

Bioinformatics Program
Technical University of Munich
Ludwig-Maximilians-Universität München

Bachelor's Thesis in Bioinformatics

Developing a One-Year Risk Score for Transcatheter Aortic Valve Implantation

Judith Y. Bennett

Bioinformatics Program
Technical University of Munich
Ludwig-Maximilians-Universität München

Bachelor's Thesis in Bioinformatics

Developing a One-Year Risk Score for Transcatheter Aortic Valve Implantation

Entwicklung eines Scores zur einjährigen Risikoprognose für Transkatheter-Aortenklappenimplantation

Author: Judith Y. Bennett
Supervisor: Prof. Dr. Jan Baumbach
Advisor: Dr. Tim Kacprowski,
Chair of Experimental Bioinformatics
Maximus-von-Imhof-Forum 3, 85354 Freising
Submitted: 15.03.2020

Bachelor's Thesis Statement of Originality

I confirm that this bachelor's thesis is my own work and I have documented all sources and material used.

Place and Date

Signature

Abstract

As personalized medicine gains more and more attention, the need for the development of scores defining and weighting risk parameters is steadily growing. Currently, there is no standard risk assessment score for mortality prediction after transcatheter aortic valve implantation (TAVI). Although several scores have been presented for this purpose, they are still not reliable and accurate enough, especially for intermediate and low surgery risk patients.

In the scope of this thesis, a score for one-year all-cause mortality after TAVI was developed using only pre-procedural variables from the Swiss-TAVI registry. It reached a good concordance of 0.741 and was robust in a 10-fold cross-validation, outperforming all but one one-year TAVI-specific mortality score. The score successfully stratified the patients into high, intermediate and low hazard groups. Eleven biomarkers were identified to have an adverse effect on the patient's survival: elevated age, male gender, low hemoglobin levels, low mean transvalvular pressure gradient, mitral regurgitation of grade 3, atrioventricular block, peripheral artery disease, chronic obstructive pulmonary disease, diuretic medication and the absence of statin medication or Aspirin as anticoagulation drug.

The two major points setting this score apart from published scores are that the variables were selected with LASSO regression, a fully computational and mathematically meaningful approach, and that it is applicable to intermediate and low surgery risk patients. In the future, the score may be improved by the addition of frailty- and calcification-related parameters.

Zusammenfassung

Im Rahmen der Bedeutungszunahme personalisierter Medizin wächst das Bedürfnis nach Scores, die Risikoparameter definieren und gewichten. Zum heutigen Zeitpunkt existiert kein Standardscore zur Abschätzung des Mortalitätsrisikos nach Transkatheter-Aortenklappenimplantationen (TAVI). Obwohl mehrere Scores zu diesem Zweck bereits publiziert wurden, sind deren Verlässlichkeit und Genauigkeit verbesserungswürdig, insbesondere für Patientengruppen mit mittlerem bis niedrigem Operationsrisiko.

In der vorliegenden Arbeit wurde ein Score zur einjährigen Risikoprognose bezogen auf die Gesamt mortalität entwickelt. Hierfür wurden ausschließlich Variablen aus dem Swiss-TAVI Register verwendet, die vor der Implantation festgehalten wurden. Der Score erreichte eine gute Konkordanz (0.741), war robust in einer 10-fachen Kreuzvalidierung und übertraf die Ergebnisse fast aller publizierten TAVI-Scores hinsichtlich einjährigem Mortalitätsrisiko. Die Patienten des Registers konnten erfolgreich in Gruppen mit hohem, mittlerem und niedrigem Risiko eingestuft werden. Elf Biomarker, die das Überleben des Patienten negativ beeinflussen, konnten identifiziert werden: hohes Alter, männliches Geschlecht, niedrige Hämoglobinwerte, niedriger mittlerer transvalvulärer Druckgradient, Mitralklappeninsuffizienz Grad 3, atrioventrikulärer Block, periphere arterielle Verschlusskrankheit, chronische obstruktive Lungenerkrankung und Medikamentation mit Diuretika. Statinmedikamentation und der Einsatz von Aspirin als Antikoagulans hatten hingegen positiven Einfluss.

Die zwei wesentlichen Vorteile, die diesen Score von anderen unterscheiden, bestehen darin, dass die Variablen mit LASSO Regression, einem vollständig automatisierten und mathematisch wohlfundierten Ansatz, ausgewählt wurden und dass der Score für Patienten mit mittlerem bis niedrigem Operationsrisiko anwendbar ist. Zukünftig könnte der Score durch die zusätzliche Berücksichtigung von Kalzifikationsparametern und Variablen, die sich auf Altersgebrechlichkeit beziehen, verbessert werden.

Acknowledgment

First of all, I would like to thank Prof. Dr. Jan Baumbach for the chance to write my thesis at his extraordinary chair. I enjoyed the time working on this project very much and learned how promising bioinformatics projects can be when so many people from different fields and places come together.

In particular, I want to thank Dr. Tim Kacprowski, my advisor, for his excellent advice and insights. In spite of his busy schedule, he always found the time to answer my questions and provide me with new ideas. He also established the contact to Dr. med. Philipp Jakob, whom I have to thank for the data and the insightful discussions in many Skype meetings. I felt honored for being allowed to work on this project.

Furthermore, I am grateful for becoming a part of Dr. Tim Kacprowskis Computational Systems Medicine research group. I want to thank the whole group and my fellow students for their input in our group meetings. I am very much looking forward to working with these people on further, interesting topics.

Moreover, I thank my friends and family for their patience and their motivation. Thank you to Julian, who was always there for me, helped me to believe in myself and never gave up to convince me of the beauty of mathematics. Last but not least, nothing of all this would have been possible without my mother and my grandmother. I thank them for their immense support and love. They gave me the strength to work my way through all my studies and never stopped motivating me to do good and thorough work.

Contents

List of Tables	iii
List of Figures	iii
List of Abbreviations	v
1 Introduction	1
1.1 Motivation	1
1.2 Overview	2
1.3 Outline	2
2 Background	5
2.1 Medical Background	5
2.1.1 TAVI procedure and indications	6
2.1.2 Biomarkers for TAVI	8
2.2 Computational and statistical methods	10
2.2.1 Regression background	10
2.2.2 Random Survival Forests	16
2.3 Related Work	18
2.3.1 Insufficiency of the STS Score and the EuroSCOREs for TAVI mortality prediction	18
2.3.2 Published Scores for TAVI mortality prediction	19
3 Material and Methods	25
3.1 Swiss-TAVI cohort	25
3.2 Statistical analyses	28
3.3 Model inference and statistical testing	28
4 Results and Discussion	31
4.1 Baseline characteristics	31
4.2 Scores	39
4.2.1 Comparison of the scores	42
4.2.2 LASSO-on-all: Smaller Model	45
4.3 Limitations	54
5 Conclusion	57
5.1 Summary and Conclusion	57
5.2 Outlook	58
Bibliography	59
Appendix	i
Supplementary Data	i

List of Tables

2.1	Aspects for decision between SAVR and TAVI	7
2.2	Overview of the performance of the STS score, the logistic EuroSCORE and the EuroSCORE II for TAVI mortality prediction.	18
2.3	Overview of the published scores for TAVI	20
2.4	Summary of the findings of SAVR versus TAVI mortality studies	24
4.1	Baseline characteristics of the Swiss-TAVI database stratified by event .	33
4.2	Overview of the developed Cox models	40
4.3	Cox model for the ‘LASSO-on-all: Smaller Model’	48
5.1	Cox model for the ‘Benjamini-Hochberg Model’	v
5.2	Cox model for the ‘LASSO-on-all Model’	vi
5.3	Cox model for the ‘LASSO-on-intermediate/ low risk Model’	vii
5.4	Cox model for the ‘Random Forest Model’	vii

List of Figures

2.1	Anatomy of the heart	5
2.2	Visualization of a linear regression	11
2.3	Illustration of the idea behind random survival forests	17
3.1	Distribution of the implanted TAVI devices	26
3.2	Flowchart of the patient and baseline variable preselection progress . .	27
4.1	Overview of the event distribution	35
4.2	Yearly distribution of the significant categorical variables	36
4.3	Random Forest: VIMP versus Minimal Depth	41
4.4	Comparison of the concordances of the 100 permutation tests and the real C-index/cross-validated iAUCs	43
4.5	Pearson correlation coefficients for the medication variables included in the Cox models	44
4.6	Kaplan-Meier curves ‘LASSO-on-all: Smaller Model’	46

4.7	Pearson's correlation coefficients for the 'LASSO-on-all: Smaller Model'	47
5.1	Percentage of devices implanted, stratified by event	i
5.2	Density plots for the follow-up data	ii
5.3	Violin Plots for the distribution of the continuous variables, stratified by event	ii
5.4	Distribution of the continuous variables over the years, stratified by event	iii
5.5	Distribution of the boolean-coded variables, stratified by event	iv
5.6	Out of Bag error rate for the number of trees.	viii
5.7	Distribution of the centered linear predictors for the 'LASSO-on-all: Smaller model'.	viii
5.8	Kaplan-Meier curves STS score	ix
5.9	Kaplan-Meier curves Euro II score	ix

List of Abbreviations

ΔPm	mean transvalvular pressure gradient
AKI	acute kidney injury
AS	aortic stenosis
AUC	area under the ROC curve
AVA	aortic valve area
BAV	balloon aortic valvuloplasty
BMI	body mass index
BSA	body surface area
CCS	Canadian Cardiovascular Society
CI	confidence interval
COPD	chronic obstructive pulmonary disease
CT	computed tomography
EuroSCORE	European System for Cardiac Operative Risk Evaluation
GFR	glomerular filtration rate
Hb	hemoglobin
LASSO	Least Absolute Shrinkage and Selection Operator
LBBB	left bundle branch block
LV	left ventricle
LVEF	left ventricular ejection fraction
MI	myocardial infarction
MRI	magnetic resonance imaging
NYHA	New York Heart Association
PCI	percutaneous coronary intervention
RBBB	right bundle branch block
SAVR	Surgical Aortic Valve Replacement
STS	Society of Thoracic Surgeons
SVi	stroke volume index
TAVI	Transcatheter Aortic Valve Implantation
TEE	transesophageal echocardiography
TTE	transthoracic echocardiography

1 Introduction

Bioinformatics is an emerging, promising field applying computational and mathematical approaches to the steadily growing masses of biological data. Especially in the field of human medicine, it is no longer feasible to conduct quantitative research without statistical programs or other computational ways to process all the data generated in large-scale studies. The idea of personalized medicine aims to treat every patient individually, tailoring therapies that match the patient perfectly in order to give him the best recovery chances possible. This requires the development of models and scores defining and weighting risk parameters which can then be individually assessed, thus leading to an individual hazard evaluation. The Cox proportional hazard model is a regression method which is often used for these kind of purposes. The one-year risk score proposed in this thesis was also calculated with this method.

1.1 Motivation

Patients with severe aortic stenosis either have to undergo Surgical Aortic Valve Replacement (SAVR), transcatheter aortic valve implantation (TAVI) or drug therapy. Various studies have shown drug treatment to be unfavorable, the median survival for this patient group was merely one year [1]. For this reason, trials for applying Transcatheter Aortic Valve Implantation (TAVI) in patients uneligible for a surgical intervention were run [2, 3]. Since these studies reached satisfactory results, demonstrating that TAVI significantly prolongs the subjects' life, TAVI became a standard method for prohibitive, extreme and high surgery risk patients. Similarly, it became a considerable option for intermediate and low surgery risk patients because it is performed minimally invasive and mostly in local anesthesia putting the patient under less physical stress.

There are commonly accepted scores for operative risk evaluation like the European System for Cardiac Operative Risk Evaluation (EuroSCORE) II [4] or the Society of Thoracic Surgeons (STS) Score [5–7] but there is no standard score for TAVI. Currently, the surgery risk scores are just used for deciding between SAVR and TAVI because they perform well for SAVR mortality prediction [8–10] but poor for TAVI [11–18].

Most published TAVI risk evaluation scores are not very accurate and can therefore only support treatment choices instead of reliably predicting outcomes [19–26]. Additionally, the derivation cohorts for the existing scores mainly consist of high and extreme surgery risk patients, making it questionable whether or not they can be extrapolated to intermediate and low surgery risk patients. Another shared weakness of

most TAVI risk scores is that the predictors were chosen on the basis of expert opinion rather than selecting them in an unbiased statistical and computational approach.

As machine learning techniques are currently attracting more and more notice and are becoming better every year, a fully computational approach for risk score development seems to be the way to go. The Swiss-TAVI cohort, a multi-center registry for TAVI patients in Switzerland, is sufficiently large so that a combination of bioinformatical methods and medical knowledge can unveil strong biological prediction features. An additional advantage lies within the fact that Swiss-TAVI mostly consists of intermediate and low surgery risk patients, setting it apart from existing studies.

1.2 Overview

For this thesis, a one-year mortality risk score for TAVI was developed, consisting only of pre-proceduraly recorded biomarkers of the Swiss-TAVI registry. It aims to reliably estimate the patient's individual risk in order to raise his awareness for potential complications so that one-year mortality can be prevented.

The presented score differs from existing TAVI risk scores by two major points: Firstly, it was constructed using a fully computational and mathematically meaningful approach for variable selection instead of relying on personal opinion. Secondly, the Swiss-TAVI cohort mainly consists of intermediate and low surgery risk patients, making the score applicable to this patient group.

The score uses 11 variables for mortality prediction, balancing the risk of a too simple and an overfitted model. The features were selected by testing different computational approaches including simple statistical testing, Least Absolute Shrinkage and Selection Operator (LASSO) regression and random survival forests. All models were compared and validated by various statistical measures, including the c-index, the Wald test, 10-fold cross-validation and permutation tests.

