NHR-model is 0,69 (IQR 0,646-0,75) en er zijn geen tekenen van miskalibratie van de voorspellingen van sterfte waargenomen.

In hoofdstuk 5 werd, naast sterfte, ook naar andere morbiditeitsuitkomsten gekeken. Er werd gevonden dat ondanks de significante uitbreiding van de uitgevoerde TAVI-procedures in de loop der jaren, de sterftecijfers van TAVI-patiënten significant daalden. De meeste ernstige complicaties zoals pacemakers, stroke, ernstige bloedingen treden nog steeds op bij oudere patiënten, en bij patiënten met een hoger chirurgisch risico.

Hoofdstuk 6 pleit voor het opstellen van een strategisch informatiemanagement plan (SIM-plan) voor het inrichten van de benodigde databeheerinfrastructuur in een hartcentrum. Het implementeren van een SIM-plan voor een hartcentrum helpt bij het identificeren van business doelen en IT-doelen. In dit onderzoek hebben we aan de hand van een bestaande richtlijn een SIM-plan opgesteld en er zijn 16 business doelen en 6 IT-doelen gevonden. Bovendien maken de analyseactiviteiten het mogelijk om de huidige situatie van het ziekenhuisinformation systeem (ZIS) te beschrijven en te beoordelen. Een SIM-plan geeft de mogelijkheid om IT-problemen te identificeren en om prioriteiten te stellen om deze aan te pakken.

Hoofdstuk 7 vat de bevindingen van dit proefschrift samen en doet aanbevelingen voor de toekomst. De meest relevante aanbevelingen zijn om na te denken over het implementeren en onderhouden van het nieuwe NHR-TAVI-predictiemodel en het organiseren en verder optimaliseren van de dataverzameling in de hartcentra om te kunnen voldoen aan de ambities in modellering van predictiemodellen en de toepassingen van kunstmatige intelligentie hierin.


CV

Curriculum vitae and Portfolio

Curriculum vitae

Hatem Al-Farra was born in Homs city, Syria, in 1978 and grew up in various countries in the Middle East. He completed his medical studies at the University of Science and Technology in Sudan, where he spent seven years. Subsequently, he pursued a higher diploma for hospital management and administration under the aegis of the Sudan medical board, which he completed in 2007.

With eight years of professional experience as a medical officer and registrar of internal medicine, orthopedics, and general surgery in the Middle East, Hatem relocated to the Netherlands in 2013. Following the completion of the pre-master program in 2015, he joined the Master's program in Medical Informatics at the Academic Medical Center of the University of Amsterdam and earned his degree in 2017. During his master's studies, he wrote his thesis entitled "Development of a Strategic Information Management Plan in the AMC Heart Center: SIM-plan in the creation phase - lessons learned."

In November 2016, Hatem commenced his doctoral research at the department of Medical Informatics and the Heart Center of the Academic Medical Center - University of Amsterdam. Between 2016 and 2022, he conducted research as described in this thesis under the guidance of Prof. Dr. Ameen Abu-Hanna, Prof. Dr. Bas A.J.M. de Mol, Prof. Dr. José P.S. Henriques, and Dr. Anita C.J. Ravelli.

In addition to his research, Hatem has been involved in mentoring bachelor's students in the Medical Informatics department on strategic information planning. He has also taken on various responsibilities at the Heart Center, such as enhancing medical data registration and extraction and devising a plan to establish a data warehouse for the Heart Center.

Hatem obtained his medical license in the Netherlands in 2019 and has since been practicing as a general physician, in the fields of psychiatry and elderly care. Aside from his clinical work, he volunteers as a mentor and trainer for foreign doctors who recently arrived in the Netherlands. Additionally, he is involved in community service activities that raise awareness about COVID-19 infection and vaccination programs.

After finishing his Ph.D., Hatem will continue working in elderly care and contribute to the ongoing scientific research in this field.

Upon completion of his Ph.D., Hatem will continue to contribute to ongoing scientific research in the field of elderly care while working in this area.

Portfolio

Name Ph.D. Student:	Hatem Al-Farra
Ph.D. Period:	November 2016 to November 2022
Promoters:	Ameen Abu-Hanna and Bas A.J.M. de Mol
Co-promoters:	José P.S. Henriques and Anita C.J. Ravelli