The final score is composed of a subset of the LASSO-selected variables which was obtained by applying further optimization to the set of features. Kaplan-Meier survival curves showed that the score successfully stratifies the patients into high, intermediate and low hazard groups, outperforming the STS and EuroSCORE II as well as all but one published TAVI one-year mortality score.

1.3 Outline

Chapter 2 is subdivided into three parts, covering the medical background, the computational and statistical methods and the related work on the field of TAVI risk scores. Section 2.1 explains the TAVI procedure and describes potential biomarkers. Section 2.2 addresses several regression methods and their statistical quality measures. The chapter also describes random survival forests. Section 2.3 first explains why the STS score and the EuroSCORE II are insufficient for TAVI mortality prediction and then presents and discusses published scores and mortality studies for TAVI.

Chapter 3, Material and Methods, starts by describing the Swiss-TAVI registry in general and which part of the data was used for this thesis. Chapters 3.2 and 3.3

portray how the data was examined with simple statistical tests, how the models were developed by different approaches and how they were compared by statistical methods. In Chapter 4, ‘Results and Discussion’, the baseline characteristics of the Swiss-TAVI cohort are examined first. In Chapter 4.2, the scores resulting from the different variable selection approaches (simple statistical testing, LASSO regression, random survival forest) are depicted and debated. Chapter 4.2.2 discusses the final score presented as the main result of this thesis. It assesses the model quality in terms of discrimination of patients, internal robustness, redundancy, and medical research. Lastly, the limitations of this thesis and its presented mortality score are discussed.

Finally, the work of this thesis is summarized in Chapter 5. Conclusions are drawn from the results of the analyses and an outlook for future work on this topic is given.

2 Background

The following section guides through the required medical and computational background for this work. It explains the clinical situation of patients undergoing TAVI and the statistical methods to examine the Swiss-TAVI database for the creation of a mortality risk score. The chapter also includes published work in the field of TAVI related mortality studies and thus leads over the scope of the thesis.

2.1 Medical Background

The human heart possesses four valves, two separating the atria from the ventricles and two connecting the aorta respectively the pulmonary trunk with the ventricles. The *vena cava inferior* and the *vena cava superior* transport the oxygen-poor blood from the body into the right atrium of the heart. From there, it passes the tricuspid valve (*valvula tricuspidalis, valvula atrioventricularis dextra*) to get into the right ventricle and finally via the pulmonary valve (*valva trunci pulmonalis*) into the pulmonary trunk and into the lungs where it becomes re-saturated in oxygen. The oxygenated blood flows back into the left atrium of the heart through the *venae pulmonales*. It leaves the atrium via the mitral valve (*valva atrioventricularis sinistra, valva mitralis, valvula bicuspidalis*) where it reaches the left ventricle. The blood is released into the body through the aortic valve (*valva aortae*) and the aorta [27] (see Figure 2.1).

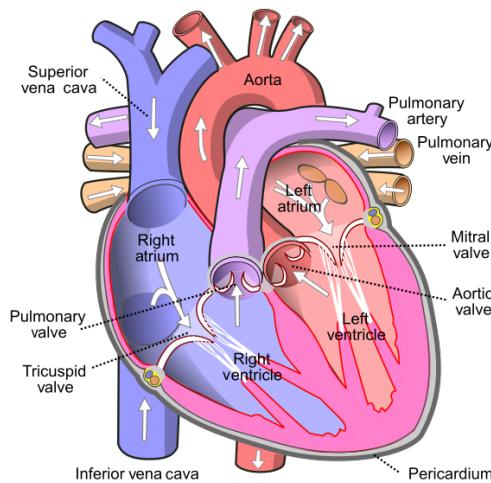


Figure 2.1: Anatomy of the heart. The blood flow is illustrated with the aid of white arrows. From Wikimedia Commons.

All of these valves can suffer from a stenosis, i.e. narrowing of their opening, or regurgitation, meaning an incomplete closure of the valve [28]. Although it is sometimes sufficient to widen the valve in case of a stenosis (valvuloplasty, referred to as balloon aortic valvuloplasty (BAV)) [1, 28], a replacement of the diseased valve with an artificial one is often required.

2.1.1 TAVI procedure and indications

The latest ESC/EACTS guidelines for the management of valvular heart disease [8] recommend an intervention in aortic stenosis (AS) patients if there is

- a severe, symptomatic AS with
 - a) a high mean transvalvular pressure gradient (ΔP_m) (≥ 40 mmHg)
 - b) a low ΔP_m , an aortic valve area (AVA) ≤ 1 cm 2 and a low flow (stroke volume index (SVi)) ≤ 35 mL/m 2)
- a severe, asymptomatic AS with a AVA ≤ 1 cm 2 and
 - a) a poor left ventricular ejection fraction (LVEF) ($< 50\%$)
 - b) normal LVEF and exercise test abnormalities like decrease of blood pressure
 - c) normal LVEF and risk factors like “peak velocity > 5.5 m/s; severe valve calcification + peak velocity progression ≥ 0.3 m/s per year; markedly elevated neurohormones ($>$ threefold age- and sex-corrected normal range) without other explanation; severe pulmonary hypertension (systolic pulmonary artery pressure > 60 mmHg)” [8]

The mean transvalvular pressure gradient is calculated from the velocity of the blood passing the aortic valve. An aortic stenosis can decelerate the blood flow and therefore build up a high ΔP_m [1, 27]. The SVi is defined as the stroke volume (volume of blood ejected from the left ventricle within one contraction) in relation to the body surface [28]. The LVEF describes the amount of blood that is ejected from the left ventricle in one heart beat, i.e. the fraction that the stroke volume makes up of the blood amount inside the left ventricle [28]. The systolic pulmonary artery pressure describes the blood pressure in the pulmonary artery [28].

Depending on various clinical characteristics, anatomical and technical aspects and cardiac conditions the guidelines either advise the multidisciplinary Heart Team of the hospital to do a surgical aortic valve replacement (SAVR) or a TAVI as intervention (see Table 2.1). In particular, TAVI is recommended for patients who are unsuitable for surgery or have an extreme or high surgery risk [1, 8, 28]. The eligibility of intermediate or low surgery risk patients is being examined [1, 29–33]. The Swiss-TAVI trial also contains an unusually high share of intermediate or low surgery risk patients.

In TAVI, the aortic valve is replaced percutaneously via a catheter. The most common access routes are transfemoral and transapical, with transfemoral being the preferred access route because it has been associated with a lower mortality rate [1]. While the femoral procedure can usually be done with local anesthesia, the transapical one normally requires general anesthesia. [28]. Other possible access routes include the subclavian or direct aortic approach.

	Favours TAVI	Favours SAVR
Clinical characteristics		
STS/EuroScore II <4%		+
Presence of severe comorbidity	+	
Age \geq 75 years	+	
Previous cardiac surgery	+	
Frailty	+	
Restricted mobility and conditions that may affect the rehabilitation process after procedure	+	
Suspicion of endocarditis		+
Anatomical and technical aspects		
Favourable access for transfemoral TAVI	+	
Sequelae of chest radiation	+	
Porcelain aorta	+	
Presence of intact coronary bypass grafts at risk when sternotomy is performed	+	
Expected patient-prosthesis mismatch	+	
Severe chest deformation of scoliosis	+	
Short distance between coronary ostia and aortic valve annulus		+
Size of aortic valve annulus out of range for TAVI		+
Aortic root morphology unfavourable for TAVI		+
Valve morphology (bicuspid, degree of calcification, calcification pattern) unfavourable for TAVI		+
Presence of thrombi in aorta or LV		+
Other cardiac conditions that could require a concomitant intervention		
Severe coronary artery disease requiring revascularization		+
Severe primary mitral valve disease, which could be treated surgically		+
Severe tricuspid valve disease		+
Aneurysm of the ascending aorta		+
Septal hypertrophy requiring myectomy		+

Table 2.1: Aspects to be considered by the Heart Team for the decision between SAVR and TAVI in patients at increased surgical risk. Adapted from [8].

The artificial aortic valve can either be self-expandable or balloon-expandable and is integrated in a stent (endoprosthesis with a lattice structure) at the end of the catheter [28]. The balloon technique widens the stenosis first by inflating a balloon which is located at the end of another catheter (balloon catheter) like in a BAV [28, 34]. The catheter with the stent and the aortic valve is inserted afterwards [34]. The aortic valve replacement comes in different sizes, depending on the patients heart. The composition of the artificial valve varies from producer to producer but can basically be divided into fully mechanical valves and valves with biological parts (mostly bovine or porcine) [28]. In preparation for the TAVI procedure, the imaging methods transthoracic echocardiography (TTE) and transesophageal echocardiography (TEE) are a clinical standard. In order to assess the heart anatomy even better, a computed tomography (CT) or magnetic resonance imaging (MRI) can additionally be performed. Thus, the proper valve size and potential access difficulties can be determined [34].

It is possible to have concomitant procedures in conjunction with TAVI like a balloon valvuloplasty or a percutaneous coronary intervention (PCI), a method for revascularization, i.e. reconstructing a regular blood flow.

2.1.2 Biomarkers for TAVI

The WHO International Programme on Chemical Safety Biomarkers in Risk Assessment defines biomarkers as “any substance, structure or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease” [35]. It has been shown that “[t]he functional status, comorbidities, and age of each individual patient [...] significantly affect outcomes [in patients undergoing SAVR]” [1]. Consequently, it seems likely that this also holds true for TAVI which has been explored by various papers [20–26, 36–42]. As the challenge of this thesis is to infer a mortality risk prediction score based on biomarkers recorded in the Swiss-TAVI database, it is essential to first explain some of these biomarkers in a medical context. The biomarkers can be divided into heart-related comorbidities, other comorbidities and blood levels. I also include medical scoring systems that are themselves derived from biomarkers into the biomarker definition. For some biomarkers, I include medical therapy possibilities because the Swiss-TAVI database also records numerous drugs.

Heart-related comorbidities

Apart from stenosis or regurgitation of the different heart valves mentioned above, there are several other noteworthy, heart-related diseases.

First, there are different types of rhythm disturbances, meaning deviances from the normal sinus rhythm. A commonly known case is atrial fibrillation which can be i.a. treated with anticoagulants (Marcoumar/Sintrom or novel anticoagulants like Rivaroxaban), beta blockers and calcium (Ca)-antagonists. Another special case of rhythm disturbance is atrioventricular block, which presents itself (contrary to atrial fibrillation) in a deceleration of the heart beat. In some cases, this can require a pacemaker implantation [28].

Secondly, there can be rather muscular problems, like an intraventricular conduction delay. Here, the ventricles contract asynchronously from each other, i.e. the left ventricle before the right one or vice versa (left bundle branch block (LBBB) resp. right bundle branch block (RBBB)) [28].

Thirdly, not only the valves but also the coronary arteries can experience stenosis. Together with other disorders, this is summarized as coronary artery disease. Medical therapies include Aspirin, beta blockers, Ca-Antagonists and Statin [28].

These disorders can, amongst other things, lead to a myocardial infarction (MI) or a syncope. The latter term describes sudden unconsciousness that can also be triggered by various other circumstances including aortic stenosis [1, 28].

Other comorbidities

If not the coronary arteries but the aorta or the arteries that support the extremities with blood are affected by a stenosis or an occlusion, the disease pattern is defined as peripheral artery disease. Medical treatments include Clopidogrel and P2Y12-antagonists [28].

Another comorbidity with potential negative effect on the heart is a pathological elevation of the blood pressure, called arterial hypertension. ACE inhibitors, ATII-antagonist, beta blockers, Ca-Antagonists and diuretics may counteract arterial hyper-

tension [28]. Diuretics are also indicated when a patient suffers from renal failure or heart failure [28].

Further important comorbidities in TAVI patients include diabetes mellitus (often treated with antidiabetical medication and insulin), chronic obstructive pulmonary disease (COPD) (sometimes treated with steroids) and dyslipidemia, a fat metabolism disorder that can be treated with statin [28].

Lastly, cerebrovascular accidents like strokes should also be considered in TAVI patients.

Blood Levels

Heart-related blood biomarkers include troponin and the creatinine kinase. The former can assume two types (type T or type I) and its measurement can be used to diagnose a MI early [28]. Elevated levels of the creatinine kinase can indicate a cerebrovascular accident, a MI or renal insufficiency [27, 28].

Since chronic renal disease has a severe influence on the probability for heart failure [43] and is therefore important for the outcome of TAVI patients, makers for renal function have to be measured. The most important markers are creatinine and the herefrom derived glomerular filtration rate (GFR). Elevated levels of creatinine or a reduced GFR indicate renal dysfunction [44].

Other standard blood levels include hemoglobin (Hb), the red blood cells, and thrombocytes which are important for blood coagulation. Low hemoglobin levels have been associated with increased mortality after TAVI [45]. Bad blood coagulation is associated with a higher bleeding risk [27] but quite common in the population since a lot of patients take anticoagulants. Albumin, the main blood plasma protein, is also gaining more attention as a biomarker. A decreased concentration is often observed when the patient experiences liver failure, critical illness, sepsis and infections but also hepatitis, diabetes, cancer and other chronic diseases [46].

Medical Scoring Systems

Two commonly used medical scoring systems in cardiology are the Canadian Cardiovascular Society (CCS) Functional Classification of Angina and the Dyspnea New York Heart Association (NYHA) Functional Classification. The former is a classification system for angina pectoris (symptom of coronary insufficiency and AS, characterized by sudden pain in the thorax region) [1, 28]. There are four CCS classes I-IV with IV being the highest one. Ca-Antagonists can also be utilised for angina pectoris treatment [28]. The NYHA classification system ranges from I (no discomfort or dyspnea) to IV (discomfort and dyspnea even at rest) for dyspnea [28].

As mentioned in the introduction, there exist two frequently used scores to assess the operative mortality risk for 30 days, the STS score and its European counterpart, the so-called logistic EuroSCORE and its improved version, the EuroSCORE II. A STS Score ranging from 0 to 4 is associated with low mortality risk, whereas a score over 12 is associated with high risk. The logistic EuroSCORE values tend to be much higher, the EuroSCORE II values only slightly higher than the STS values. Additionally, the EuroSCORE uses significantly less variables than the STS Score (20 versus over 100).

2.2 Computational and statistical methods

This section provides a deeper understanding of the computational and statistical methods used in this thesis. More precisely, it leads over to the definition of Cox regression, variable selection via LASSO regression and random survival forests. Since Cox regression is based on logistic regression, which is a variant of linear regression itself, the chapter containing the regression background explains these two methods first. Secondly, LASSO regression is defined in the context of regularized regression, meaning LASSO regression as well as ridge and elastic net regression. Finally, the basic principles of random forests and its variant used in this thesis, random survival forests, are explained.

2.2.1 Regression background

In general, regression methods describe the relationship between a dependent (or response) variable and one or more independent (or predictor) variables. In this case, the predictor variables consist of the biomarkers denoted in the Swiss-TAVI database. The response variable is the variable that one wants to predict, in our case all-cause mortality within one year. The regression method chosen for this task is Cox regression. Cox regression is based on logistic regression which is itself based on linear regression. The predictor variables were chosen with the help of LASSO regression, a form of regularized regression.

The main difference between linear regression, logistic regression and Cox regression is the nature of the response variable. In linear regression, it is continuous (e.g. predict blood pressure from age, weight and arterial hypertension), in logistic regression it is the probability of a dichotomous variable (e.g. predict probability for presence of atrial fibrillation with age, sex, diabetes and heart-related comorbidities). Cox models work with survival or time-to-event data and predict the hazard for experiencing the event. They are applied when logistic regression cannot be used because not all events have occurred in the observed time frame. If, for example, a new drug for a specific disease is being tested and all subjects are followed up for one year, it is very likely that not all patients will have died after one year. Additionally, some patients could have dropped out of the study during the year. These data points are called censored. It would be wrong to discard all censored data and then conduct a logistic regression because it would lead to severely biased, wrong results. Cox regression can work with censored data and is therefore used in this thesis. [47].

This chapter explains the mathematical regression background step by step, starting with linear regression and logistic regression and then leading over to Cox regression and regularized regression. It also provides several quality measures for the resulting models.

Linear Regression

Let $\mathbf{y} \in \mathbb{R}^n$ be the continuous dependent variable and $\mathbf{X} \in \mathbb{R}^{n \times k}$ be the k independent predictors in a population of n subjects. The linear regression equation for subject i ,

$i \in [1, n]$, is denoted by

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \epsilon_i \quad (2.1)$$

and summarized for all subjects by

$$y = X\beta + \epsilon \quad (2.2)$$

When all predictors have a value of zero, the value of y is β_0 . It is also referred to as the Y-intercept. The value of the coefficients β_i provide information about the amount of change in the dependent variable when x_i increases by 1 while the other predictors stay constant.

The predicted values are calculated by estimating β . A common way for doing this is to minimize the sum of the squared differences of y and \hat{y} (*least squares estimator*):

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.3)$$

The individual error (residual) is denoted by ϵ . Its value is the deviation of the observed values of y and the predicted values of y , denoted by \hat{y} [47, 48].

Revisiting our example above (predicting blood pressure from age, weight and arterial hypertension), blood pressure would be y , age x_1 , weight x_2 and arterial hypertension x_3 . Let $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)^T$ be the estimators of β with $\hat{\beta}_1$ having a value of 0.1. If subject A is 75 years old and subject B is 76 years old, both weigh 70 kg and do not have arterial hypertension, the predicted blood pressure of subject B would be 0.1 higher than the blood pressure of subject A due to the nature of the regression equation.

An illustration of the linear regression model can be found in Figure 2.2.

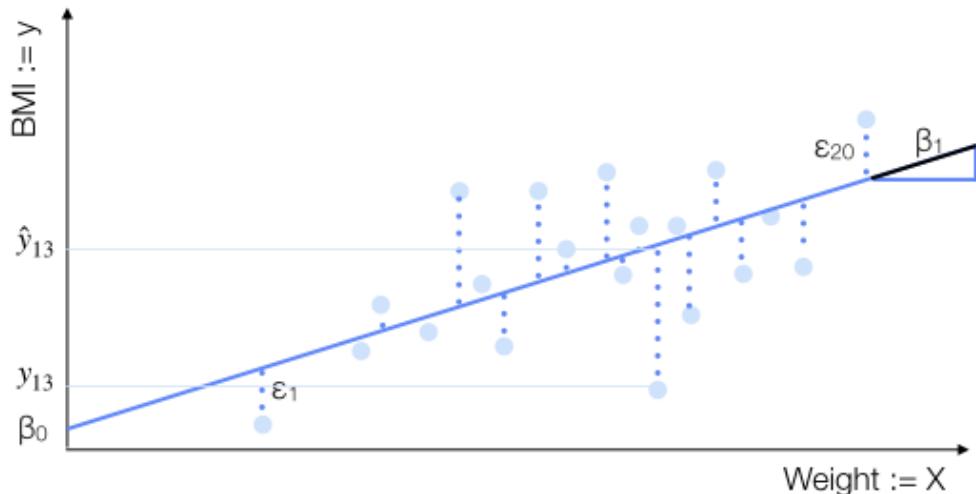


Figure 2.2: Visualization of a linear regression with just one independent variable, weight. The dependent variable, y , is the BMI. In this case, β_0 would be the value of the BMI if the person weighed 0 kg. β_1 is the slope of the regression line when there is just one variable. There are 20 persons whose measurements are symbolized by points. The dotted lines correspond to the personal error.

Logistic Regression

Like in linear regression, the underlying model of the logistic regression is linear. In a logistic regression, the logarithm of the odds that the variable has one of the two levels (here denoted by 0 and 1) are calculated. Let p be the probability that the variable is 1. Then $1-p$ is the probability that the variable is 0. The odds are defined as $\frac{p}{1-p}$. Therefore, the logistic regression formula is

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (2.4)$$

which is equivalent to

$$p = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)} \quad (2.5)$$

For estimating the β_i s, a *maximum likelihood* procedure is used [47]. It is an iterative procedure that is commonly initialized with the *least squares estimator* and then finds new values for the coefficients which have a greater likelihood of being optimal.

Cox Regression

Cox regression is a regression method specifically designed for survival and time-to-event data. It can either be used to compare the survival between two groups or to compare the survival of one population against the so-called baseline hazard, i.e. the risk of experiencing the event at time t when all covariates are zero. The latter case is described by the following equation:

$$\ln\left(\frac{h(t)}{h_0(t)}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (2.6)$$

which is equivalent to

$$h(t) = h_0(t) \cdot \exp(\beta_1 x_1 + \dots + \beta_k x_k) \quad (2.7)$$

If one is interested in comparing the survival between two groups, the baseline hazard just has to be substituted with the hazard of the other group. The Cox model is also called proportional hazards model because the underlying assumption is that:

$$\forall t \in \mathbb{R} : \frac{h_0(t)}{h(t)} = c \quad (2.8)$$

with $c \in \mathbb{R}$ being a constant. The $\hat{\beta}_i$ are obtained by a maximum likelihood procedure like in logistic regression [47, 49].

Regularized Regression

When dealing with a big data set, it is very likely, especially for biological data, that there are many possible covariates that potentially correlate with each other. As both of these points are counterproductive for regression methods, it was proposed to introduce

a penalty term to the equation that estimates the coefficients. This kind of regression is called regularized regression. Common methods include ridge, LASSO and elastic net regression. The latter is a combination of the two former methods [50].

Ridge regression obtains values for $\hat{\beta}$ by minimizing the following equation:

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \|\beta\|_2^2 = \arg \min_{\beta} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^k \beta_j^2 \quad (2.9)$$

It adds a penalty term to the *least squares estimator*, consisting of a L2-norm of β that is multiplied by the tuning parameter $\lambda \geq 0$. In order to find the minimum, one has to solve for the derivative w.r.t. β :

$$\begin{aligned} \hat{\beta} &= \arg \min_{\beta} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \|\hat{\beta}\|_2^2 \\ &= \arg \min_{\beta} \|y - X\hat{\beta}\|_2^2 + \lambda \|\beta\|_2^2 \\ &= \arg \min_{\beta} y^T y - 2y^T X\hat{\beta} + \hat{\beta}^T X^T X\hat{\beta} + \lambda \|\hat{\beta}\|_2^2 \end{aligned}$$

differentiate w.r.t. β for minimization

$$\begin{aligned} 0 &= -2X^T y + 2X^T X\hat{\beta} + 2\lambda\hat{\beta} \\ (2X^T X + 2\lambda I)\hat{\beta} &= -2X^T y \\ \hat{\beta} &= (X^T X + \lambda I)^{-1} X^T y \end{aligned}$$

It can easily be seen that $\hat{\beta} \xrightarrow{\lambda \rightarrow \infty} 0$, but no β_i can ever become zero. This is due to the fact that the numerator, $X^T y$, can not become zero as $\lambda \rightarrow \infty$, but as $\lambda \rightarrow \infty$, the denominator $(X^T X + \lambda I) \rightarrow \infty$, letting the whole term and therefore $\hat{\beta}$ approach zero. If $\lambda = 0$, the normal *least squares estimator* is obtained.[51]

LASSO regression is very similar to ridge regression, except that it substitutes the L2-norm with a L1-norm:

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \|\beta\|_1 = \arg \min_{\beta} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^k |\beta_j| \quad (2.10)$$

We repeat the steps above in order to find the minimum. Since it is rather complicated to show the steps for any vector size, I will just show it for one predictor. Let $\hat{y} := \beta_1 x$ and $x \in \mathbb{R}^n$:

$$\begin{aligned} \hat{\beta}_1 &= \arg \min_{\beta} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \|\hat{\beta}\|_1 \\ &= \arg \min_{\beta} \|y - x\hat{\beta}_1\|_2^2 + \lambda |\hat{\beta}_1| \\ &= \arg \min_{\beta} y^T y - 2y^T x\hat{\beta}_1 + \hat{\beta}_1 x^T x\hat{\beta}_1 + \lambda |\hat{\beta}_1| \end{aligned}$$

differentiate w.r.t. β for minimization

for $\hat{\beta}_1 > 0$:

$$0 = -2y^T x + 2x^T x \hat{\beta}_1 + \lambda$$

$$2x^T x \hat{\beta}_1 = 2y^T x - \lambda$$

$$\hat{\beta}_1 = \frac{y^T x - \lambda/2}{x^T x}$$

for $\hat{\beta}_1 < 0$:

$$0 = -2y^T x + 2x^T x \hat{\beta}_1 - \lambda$$

$$2x^T x \hat{\beta}_1 = 2y^T x + \lambda$$

$$\hat{\beta}_1 = \frac{y^T x + \lambda/2}{x^T x}$$

We can see, that it is now possible to tune λ in such a way that $\hat{\beta}_1$ can become zero. This also holds true for more than one predictor. This is why LASSO is able to eliminate certain predictors and therefore perform a variable selection while ridge regression is just used for shrinking them. In both regressions, the tuning of the λ parameter is used to find the optimal regression model while the minimization of the equation guarantees the minimal β_i for the chosen λ . This is why the regressions use multiple iterations until a good value for λ is found [51]. As seen above from the equations for finding $\hat{\beta}$, the larger λ is chosen, the more does ridge regression shrink the β_i toward zero and the more parameters are eliminated by LASSO regression. Since it is an iterative process, we can test the resulting model at every chosen value for λ and use the λ that, e.g. minimizes the mean squared error, the log likelihood, the area under the ROC curve or another chosen measure.

Elastic net regression combines LASSO and ridge regression in the following way:

$$\hat{\beta} = \arg \min_{\beta} \|y - X\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2 \quad (2.11)$$

Elastic net defines a parameter $\alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2}$, which transforms this equation into solving the problem:

$$\hat{\beta} = \arg \min_{\beta} \|y - X\beta\|_2^2, \text{ s.t. } (1 - \alpha)\|\beta\|_1 + \alpha\|\beta\|_2^2 \leq t, t \in \mathbb{R} \quad (2.12)$$

It is apparent that for $\alpha = 1$, the equation becomes a ridge regression and for $\alpha = 0$ a LASSO regression. For all $\alpha \in (0, 1)$, both ridge and LASSO regression is performed in proportions specified by α . [50]

There are implementations that allow those methods to be used for variable selection and coefficient shrinkage in Cox regression [52–54].

There are various other ways to select variables for the regression model. One example is forward stepwise selection (FS), in which one independent variable is added per iteration. The first variable to enter the model has the highest correlation with \mathbf{y} , the second one must have “the highest correlation with the residuals from the earlier model” [47] and so on [47, 48]. Backwards elimination (BE) works analogously, starting with the inclusion of all available variables in the model. At each step, one variable

that falls below the defined threshold the most is eliminated from the model. There is some criticism about stepwise procedures, including that it is not guaranteed “that the ‘best’ subset of a given size will be revealed” [55] and that the discrete process can overlook a very good combination of variables and leads to high variability [48, 55]. This is why LASSO regression was chosen for variable selection in this thesis.

Statistical assessment of the model quality

Routine statistical methods for the assessment of the model quality include confidence intervals for the coefficients, the Wald test, the likelihood ratio test, the goodness-of-fit statistics and ROC curves [47].