Ph.D. Training and courses

Training and courses	Year	Workload (EC)
The AMC World of Science	2017	0.70
Basic Course Legislation and Organization - eBROK	2017	0.80
Practical Biostatistics (e-learning)	2017	1.10
Advanced Topics in Biostatistics	2018	2.10
Computing in R - online course	2018	3.00
Tutorial prediction models	2018	0.50
Data Analysis and Visualization– Online course - Udacity	2018	0.30
R, ggplot, and Simple Linear Regression- Udemy	2018	0.40
Communication with patients, for MDs	2018	1.50
Clinical examination: training in Dutch	2018	2.10
Computing in R, the AMC Graduate School	2019	4.00
R Programming online course, "Coursera"	2019	2.00
Cardiovascular Diseases – Webinar, Mayo Clinic	2019	0.30
Special Topics in Data Science in Medicine	2019	2.10
Learn R, DataCamp online course	2019	1.10
Amsterdam Public Health (APH): Methodology tutorial, Missing Data: consequences and solutions	2019	1.00
Amsterdam cardiovascular sciences	2019	1.50
Psychiatry: common diseases in elderly people "GGZ-Leiden"	2019	2.20
Data Science: Statistics and Machine Learning Specialization: Regression Models; Coursera	2019	1.50
Project Management: online course	2020	2.00

Conferences, workshops, and activities	Year	Workload (EC)
Zorg & ICT Conferences - Utrecht	2017, 2018, and 2019	1.50
Value-Based Health Care: Michael Porter presentation, Utrecht	2017	0.50
The Medical Informatics Conference - MIC	2017	2.00
Scientific Talk(s) Heart Center AMC (Every 3 months)	2018, and 2019	0.50
Journal club data science	2018, and 2019	0.90
Teaching SIM-plan bachelor students (BAM 3.5)	2018	0.20
APH meetings: Annual and Spring Meetings, Amsterdam	2019	0.70
Teaching SIM-plan bachelor students (BAM 3.5)	2019	0.20
Ph.D. Day 2019: A future in biostatistics	2019	0.70
Antibiotics: safety and allergic reactions "Online training and workshop" - LUMC - Leiden	2021	0.50
ECG: basic and advice training, LUMC – Leiden	2021	4.50
BIG5 Congress of Elderly Medicine, Ede	2021, and 2022	8.00

Acknowledgement

Acknowledgement

Undertaking a Ph.D. is no easy feat, and the successful completion of this thesis would not have been possible without the support and contributions of many. I am incredibly grateful for the invaluable help and support extended to me throughout this journey, and I would like to take this opportunity to express my heartfelt gratitude to everyone who played a role in my success.

First and foremost, I would like to extend my gratitude to all the participants in the studies mentioned in this thesis. I was allowed to use data from 16 heart centers registered in the NHR. Without their cooperation, there would have been no data to analyze, and my research wouldn't have been possible.

I would also like to express my sincere gratitude to the members of the reading committee, Prof. dr. H.A. Marquering, Prof. dr. F.W. Asselbergs, Prof. dr. Martijn Schut, Prof. dr. M.M. Vis, and dr. Giovanni Cina, for their time and efforts in critically reading and assessing my dissertation.

My sincere thanks to my Ph.D. advisory team, Prof. dr. A. Abu-Hanna, Prof. dr. Bas A.J.M. de Mol, Prof. dr. José P.S. Henriques, and dr. Anita C.J. Ravelli, for their invaluable guidance, support, and inspiration. I am grateful for their analytical thinking, expertise, and knowledge in the field of research, which helped me to complete this project successfully.

A. Abu-Hanna thank you for your support, critical thinking and all your knowledge and involvement in research in the field of artificial intelligence and prediction models. Bas A.J.M. de Mol, thank you for inspiring me with innovative ideas that initially puzzled me but eventually proved to be invaluable. You revived my passion and revitalized my stagnant medical practice. I am also extremely grateful to Prof. dr. José P.S. Henriques and dr. Anita C.J. Ravelli for their continuous guidance and unwavering support throughout my Ph.D. trajectory, always being available to assist me in various aspects of my research.

I would also like to express my appreciation to the Ph.D. advisory team for their support during difficult times and for encouraging me to prioritize my well-being.

I extend my gratitude to Mr. W.J.P.P. ter Burg for his enthusiasm for strategic information management, and his contributions to my scientific research, which eventually led to the start of my Ph.D. trajectory.

I would also like to extend my gratitude to Dr. Stephanie K. Medlock, who was instrumental in helping me during my Master's degree. Her guidance and mentorship provided me with a strong foundation and helped me develop the skills necessary to pursue my Ph.D. Without her support, I would not have been able to make the progress that I did, and I am grateful for all that she has done for me.

My sincere thanks also go to Mr. Noman Dormosh for his expertise, time, and invaluable support.

Lastly, I am deeply grateful to my family, particularly my parents, and my beloved wife Dr. Wesam Alloh and children Asser, Lour and Naya, for their constant love and support throughout my journey. Their unwavering encouragement and motivation have been pivotal in my achievements. Additionally, I extend my heartfelt thanks to my friends, particularly Eng. A. Elkahlout and Dr. Rasheed Elberir, for providing a much-needed break from the rigors of my research work. It was a great pleasure to spend time with you engaging in diverse activities, such as singing, chatting, and helping me renovate my new home.

Once again, thank you to all those who have contributed to this journey.