The Wald test tests the null hypothesis H_0 that the coefficient has the value zero:

$$H_0 : \beta_i = 0 \text{ versus } H_1 : \beta_i \neq 0$$

The test statistic is defined as $\frac{\hat{\beta}_i}{se(\hat{\beta}_i)}$ and follows a normal distribution under H_0 . Like with every hypothesis test, we want to reject H_0 , because we want x_i to be important in the model. The Wald test can also be applied for the whole model [47].

The likelihood ratio test is a common alternative to the Wald test. It is used for assessing whether the removal of variables from the so-called full model makes the model significantly worse. Its test statistic is defined as

$$\chi^2_{LR} = -2 \ln\left(\frac{\text{Likelihood of the reduced model}}{\text{Likelihood of the full model}}\right) \quad (2.13)$$

and follows a χ^2 distribution under H_0 . The degrees of freedom correspond to the difference of the number of parameters of the full model and the reduced model, respectively. One can also always compare the developed (full) model with one or more variables against the model with just β_0 which would correspond to the hypothesis that all β_i are zero. In this case, the Wald test and the likelihood ratio test yield similar test statistics and p-values [47].

The confidence interval (CI) of a coefficient defines the interval in which the coefficient has to be with a probability of $1-\alpha$. In many statistic programs, the confidence interval marks the $1-\alpha/2$ quantile, in which the Wald statistic of the coefficient has to be. It is calculated using z-scores and the estimates of the coefficients:

$$CI = [\exp(\hat{\beta}_i - z_{1-\alpha/2} \cdot se(\hat{\beta}_i)), \exp(\hat{\beta}_i + z_{1-\alpha/2} \cdot se(\hat{\beta}_i))] \quad (2.14)$$

If one is e.g. interested in a 95%-CI, α equals 0.05. The exponentiated coefficients of a regression equation can be easily interpreted: If the exponent is greater than 1, then the variable has a positive influence on the odds. If it is smaller than 1, the odds decrease. A coefficient is not statistically relevant, if its CI contains 1 [47].

Another commonly used statistical assessment is the ROC curve which is obtained by plotting the sensitivity against 1-specificity for different thresholds. The area under the ROC curve (AUC) is obtained by integrating over the ROC curve. This yields values between 0 and 1, where 0.5 would be the AUC of a random model and 1.0 would be the AUC of a perfect model. In medicine and biostatistics, an AUC between 0.8 and 0.9 is considered excellent [47].

There is a special statistical measurement for survival data, the concordance index or (Harrell's) C-index, which is closely related to the AUC. “The c index is defined as the proportion of all usable patient pairs in which the predictions and outcomes are concordant” [56]. It is computed by first finding all pairs of patients (P, Q) where at least one of the two patients has died and one patient lived longer than the other one. Let P be the patient that died at t_p and let Q be the patient that outlived P . Let t_q , $t_q > t_p$, be the last known time point for patient Q . There are now three possibilities for the predicted survival times \hat{t}_p and \hat{t}_q , that contribute differently to the C-index:

- $\hat{t}_p < \hat{t}_q$: The prediction is concordant to the actual outcomes. A 1 is added to the numerator as well as the denominator of the C-index [56].
- $\hat{t}_p = \hat{t}_q$: $\frac{1}{2}$ is added to the numerator, 1 is added to the denominator of the C-index [56].
- $\hat{t}_p > \hat{t}_q$: The prediction is wrong. Nothing is added to the numerator, 1 is added to the denominator of the C-index [56].

From this definition, it is clear that the C-index would be 0.5 for a random model and 1.0 for a model with perfect predictions.

2.2.2 Random Survival Forests

The concept of random forests is based on *bagging*, which essentially averages the results of numerous models created by bootstrapping [48]: Let $\mathbf{X} \in \mathbb{R}^{n \times k}$ be our sample data and $\mathbf{X}^{*b}, b = 1, 2, \dots, B$ be the bootstrapped sample data. The obtained prediction for this model is denoted by \hat{y}^{*b} . Then, the bagging estimate is:

$$\hat{y}_{bag} = \frac{1}{B} \sum_{b=1}^B \hat{y}^{*b} \quad (2.15)$$

A random forest follows this idea by drawing a bootstrap sample of the predictor variables, fitting a decision tree with that variables by finding the optimal splitting variable at each step and therefore creating a set of trees, a forest.

A tree is a directed graph without cycles, starting at one root node. Each node can have one or more children nodes but every children node can only have one parent. The root node of a decision tree contains the whole data set. The data set is then splitted by the value of a certain variable, e.g. atrial fibrillation. Since atrial fibrillation can either be present or not present, two children nodes are generated, one containing only the patients with atrial fibrillation, one containing only the patients without atrial fibrillation. If the variable can assume more than 2 discrete values, e.g. NYHA class (I-IV), four new children nodes are generated from the parent node. For continuous variables like age, thresholds have to be set, e.g. ≤ 30 , $30-50$, $50-70$, ≥ 70 .

Which variable is optimal for splitting the data set of the parent node is determined by a splitting rule like mean squared error for regression problems, AUC for classification problems or log-rank test for survival data. A random forest can be used for all of these problems [48].

For this thesis, a random survival forest was computed, i.e. a random forest that picks the best variable/split-point by comparing the survival in the formed groups. In this case, the splitting rule was a log-rank test (see Figure 2.3).

The algorithm for random forests is defined as follows:

Algorithm 1: Random Forest Algorithm, adapted from [48]

Input: Table of k predictor values for n patients $\mathbf{X} \in \mathbb{R}^{n \times k}$

for $i \leftarrow 1$ **to** B **do**

- $X^{*b} \leftarrow$ Bootstrap sample of size N from \mathbf{X}
- Grow a random forest tree T_b to the bootstrapped data
- Let n_{min} be the minimum node size
- foreach** $t \leftarrow \text{terminal node} \in T_b$ **do**

 - $n \leftarrow$ node size of t
 - while** $n > n_{min}$ **do**

 - i. Select m variables at random from the p variables.
 - ii. Pick the best variable/split-point among the m .
 - iii. Split the node into two daughter nodes.

 - end**

- end**

end

Output $\{T_b\}_1^B$

if regression problem **then**

- $\hat{y}_{rf}^B \leftarrow \frac{1}{B} \sum_{b=1}^B T_b$

end

else if classification problem **then**

- Let C^b be the class prediction of T_b
- $\hat{C}_{rf}^B \leftarrow \text{majorityVote} (\{C^b\}_1^B)$

end

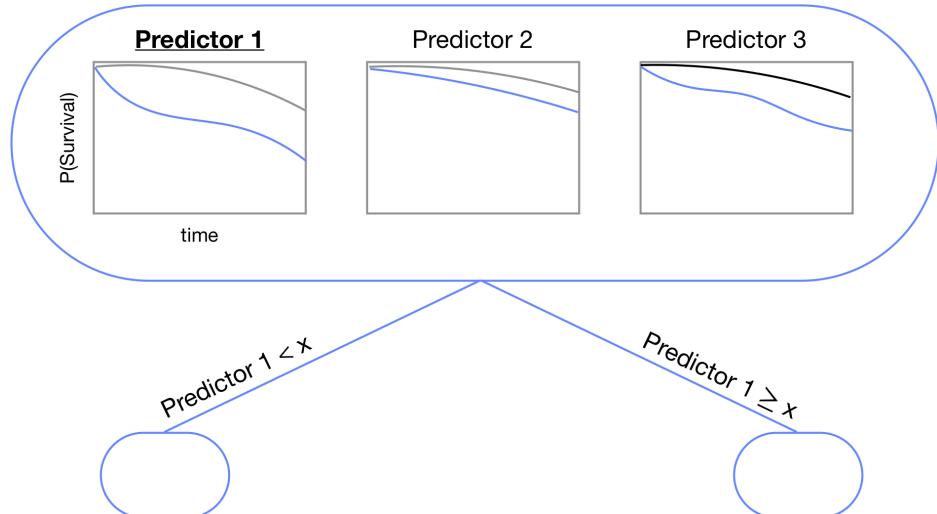


Figure 2.3: Illustration of the idea behind random survival forests. Here, a random set of $m=3$ predictor variables was selected. The survival curves for the groups, each stratified by the predictor variables, are pictured. The log-rank test would yield the best result for the first predictor. Therefore, the first variable is selected for splitting the node into two daughter nodes. In this example, variable 1 is a continuous variable with one split-point.

2.3 Related Work

The goal of this thesis is to create a risk score for transcatheter aortic valve implantation for one-year mortality prediction. This part of the chapter guides through related work that has already been done in the field. It first answers the question why the existing, widely used operative mortality prediction scores - namely the STS Score and the EuroSCOREs - are not sufficient for TAVI mortality prediction even though they showed good discriminatory power for SAVR [9, 10]. Secondly, published scores for TAVI mortality prediction are presented. Here, I discuss their weaknesses and address the question, why we are, in spite of these publications, in need of a new score. Apart from published scores, there also exist TAVI-related mortality studies that did not lead to a score but still show insights into possible predictors of TAVI-related mortality.

2.3.1 Insufficiency of the STS Score and the EuroSCOREs for TAVI mortality prediction

There have been numerous studies investigating the discriminatory power of the STS score, the logistic EuroSCORE and the EuroSCORE II for TAVI mortality prediction [9–18]. These findings are summarized in Table 2.2. Most of the studies concluded that the operative risk assessment scores do not provide satisfying results. The average logistic EuroSCORE AUC was about 0.616 [11–13, 15–18], the average EuroSCORE II AUC 0.67 [10, 12–15, 17] and the average STS score AUC 0.634 [9–18]. The results of the examination of the logistic EuroSCORE reached from an AUC of 0.49 [11] to 0.703 [17]. The EuroSCORE II performed similarly with AUCs of 0.54 [14] to 0.711 [17]. The STS score had even worse predictions for TAVI mortality. Its AUC just reached values from 0.52 [14] to 0.69 [11]. Two papers suggested the problem originating from the disproportional risk of non-femoral access points. After discarding this group of patients, the STS AUC indeed rose from 0.674 to 0.789 [9] and the EuroSCORE II AUC from 0.68 to 0.74 [13]. Even though this discovery suggests a strong correlation it might just be coincidental because the latter paper also used this approach for the STS score and the logistic EuroSCORE which just resulted in a decrease of the logistic EuroSCORE AUC from 0.65 to 0.61 and a constant remaining STS score AUC of 0.6.

Score	Average AUC	Range	Publications
Logistic euroSCORE	0.616	0.49-0.703	[11–13, 15–18]
euroSCORE II	0.67	0.54-0.711 (only transfemoral access: 0.74)	[10, 12–15, 17]
STS Score	0.634	0.52-0.69 (only transfemoral access: 0.789)	[9–18]

Table 2.2: Overview of the performance of the STS score, the logistic EuroSCORE and the EuroSCORE II for TAVI mortality prediction.

2.3.2 Published Scores for TAVI mortality prediction

An overview of the published scores for TAVI is given in Table 2.3. The scores can mainly be separated in scores for the prediction of in-hospital mortality (NIS TAVR [36] and STS/ACC TVT Registry Score [57]), 30-day all-cause mortality (CoreValve [37], OBSERVANT [25], UK-TAVI [21] and FRANCE-2 score [20]) and 1-year all-cause mortality (CoreValve [37], TAVI₂ [24], C₄CAPRI [22] and TARIS score [23]). Apart from that, there is also the Predictor Of Poor Outcomes score that defines poor outcome as either death, bad quality of life or decrease in quality of life. It is derived from the PARTNER trial population [26].

The Predictor of Poor Outcomes has a bad discrimination, even in the development cohort. For 6 months and for 1 year, the C-index of the derivation cohort was 0.66 (0.64 and 0.62 in the respective validation cohort). Both cohorts contained only high or prohibitive surgery risk patients (mean STS score 11.5). The model was computed with an iterative backward selection strategy, starting with 25 candidate baseline variables. Since there are many more interesting biomarkers than 25, one can argue that this preselection led to the bad result. Furthermore, the results cannot be applied to intermediate or low surgery risk patients [26].

Below, the in-hospital mortality scores, the 30-day mortality scores, the 1-year mortality scores and further biomarker mortality studies are presented and discussed. Overall, TAVI scores were mainly computed for extreme and high surgery risk patients. This is the case because the topic whether TAVI is non-inferior or even superior to SAVR is still widely discussed in literature. Since the median survival of medically treated AS patients is solely one year [1], TAVI was therefore first tried in prohibitive and extreme surgery risk patients. After the PARTNER and the CoreValve trial showed that TAVI was indeed non-inferior/ superior [2, 3], it was also tried for high, intermediate and low surgery risk patients. The results from these studies are summarized in Table 2.4. In comparison to SAVR, TAVI was non-inferior in all nine discussed studies for extreme, high, intermediate and low surgery risk patients. In six of these nine studies, TAVI was even superior to SAVR regarding the mortality rate. It must however be noted that there were more major strokes, major vascular events and aortic regurgitations found in the TAVI group for most of the mentioned studies. For SAVR, there are mainly more reported severe bleedings, new-onset atrial fibrillations, strokes and rehospitalizations.

In-hospital mortality scores

Even though the STS/ACC TVT Registry has the biggest cohort size of all (N = 20586), the C-index of the developed score (0.66) is not as good as one could expect. Despite the fact that a logistic regression with forward and backwards steps was used for the model, which is a broadly accepted method, the initial candidate covariates were picked on the basis of literature and expert opinion. From a computational point of view, this could explain the moderate discrimination of the final model. Like the Predictor of Poor Outcomes, the registry only included high and extreme surgery risk patients, making the model inapplicable for intermediate and low surgery risk patients [57].

The discrimination of the NIS TAVR score is excellent (validated C-index of 0.92) and far better than all other published scores (see Table 2.3). One main reason for this

Score, Year	Endpoint	Validated C-index	Cohort size	Validation	Variables
NIS TAVR, 2019 [36]	In-hospital mortality	0.92	10883	2:1 splitting	AKI, cardiogenic shock, fluid and electrolyte disorders, cardiac arrest, sepsis, dyslipidemia, hypertension, coagulopathy, smoking, vascular complications
CoreValve, 2016 [37]	1-year all-cause mortality	0.79	3687	2:1 splitting	Home oxygen use, Albumin<3.3g/dL, falls in the past 6 months, STS score>7, severe Charlson Comorbidity Score
CoreValve, 2016 [37]	30-day all-cause mortality	0.75	3687	2:1 splitting	Home oxygen use, assisted living, Albumin<3.3g/dL, Age>85
TAVI ₂ , 2015 [24]	1-year all-cause mortality	0.715	511	100 x bootstrapping	age, gender, porcelain aorta, anemia, LVEF, recent MI, Δ Pm, renal dysfunction
OBSERVANT, 2014 [25]	30-day all-cause mortality	0.71	1878	2:1 splitting	GFR<45mL/min, critical state, pulm. hypertension, diabetes, NYHA 4, prior BAV, LVEF<40%
C ₄ CAPRI, 2019 [22]	1-year all-cause mortality	0.67	2227	ca. 6:1 splitting	age, gender, GFR, COPD, peripheral vasc. disease, previous stroke, LVEF, pulm. pressure, Δ Pm, dyspnea, mitral regurg. level, thoracic aortic calcification
UK-TAVI, 2018 [21]	30-day all-cause mortality	0.66	6339	bootstrapping	age, gender, BMI, BMI^2 , GFR, pulm. disease, extracardiac arteriopathy, sinus rhythm, prior BAV, critical status, poor mobility, KATZ score, pulm. artery pressure, elective, access route
STS/ACC TVT Registry, 2016 [57]	In-hospital mortality	0.66	20586	2:1 splitting	age, GFR, dialysis, NYHA 4, severe chronic lung disease, access route, procedural acuity categories
Predictor of Poor Outcomes, 2014 [26]	poor outcome after 6 months	0.64	2137	2:1 splitting	gender, diabetes, major arrhythmia, creatinine, mean art. pressure, BMI, oxygen-dependent lung disease, Δ Pm, Mini-Mental Status Examination, 6-Min Walk Test distance
Predictor of Poor Outcomes, 2014 [26]	poor outcome after 1 year	0.62	2130	2:1 splitting	above except for diabetes and mean art. pressure
TARIS, 2014 [23]	1-year all-cause mortality	0.6	847 + 333	external validation	age, gender, BMI, pulm. hypertension, GFR, Hb, Δ Pm, LVEF<45%
FRANCE-2, 2014 [20]	30-day all-cause mortality	0.59	3833	2:1 splitting	Age≥90, BMI, NYHA 4, pulmonary oedema, pulm. hypertension, respiratory insufficiency, dialysis, access route, critical state

Table 2.3: Overview of the published scores for TAVI. 2:1 splitting into derivation and validation cohort.

could be that “rather than considering fixed assumptions on data behavior and variable preselection [...] [NIS TAVR uses machine learning algorithms that] allow the data to create the model by detecting or learning underlying patterns” [36]. The authors tried different machine learning algorithms, namely logistic regression, artificial neural networks, naïve Bayes and random forests. Logistic regression outperformed the other learning techniques, but not by much. The features were picked by a regularized logistic regression with an L2 norm penalty as feature scoring function. Looking at the features of the final model, one can easily understand why the model performs so well. It includes variables that were measured after the TAVI procedure like acute kidney injury (AKI), cardiogenic shock, cardiac arrest and sepsis [36]. If a patient experiences only one of those events, it is obvious that his in-hospital mortality probability rises. As the real challenge consists of predicting TAVI-related mortality before the procedure, this score does not really help with making a decision whether to perform TAVI.

30-day mortality scores

The CoreValve 30-day score behaves quite well. With a concordance of 0.75 in the validation cohort and 0.76 in the derivation cohort, it is very robust even though it consists of just four variables. The variables are mainly frailty related (home oxygen use, assisted living, age > 85) [37]. As mentioned in 2.1.2, a decreased albumin concentration can be observed in many forms of illness. This might also be a downside, though, because one cannot infer the original risk parameter. Thus, the risk might come from liver disease, infections, diabetes, cancer, hepatitis or other diseases [46]. As the cohort consists mainly of high and extreme surgery risk patients (mortality rate of 5.8% at 30 days and 22.8% at 1 year) and 37.6% have diabetes, 2.4% have liver cirrhosis, 12.1% have chronic kidney disease and 57.3% have a Charlson Comorbidity Index ≥ 5 (“associated with a 1-year mortality rate of 85%” [37]), the low albumin level could represent a lot of comorbidities. Like in other papers, the initial set of variables was selected by hand on the basis of “clinical utility” [37]. Then, univariate predictors with a p-value < 0.05 in a Cox model were kept and put into a multivariate Cox model that was modified by stepwise analyses with an entry and exit threshold [37].

The OBSERVANT score is also based on a high surgery risk cohort (6% mortality at 30 days, 65.2% with NYHA class III/IV, mean age 81.9) [25]. The parameters for the multivariable model were selected by univariate logistic regression analysis. After that, a stepwise logistic regression analysis combined with bootstrapping of the data was performed in order to get to the final model. Even though LVEF and prior BAV did not have the required p-value, they were forced into the final model because of literature evidence. This decision somewhat undermines the point of the initial decision to search for relevant parameters with the help of regression methods, but still led to a good concordance. Frailty measurements were also considered but were not found to be univariately associated with 30-day mortality and did not ameliorate the model performance [25].

For the UK-TAVI population, no STS score was calculated, but as the 30-day mortality was 5.14% and 17.2% have a NYHA class 4, one would suppose that the surgery risk was slightly less high than in the other cohorts mentioned above. For the model, all

available baseline variables were used, but the registry did not record frailty measures like poor mobility and the KATZ score (KATZ Index of Independence in Activities of Daily Living) until 2013. The whole population is derived from data between 2009 and 2014. Since the authors wanted to include the frailty measures, they first performed backwards selection using Akaike's information criterion for the data from 2009-2014 with the baseline variables that were available before 2013. Then, they updated the model by fitting a logistic regression model for the 2013-2014 data with the predictors from the 2009-2014 model and the new, frailty related covariates. Contemporary trends in the data should also have been taken into account. They also experimented with "quadratic and cubic transformations of the continuous variables" [21], which is how the BMI and the BMI^2 ended up in the model. Since, however, these values correlate highly, there should be no additional benefit from including both parameters in the model.

The FRANCE-2 score exceeds a random model by only 9 percent points, making it the worst model for TAVI mortality prediction. The population consists of high or prohibitive surgery risk patients. The authors, too, first performed an univariate analysis with a logistic model, included the covariates with a p-value <0.2 in a multivariate model, which they adjusted with a backwards selection procedure. Some predictive power could have been lost by dichotomizing all continuous variables with clinically relevant thresholds. Another reason may be that most of the 10% who died within the first 30 days, died on the day of surgery or in the first week [20]. Defining "in-hospital mortality" as endpoint might have achieved better results.

1-year mortality scores

The CoreValve 1-year mortality score performs even better than its 30-day counterpart with a C-index of 0.79 in the validation cohort and 0.83 in the development cohort. Additionally to the 30-day predictors, it adds falls in the past 6 months, which is also frailty related, an STS score greater than 7 and a Charlson Comorbidity Index ≥ 5 [37]. Aside from the criticized points mentioned above, it is also important to take a look at the last two variables. Even though the 1-year score might give the impression that only five variables are needed to compute it, it really consists of many more. The STS score alone has over 60 possible variables with nearly 100 subcategories [5–7]. The Charlson Comorbidity index is composed of 18 variables [37]. Thus, the actual number of covariates integrated in this score is over 180, making the score considerably more intense in terms of computation.

The TAVI₂ cohort is the smallest one (511 patients) which could be regarded as a weak point. The mean STS score was 16.6, making the cohort a high and prohibitive surgery risk cohort. Like other models before, the authors of TAVI₂ used univariate Cox models for initial variable selection and included the variables with a p-value <0.05 in a multivariate Cox model which they reduced by backwards elimination. Because of the small cohort size, bootstrapping was performed 100 times in order to correct for overfitting [24]. The variables seem quite reasonable in comparison to the other scores and based on literature research, even though the model can not be extrapolated for intermediate or low surgery risk patients.

C₄CAPRI is the first score that takes thoracic aortic calcification into account. Looking

at the fact that C₄CAPRI is based on the same cohort as the FRANCE-2 score, but achieves a much better C-index, one can infer that several things were optimized in the process of score development. Data from 2010-2014 was used as development cohort. The score was tested on all patients that entered the registry in 2015. As initial variables, biomarkers that were prognostic according to literature were chosen (in addition to calcification). Since some biomarkers correlate, the Spearman, the tetrachoric resp. the polychoric correlation coefficients were computed for all of the variables and only the nonredundant ones were chosen. The final Cox model was computed by integrating variables based on their prediction improvement (quantified by likelihood ratio test and integrated discrimination improvement index). The calcification level was very significant (p-value < 0.01). The AUC value was additionally corrected by bootstrapping [22].

The TARIS cohort includes also only high and extreme surgery risk patients (mortality rate 24.5% at 1 year). The model was computed by an age- and sex-adjusted Cox model. Instead of computing univariate models first, variable selection was done by applying LASSO regression to the Cox model variables. The C-index of the development cohort was 0.66. This is the only score that verifies its results with an external cohort - a different registry with a high risk population [23].

Further mortality biomarker studies

In general, it has been suggested that advanced age, a poor LVEF, NYHA class III or IV, a MI less than 24h before surgery, concomitant procedures (increase the operative time and complexity of the procedure), a small body surface area (BSA), previous cardiac operations, a calcific AS and presence of comorbidities influence the outcome of AS patients negatively [1]. Particularly, a low LVEF seems to be a major predictor of outcomes [1].

In SAVR-treated patients, the ones with “normal preoperative LVEF have an estimated 15-year survival of 62.3%, the presence of left ventricle (LV) dysfunction significantly reduces the long-term survival after aortic valve replacement (AVR). It is 47.6% for those with LVEF 35%-49%, 41% for LVEF 20%-34% and 36.9% for LVEF <20%.” [1]. Apart from the already mentioned biomarkers, 30-day mortality predictors for SAVR included coronary artery disease, wrong prosthesis size, emergency surgery and female gender [1]. For long-term survival, low ΔP_m , atrial fibrillation and systemic hypertension were also relevant [1].

For TAVI, Moat *et al.* pointed out that “survival was significantly adversely affected by renal dysfunction, the presence of coronary artery disease, a nontransfemoral approach” as well as by “left ventricular function (ejection fraction <30%), the presence of moderate/severe aortic regurgitation, and chronic obstructive pulmonary disease” [39].

Seiffert *et al.* discovered a BMI < 20, the NYHA class IV, a LVEF <30%, the STS Score, and age <75 to be the most valuable predictors for all-cause mortality prediction at 1 year [40]. Schymik *et al.* however, defined a nontransfemoral approach, renal insufficiency, liver disease, moderate or severe tricuspid regurgitation, presence of a porcelain aorta and atrial fibrillation as the best markers for 1 year mortality prediction after TAVI [38].

Surgery risk group	Non-inferiority of TAVI reported	Superiority of TAVI reported	Complications in TAVI	Complications in SAVR (resp. medical treatment in first group)
Prohibitive / extreme High	Yes [2, 3]	Yes [3]	major strokes, major vascular events [3]	repeat hospitalization, cardiac symptoms [3]
	Yes [58–61]	Yes [60, 61]	major strokes, major vascular complications, aortic regurgitation [58, 59, 61]	major bleedings, new-onset atrial fibrillations, strokes [58, 61]
Intermediate	Yes [29–31]	Yes [29, 31]	major vascular events, aortic regurgitation, pacemaker implantations [29, 30]	bigger AVAs, AKIs, severe bleedings, transfusions, new-onset atrial fibrillation, strokes, regurgitation [29–31]
Low	Yes [32, 33]	Yes [32]	aortic regurgitation, pacemaker implantations, higher NYHA class [33]	strokes, rehospitalization, bleedings, cardiogenic shock, AKIs, atrial fibrillation [32, 33]

Table 2.4: Summary of the findings of SAVR versus TAVI mortality studies.

Rodés-Cabau *et al.* examined biomarkers for 3-year mortality. Their results show that frail patients with COPD, chronic kidney disease and chronic atrial fibrillation would be at most risk [42]. Duncan *et al.* defined renal dysfunction, atrial fibrillation, respiratory dysfunction and LVEF<30% as most prognostic biomarkers for 3 years and together with coronary artery disease and age for 5 years [41].

Conclusion

In the studies for mortality biomarkers, there are a few variables reported several times. Poor LVEF, NYHA class IV, atrial fibrillation, COPD, renal dysfunction, age and a non-transfemoral approach are among these biomarkers. All of these biomarkers also occur in published scores for TAVI mortality prediction (see Table 2.3).

As most of the scores only have a high or extreme surgery risk cohort, the scores cannot be applied to intermediate and low risk patients. The need to develop a score that also includes this group is urgent with TAVI becoming a standard procedure.

A further weakness of most existing scores consists in picking the initial variables on the basis of expert opinion rather than automating the whole process. Since the expert opinion was already part of the registry design, variables should not be discarded from the registry without statistically examining their predictive power. As machine learning techniques are currently attracting more and more notice and are becoming better every year, a fully computational approach like in NIS TAVR seems to be the way to go for such big registries. Of course, these learning techniques still have some underlying flaws, which is why a medical background is required to judge the results. Therefore, it is the ambition of this thesis to combine a mathematical sophisticated approach with an evaluation in light of biological knowledge.

3 Material and Methods

In this chapter, the data used from the Swiss-TAVI registry and the workflow of the data analysis are portrayed. Chapter 3.1 first describes the Swiss-TAVI study per se including study sites, patient in- and exclusion criteria, basic procedure information, as well as patient information collection and follow-up circumstances. Second, the in- and exclusion of patients and features specifically for this analysis are illustrated.

In the following two chapters, the statistical methods and the different approaches for developing a Cox model are addressed. All data preprocessing and computational analyses were made with R in RStudio [62, 63].

3.1 Swiss-TAVI cohort

The Swiss-TAVI registry is listed as “prospective, national, multi-center registry of patients undergoing transcatheter aortic valve implantation” [64]. It is controlled by the University Hospital Inselspital in Bern, who invited all Swiss cardiovascular centers implanting CE approved TAVI devices to contribute [64]. This includes the university hospitals of Bern, Basel, Geneva, Lausanne and Zurich as well as other centers in Lucerne, Lugano, Basel and Zurich [65]. For this thesis, only the patients registered at the university of Zurich were used.

The official study start date is February 2011, the estimated completion date May 2020. The first procedure recorded in the Zurich subset took place in April 2012, the last one in November 2019. In this time span, 1651 patients were registered at the university hospital of Zurich.

All patients with “symptomatic, severe aortic stenosis [or] degenerated aortic bioprosthesis requiring treatment” [64] who were deemed to undergo TAVI were eligible for enrollment. They were excluded from the study if they either refused participation or had a “high probability of non-adherence to the follow up - requirements” [64].

For the baseline table, the Swiss TAVI registry first assessed basic patient characteristics as well as many laboratory measurements. Secondly, important heart- and non-heart-related comorbidities were recorded. Following the clinical standard, each patient received an electrocardiography as well as a TTE or TEE. From these measurements, several anatomical aspects of the heart were assessed and recorded. Sometimes, a CT or MRI was performed as well. For each patient, the STS score and the EuroSCORE II were computed and a NYHA and a CCS class was assigned to the patient. Lastly, procedure-specific variables were recorded in the baseline table.

As this thesis uses only the baseline variables that were recorded before the TAVI procedure, all variables relating to procedure complications were excluded as well as

all measures that were made after the placement of the artificial valve. Furthermore, variables that indicated dates, units, the study site and variables that either had 10% NAs or nearly no variance at all were excluded for the regression analysis. Since categorical variables, where the category is indicated by a number, are weighed wrongly by the regression method, they were recoded to boolean variables indicating whether a certain category applies. An overview of the variable preselection can be found in Figure 3.2.

96% of the patients received the artificial aortic valve with a transfemoral approach. Of these, 83% were performed in local anesthesia. The preferred access site was through the right femoral artery. Only 32 patients were approached transapically and only 2 directly aortical. The most commonly implanted valve sizes were 29mm, 26mm and 23mm. In total, 12 different devices from six companies (Allegra, Boston Scientific, Direct Flow Medical, Edwards, Medtronic and St. Jude Medical) were implanted. All Edwards devices are balloon-expandable, the rest are self-expandable.

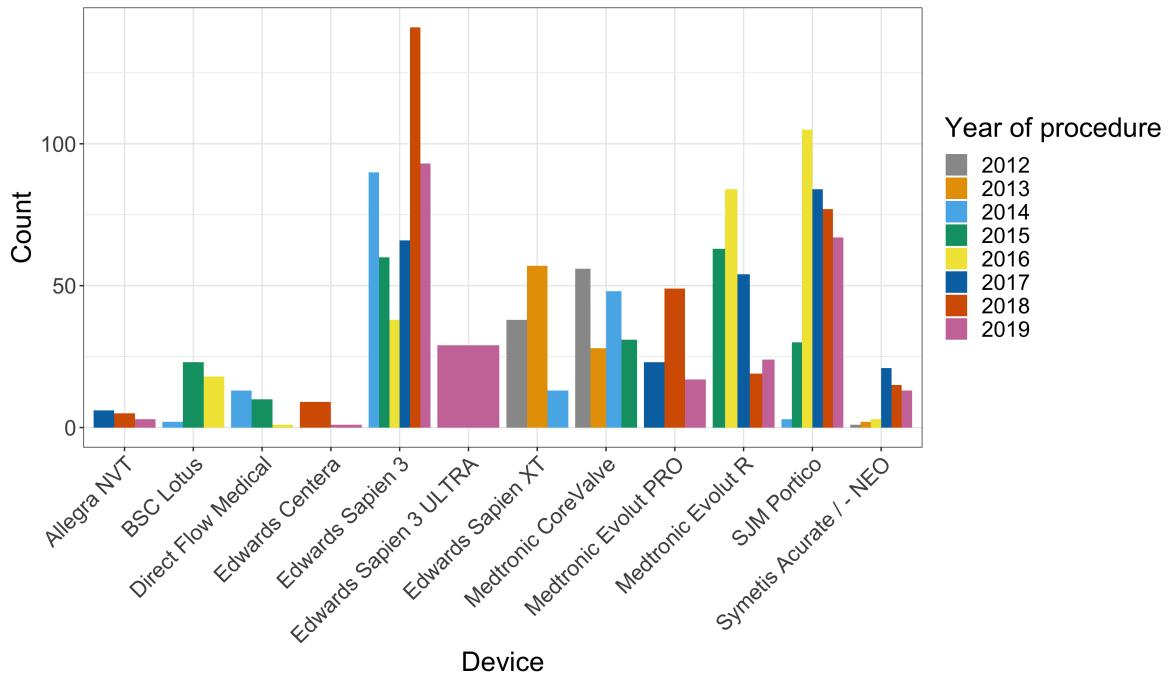


Figure 3.1: Distribution of the implanted TAVI devices. The distribution changes over the years as devices were superseded by their successors (e.g. Edwards Sapien XT by Edwards Sapien 3) or as new companies entered the market (e.g. Allegra NVT).

Figure 3.1 shows that the rate in which each device was implanted changes over the years for different reasons. One reason is that the companies developed new valves that replaced the old ones. E.g., Edwards first produced the Edwards Sapien XT, which was used from 2012 to 2014. It was superseded by the Edwards Sapien 3 (2014-2019), Edwards Centera (2018 and 2019) and Edwards Sapien 3 ULTRA (2019). In contrast, Boston Scientific withdrew all their Lotus valves from the market in 2017 due to reported malfunctions [66], whereas Direct Flow Medical was forced to close in 2017 after a funding deal had collapsed [67]. Since these fluctuations introduce a strong bias to any regression model, all devices were excluded from the Cox models in

this thesis after not having seen any specific, significant correlation to the events (see Supplementary Figure 5.1).

The patients were followed up for 30 days and yearly after the procedure, up until five years. For four patients, a seven year follow-up was made. In Supplementary Figure 5.2, it can be seen that most of the patients were just followed up for one year and most of the follow-ups were made 30 days after the procedure. For some of the patients, a follow-up was recorded but with the status 'vital status unclear, patient not traceable'. All of these follow-ups were discarded for the analysis and are therefore not displayed in Supplementary Figure 5.2. For several patients, a follow-up could not be made but it was noted whether the patient was still alive. These follow-up time points were kept since the goal of this thesis was to design a mortality score.

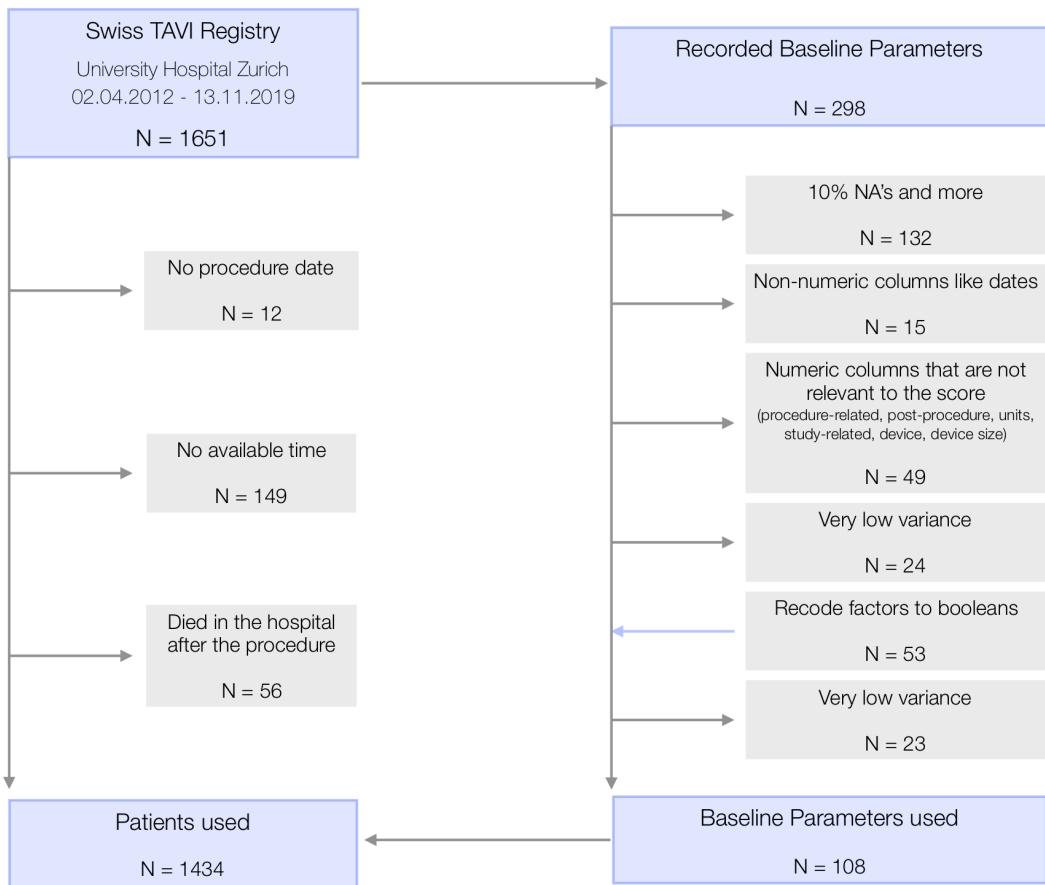


Figure 3.2: Flowchart of the patient and baseline variable preselection progress.

Missing data was handled by replacing the NA's with the median of the particular biomarker. The median was chosen because on the one hand, it is not that sensible to outliers as the mean, on the other hand, because it not only works for continuous but also for discrete values since it chooses the datapoint that separates the upper half of the values from the lower half. It is therefore always a value that exists in the data set. Patients who died in the hospital were discarded as the mortality risk score should not assess the in-hospital mortality. Furthermore, these patients most likely died of surgery complications that do not directly relate to the TAVI valve approach per se. Since a time-to-event analysis was performed, all patients who did not have a single

follow-up were also discarded. Lastly, twelve patients did not possess a value for the variable “Procedure Date”, which is why they were also excluded from the analyses. For the Cox models, the data was censored at 365 days. If the patient survived the 365 days, no event was recorded. If there was no follow-up available that spanned the time frame of a year, the last known follow-up status was recorded for the patient. Altogether, the analyses were therefore conducted with 1434 of 1651 patients and 78 of 298 variables. Because of the transformation of the categorical (factor) variables to boolean variables, the number of parameters was 108 instead of 78. The whole process is summarized in Figure 3.2.

3.2 Statistical analyses

The important baseline variables are described either by mean \pm standard deviation for the continuous parameters or by the relative frequency for categorical/boolean parameters in Table 4.1. In order to assess whether there is a significant difference in the parameter distribution when stratified by the event (all-cause mortality), a t-test was conducted for the continuous features. Since the standard t-test expects equal variance in both groups, the F-test, which tests the null hypothesis that the ratio of the variances of both groups is 1, was conducted first. When the null hypothesis of the F-test could be rejected ($\alpha = 0.05$), Welch’s t-test was performed instead of the standard t-test. For the boolean-encoded variables, Fisher’s exact test was performed. Since both statistical testings require multiple testing correction, the resulting p-values of both the t-tests and the Fisher’s exact tests were corrected with the Benjamini-Hochberg method. Corrected p-values below a significance level of 0.05 were classified as significant and are marked in Table 4.1. All of these statistical testings were performed by the *stats* R package [68].

Visualizations of the baseline variable distributions can be found in the supplementary (Figures 5.4, 5.3, 5.5).

3.3 Model inference and statistical testing

For variable selection and coefficient shrinkage in Cox regression, the *glmnet* R package was used [53, 54]. The α parameter of the elastic net was always set to 1 (simple LASSO regression). Variable selection was done with the 10-fold crossvalidation function of the LASSO regression of *glmnet* (*cv.fit*), the resulting variables (λ parameter not zero) were used for the Cox models.

The Cox models were fitted with the *coxph* function of the *survival* R package [69, 70]. The package provided several statistical methods for model evaluation. These included the Concordance- / C-index, the Likelihood ratio test and the Wald test. The Likelihood of the model was used for the calculation of the Akaike information criterion (AIC) and the Bayes information criterion (BIC) with the *stats* package of R [68]. For the individual coefficients, the standard error, the Wald statistic, the p-value of the Wald statistic and 95% confidence intervals were computed by the *survival* package. The models were also tested with a 10-fold cross validation (*iAUC* from *AUC.cd*, *survAUC* package [71]) and 100 permutation tests. Furthermore, the *coxph*

function provided the centered linear predictors of the regression for each individual. With the predictors, Kaplan-Meier curves (*survfit* from *survival*) were calculated as a sanity check for the success of the Cox regression. Additionally, the correlation between the coefficients in the Cox model was calculated with the *stats* R package (function *cor*) and visualized with *pheatmap* [72]. This served as a sanity check for the success of the LASSO variable selection which was supposed to cancel out redundant variables. A random survival forest was fit to the data with the *ggRandomForests* R package [73]. A total of 2000 trees was built using a logrank test as splitting rule. The Out-of-Bag (OOB) error rate was calculated for the set of trees. The package computes two kinds of measures for variable influence: VIMP (variable importance) and minimal depth. The former is computed by permutating the variable in the Out-of-Bag sample, “dropp[ing it] down the tree [and calculating the] OOB estimate of prediction error [...]. The difference between this estimate and the OOB error without permutation, averaged over all trees, is the VIMP of the variable” [74]. The minimal depth measure focuses on the minimal distance between the root node and the variable of interest in a certain tree. A variable close to the root indicates a very significant logrank test and therefore a significant variable. Minimal depth averages these minimal distances over all built trees [74]. The *ggRandomForests* package also provides a comparison between the minimal depth and the VIMP measure.

The most interesting variables from there were picked for the calculation of a Cox model. Apart from this Cox model and from the model using all available variables from the LASSO regression, several Cox models using other subsets of variables or patients were computed. One Cox model used the variables that had significant Benjamini-Hochberg corrected p-values. Another model used all variables but only patients that had a STS score equal or smaller than 8 because all known scores were inferred from high or extreme surgery risk patients.

4 Results and Discussion

In this section, the baseline characteristics and the statistical tests conducted on the Swiss-TAVI baseline table are presented and discussed first. In Section 4.2, all five developed models are described and compared. The best of these five models is examined in more detail in Section 4.2.2. At the end, limitations of the derived score are debated.

4.1 Baseline characteristics

	Overall (n=1434)	Alive (n=1310)	Deceased within one year (n=124)	BH- adjusted p-value
Continuous Variables				
Age [years]	80.5 ± 7.5	80.3 ± 7.6	82.1 ± 6.7	0.027
Height [cm]	165.3 ± 9.1	165.3 ± 9.1	165.9 ± 9.1	0.483
Weight [kg]	74 ± 15.5	74.1 ± 15.5	72.4 ± 15.7	0.294
BMI	27 ± 5	27.1 ± 5	26.2 ± 5	0.112
BSA	1.9 ± 0.2	1.9 ± 0.2	1.8 ± 0.2	0.338
EuroSCORE II	4.5 ± 4.4	4.3 ± 4.2	6.3 ± 6	0.002
STS Score	4.6 ± 3.6	4.4 ± 3.4	6.3 ± 4.7	<0.001
Creatinine [μmol/L]	112.7 ± 74	110.9 ± 74.2	132.4 ± 68.9	0.007
GFR	53.4 ± 24.8	54 ± 23.4	46.6 ± 36	0.051
Hb [g/L]	123.7 ± 18.2	124.4 ± 18	116.2 ± 18.8	<0.001
Thrombocytes [× 10 ⁹ /L]	228.6 ± 77.8	227.8 ± 76.8	236.2 ± 87.4	0.338
Creatinine kinase [U/L]	96.5 ± 73	97.5 ± 72.7	85.8 ± 75.6	0.128
Heart rate	74.1 ± 14.7	73.9 ± 14.7	75.5 ± 14.9	0.294
LVEF [%]	54.3 ± 13.4	54.7 ± 13.1	50.8 ± 15.4	0.019
ΔPm [mmHg]	40.8 ± 16.2	41.2 ± 16.1	36.7 ± 16	0.008
Categorical Variables				
Male gender	54.3%	53.4%	64.5%	0.065
Troponin T	85.4%	84.5%	94.4%	0.015
Troponin hs	13.7%	14.6%	4.8%	0.013
Heart-related comorbidities				
Atrial fibrillation	19.7%	18.5%	32.3%	0.007
Atrio-ventricular block	9.9%	9.5%	13.7%	0.34
LBBB or RBBB	13.6%	13.5%	15.4%	0.725
Coronary artery disease	48.7%	48.3%	52.4%	0.611
Prev. pacemaker implantation	7.7%	7.7%	8.1%	0.399

Prev. aortic valvuloplasty	5.9%	6%	4.8%	0.789
History of PCI	23.9%	23.6%	27.4%	0.610
History of MI	9.8%	9.1%	16.9%	0.039
Syncope	9.6%	9.3%	12.9%	0.379
Aortic regurgitation grade 1	17.6%	17.8%	16.1%	0.799
Aortic regurgitation grade 2	69.9%	69.7%	72.6%	0.7
Aortic regurgitation grade 3	7.5%	7.6%	6.5%	0.91
Aortic regurgitation grade 4	4.9%	4.9%	4.8%	1.0
Mitral regurgitation grade 1	10.5%	11%	5.6%	0.197
Mitral regurgitation grade 2	66.2%	67.3%	54.8%	0.032
Mitral regurgitation grade 3	18.7%	17.6%	29.8%	0.013
Tricuspid regurgitation grade 1	27.1%	27.5%	23.4%	0.586
Tricuspid regurgitation grade 2	62.1%	62.7%	56.5%	0.364
Tricuspid regurgitation grade 3	8.3%	7.6%	15.3%	0.03
History of cardiac surgery	15.2%	15%	17.7%	0.631
NYHA class I	17.7%	18.2%	12.1%	0.256
NYHA class II	33.5%	34.2%	26.6%	0.233
NYHA class III	42%	41.4%	48.4%	0.34
NYHA class IV	6.8%	6.2%	12.9%	0.034
CCS class 0	76.7%	76.9%	74.2%	0.679
CCS class 1	9%	9.2%	7.3%	0.74
CCS class 2	9.4%	9.2%	11.3%	0.628
CCS class 3+4	4.9%	4.7%	7.3%	0.373
Other comorbidities				
Peripheral artery disease	14.2%	12.7%	29.8%	<0.001
Arterial hypertension	76%	76.3&	73.4%	0.679
Diabetes	26.4%	25.6%	34.7%	0.105
COPD	12.6%	11.7%	22.6%	0.012
Dyslipidemia	48.6%	48.9%	46%	0.725
History of cerebrovascular accident	11.2%	11.1%	11.3%	1.0
Procedure-related factors				
Local anesthesia	80.5%	81.1%	74.2%	0.213
Concomitant procedure	14.3%	13.9%	18.5%	0.364
Concomitant PCI	5%	5%	4.8%	1.0
Other concomitant procedure	8%	7.4%	14.5%	0.036
Expandable valve type	65.4%	65.1%	68.5%	0.61
Additional CT performed	75.6%	76.6%	64.5%	0.026
Femoral access	96%	96.4%	91.9%	0.092
Medication				
Aspirin	56.8%	57.4%	50.8%	0.368

P2Y12	18.3%	17.9%	22.6%	0.404
Clopidogrel	16.3%	16.1%	18.5%	0.689
Marcoumar/Sintrom	17.3%	16%	30.6%	0.005
Novel Anticoagulation drugs	31.5%	30.5%	42.7%	0.03
Rivaroxaban	8.9%	9.2%	6.5%	0.617
Statin	58.7%	59.2%	54%	0.51
ACE inhibitor	29.9%	29.2%	37.1%	0.223
ATII antagonist	22%	22.4%	16.9%	0.364
Betablocker	47.7%	46.6%	58.1%	0.065
Ca-antagonist	21.1%	21.5%	17.7%	0.601
Diuretics	56.6%	54.7%	76.6%	<0.001
Steroids	11.3%	10.8%	16.1%	0.249
Insulin	8.6%	8.2%	13.7%	0.135
Oral antidiabetics	17.3%	17%	20.2%	0.61
Anticoagulants: Only Aspirin	39.3%	40.7%	25%	0.007
Anticoagulants: Only Novel Anticoagulation drugs	24.9%	24.3%	31.5%	0.223
Anticoagulants: Aspirin + P2Y12, no novel anticoagulation drugs	12.2%	11.8%	16.1%	0.373
No anticoagulants	13.6%	13.7%	12.9%	0.932

Table 4.1: Baseline characteristics of the Swiss-TAVI database stratified by event. Continuous variables are represented by mean \pm standard deviation. Categorical variables are given in percent. For the continuous variables, a t-test was computed in order to compare the two groups (Welch's t-test for variables for which the null hypothesis of the F-test could be rejected). For the categorical variables, Fisher's exact test was used. The p-values of both tests were each adjusted with the Benjamini-Hochberg (BH) correction.

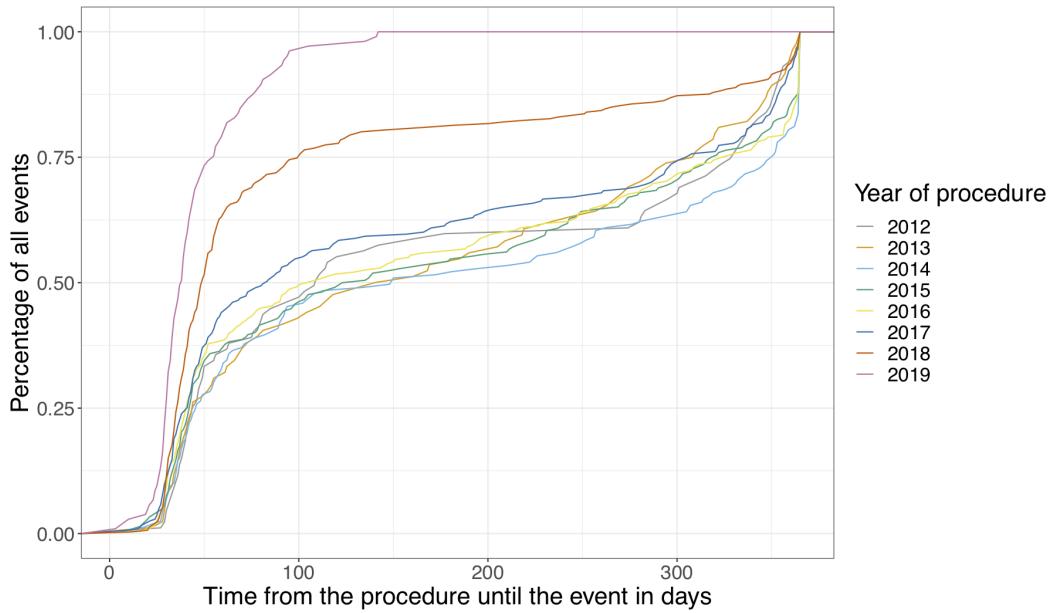
In the cohort used for the analyses (n=1434 patients), 124 patients (8.65%) have died within one year after their TAVI procedure. Figure 4.1b shows, that the event (all-cause mortality) rate decreased over the years. In the first three years of the registry, the percentage of all-cause deaths within one year was at about 14%. From 2015 to 2017, it dropped to about 8.2%, in 2018 it was at just 5.5% and in 2019 at 3.8%. The distribution of the all-cause deaths in relation to time after the procedure can be found in Figure 4.1a. It can be seen that the distribution of the years 2018 and 2019 deviates from the other years. This is mainly due to missing follow-ups for the patients that underwent TAVI in these years which also explains the significant decrease of all-cause mortality in Figure 4.1b. For the other years, there is no visible trend in the data that would suggest longer survival or a shift in the post-procedural risk over the years for the patients that died. This consistency is important for a meaningful regression analysis, because the regression does not take the year of the procedure into account. Overall, half of the events took place roughly 100 days after the procedure, 70 percent after 300 days.

It is desirable for a meaningful regression analysis that the distribution of the underlying parameters does not show great variation over the years or that a change in distribution can be linked to the change in event rate. Therefore, the distribution was

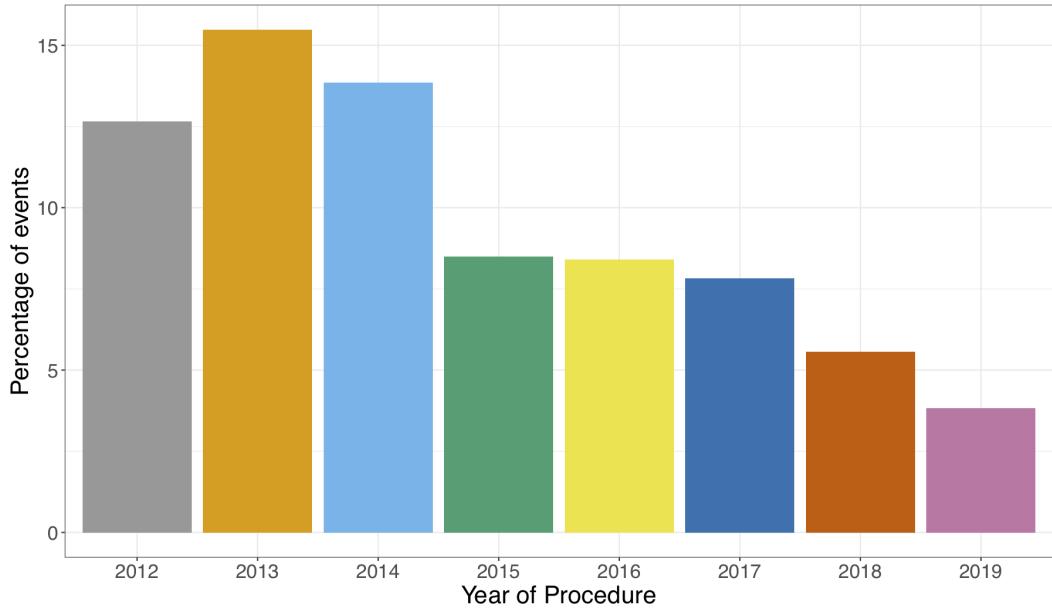
examined and visualized in the Supplementary Figures 5.4 and 5.5. The continuous variables in the baseline table do not possess great variation over the years except for a slight decrease of the mean STS score from 6.7 to 3.5 (see Supplementary Figure 5.4), corresponding to the observed decrease of the event rate. Outliers in continuous data are important, too, as the regression coefficient leads to a strong weighting of the affected variable and thus the hazard assessment of the patient. Apart from that, it would bias the model learning strongly. Outliers must therefore be treated with caution.

Supplementary Figure 5.3 shows some unexpectedly high values (>20) for the STS score and the EuroSCORE II that are most probably wrong because they could not be reconstructed with the given values of the baseline table. Nevertheless, these scores were not discarded since both scores were not incorporated in any model. Apart from that, there are some outliers in the measurements of creatinine ($> 300 \mu\text{mol/L}$ or $<100 \mu\text{mol/L}$), GFR (<20 or >150), thrombocytes ($> 500 \times 10^9/\text{L}$) and creatinine kinase ($>300 \text{ U/L}$) values. Creatinine can rise to over $300 \mu\text{mol/L}$ in patients with renal failure or acute kidney injury [75] and since all of these patients had a renal failure in the baseline table and a correspondingly low GFR, the values were regarded as properly measured. On the contrary, creatinine can also fall below $100 \mu\text{mol/L}$, often in conjunction with low muscle mass [76]. These values were responsible for the exceedingly high GFR values. All of these patients except for one had a BMI of over 37 (116 kg at a height of 176 cm) which supports the assumption of low muscle mass. The other patient had a weight of 70 kg at 180 cm and was 82 years old which could also coincide with the hypothesis. As for the thrombocyte outliers, patients with thrombocytosis can have over $450 \times 10^9/\text{L}$ thrombocytes due to several reasons [77]. However, these reasons could not be verified with the help of the registry variables. Since they could however also not be falsified, the values were kept since they can occur in patients. As there are also reported cases of such high creatinine kinase levels [78] and the creatinine kinase works as an indicator for injuries of the heart muscle or other muscles [27], these values were also kept.

In Table 4.1, the continuous variables are represented by mean \pm standard deviation. Welch's t-test was performed for the STS score, the EuroSCORE II, the GFR, the thrombocyte values and the LVEF values (corresponding p-values of the F-test: 3.81e-09, 1.77e-07, 1.18e-13, 0.04, 0.01). The Benjamini-Hochberg correction produced significant ($\alpha = 0.05$) p-values for the parameters age, EuroSCORE II, STS score, creatinine, hemoglobin, LVEF and gradient mean. Since the t-test compares the mean between two groups and the mean is very sensitive to outliers, one could hypothesize that the significant values for the two scores and the creatinine could be due to the observed outliers. This hypothesis can be rejected by looking at Supplementary Figure 5.3, where one can see that the distribution of the variables truly varies when stratified by event. Supplementary Figure 5.4 reassures that the significance of the variables is constant over the years as the line for the mean in the no-event group rarely intersects with the event-group line and for most of the years, a clear distance between the lines is visible. It suggests that lower age, lower STS and EuroSCORE II, lower creatinine, higher Hb, higher LVEF and higher ΔPm are more favourable for the patient's survival probability. All of these observations are concordant with previous findings [1, 8, 20–28, 37, 39–41, 57, 79].



(a) Empirical Cumulative Density Plot of the time between the procedure and the event for the 124 patients, stratified by year. Except for the years 2018 and 2019, there is not much variation between the distributions. For the years 2012 to 2017, 50 percent of the patients deceased circa 100 days after the TAVI procedure.



(b) Distribution of the event rate over the years. The event rate decreased over the years.

Figure 4.1: Overview of the event distribution in the analysed data set.

The categorical variables in Table 4.1 are represented by percent. Their distribution in the event and the no-event group was compared with Fisher's Exact test. The resulting p-values were adjusted with the Benjamini-Hochberg correction, Supplementary Figure 5.5 shows the distribution of the features over the years stratified by event, Figure 4.2 shows it for the significant variables.

The variables show various stronger changes over the years. There was an augmentation

of the proportion of males (2012: 49.4%, 2019: 61.9%). Even though the p-value was not significant, Figure 4.2 shows that there is a strongly visible distance between the event and the no-event line (except for the year 2017), suggesting a higher event rate for males. Additionally, since 2017, the measurement of troponin T was slowly replaced by measurement of troponin T hs (high sensitive). Both are significant in the multiple testing corrected Fisher's Exact test but Figure 4.2 shows neither is able to produce disparate curves for 2017 and 2018. They were therefore discarded from further analyses.

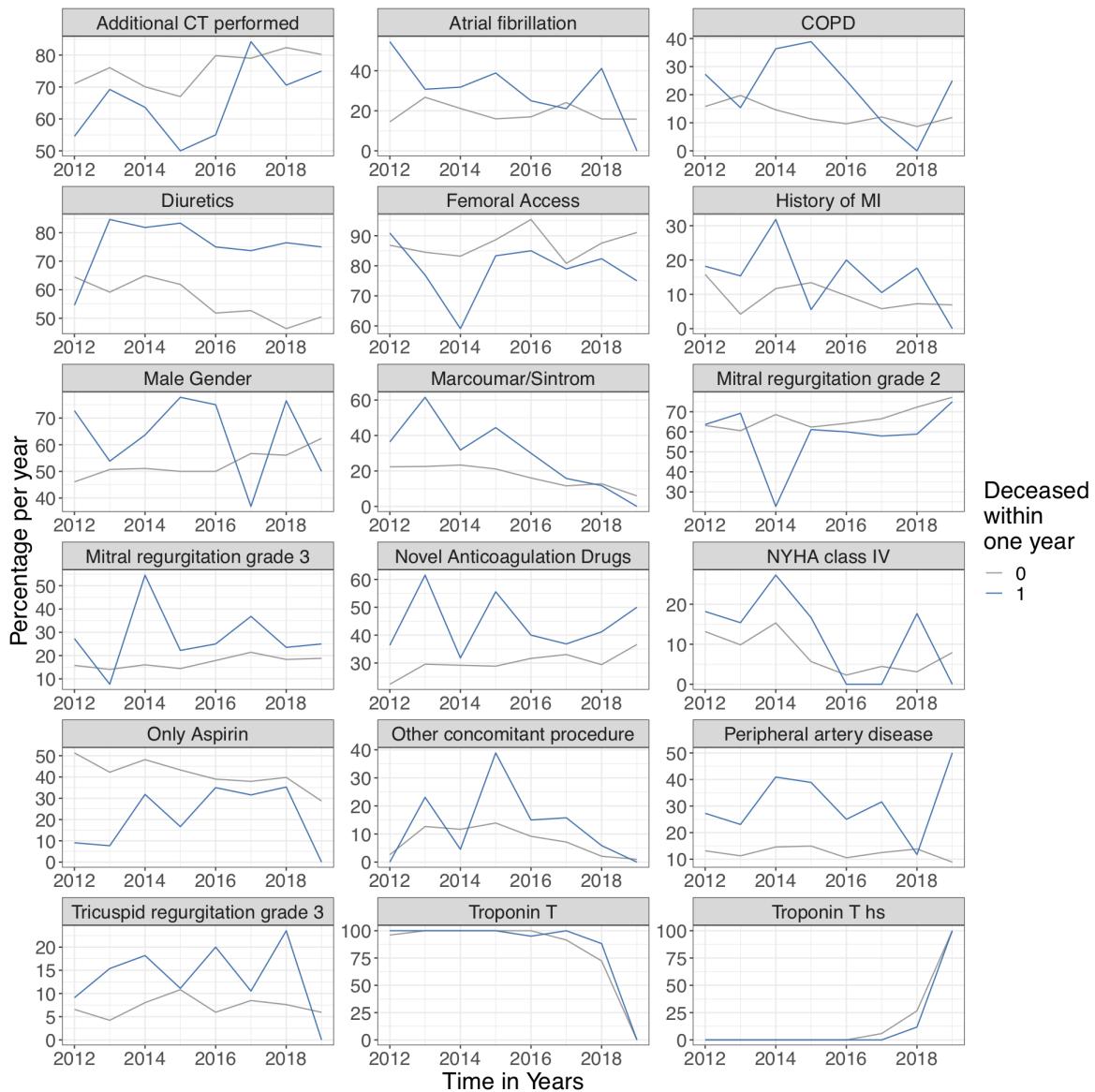


Figure 4.2: Yearly distribution of the boolean-coded variables that had a significant multiple testing corrected p-value.

Regarding procedure-specific parameters, local anesthesia became much more usual (2012: only 1%, 2014: 69%, 2019: 90%). This partially corresponds to the fact that the femoral access became the absolute standard route and for this access only local anesthesia is required. Whereas in 2012 only 88.4% of the procedures were performed

with a femoral access, in 2019 98.4% of the TAVIs were implanted this way. The rest of the observation could be explained by taking into account that the TAVI procedure was relatively new in 2012 and that with more experience more artificial valves were implanted under local anesthesia. Furthermore, the number of concomitant procedures first rose from 9% in 2012 to 24.5% in 2015 and then fell again to 4% in 2019. Table 4.1 shows a significant multiple testing corrected p-value for other concomitant procedures. Thus, it seems possible that this observation motivated the decrease of concomitant procedures. The same can be hypothesized for CT imaging. The number of additionally performed CTs rose from 69% in 2012 to 80% in 2019. An additional CT image helps to assess the heart anatomy even better and allows to place the TAVI valve more accurately. Even though this parameter was not included in the final Cox models since an additional CT image became a clinical standard, this thesis wants to emphasize its significance in prior models, in the multiple testing corrected p-value and in the differences in the yearly distributions (see Figure 4.2). Additionally, it has to be pointed out that even though the p-value of the femoral access route is not smaller than 0.05, it still is at 0.09, separates the curves in Figure 4.2 and was very significant in prior models. Yet, it was not included in the final models presented here for the same reason the CT parameter was not included.

In the subcategory of baseline variables denoting heart-related comorbidities, six features had significant multiple testing corrected p-values: Atrial fibrillation, history of MI, mitral regurgitation grade 2 and 3, tricuspid regurgitation grade 3 and NYHA class IV. Figure 4.2 confirms this as there are rarely intersections of the event line and the no-event line but mainly a visible distance between the lines. Supplementary figure 5.5 shows that there were less assignments of patients to the NYHA categories III or IV (2012: 65.5%, 2019: 42.8%) which corresponds to the decrease in the STS score and the event rate.

For the other comorbidities, the Fisher's Exact test was significant for peripheral artery disease and COPD which is also clearly visible in Figure 4.2. It should additionally be noted, that the rate of COPD patients fell from 17% in 2012 to 8% in 2018 which correlates with the decrease of the event rate. A rather peculiar change in the yearly distribution is the strong, continuous increase of dyslipidemia patients (2012: 24%, 2019: 68%). Intuitively, this should not be part of the cause why the event rate decreased.

As for medication, four patient subgroups were distributed significantly differently in the event group and the no-event group: patients with Marcoumar/ Sintrom medication, Novel Anticoagulation medication or Diuretics and patients who solely took Aspirin as anticoagulants. It is important to add that the last four categories of table 4.1 ('Anticoagulants: Only Aspirin', 'Anticoagulants: Only Novel Anticoagulation drugs', 'Anticoagulants: Aspirin + P2Y12, no novel anticoagulation drugs', 'No anti-coagulants') describe the category of anticoagulants which are not the usual anticoagulants Marcoumar/ Sintrom. As previously described in Section 2.1.2, anticoagulation drugs are i.a. used for treating atrial fibrillation. It is therefore plausible that Marcoumar/ Sintrom medication has a corrected p-value of 0.005, that novel anticoagulation drugs have a corrected p-value of 0.03 and that aspirin has a corrected p-value of 0.007 when atrial fibrillation had a corrected p-value of 0.007. More precisely, 120/248 Marcoumar/ Sintrom patients and 215/452 Novel Anticoagulation patients had atrial

fibrillation (overall 283 patients). The aspirin medication served another purpose (only 30/564 of these patients had atrial fibrillation). It is used as preventative treatment for peripheral artery disease, strokes, myocardial infarctions or after cardiac surgeries [28]. It should additionally be noted that there was a decrease in Marcoumar/ Sintrom medication over the years (2012: 24.1%, 2019: 6%) while there was an increase in novel anticoagulation drugs (2012: 24%, 2019: 37%). As stated earlier, diuretics are used for treating patients with arterial hypertension, heart failure and renal failure. Since arterial hypertension does not have a significant p-value (0.679), it is more likely that the patients who contributed to the significance of the diuretics value experienced either heart failure or renal failure. This hypothesis is supported by the significance of the variables creatinine (0.007; indicating renal failure) and NYHA class IV (0.034; indicating heart failure). Indeed, 618 of 811 diuretics patients had a GFR under 60 (normal renal function: GFR > 90 [44]) and 469 had a NYHA class III or IV. It should be added that there were less patients with diuretic medication over the years (2012: 63%, 2016: 54%, 2019: 51%). Furthermore, there was a rise in statin medication (2012: 46%, 2019: 60%) which could i.a. be the result of the increase of dyslipidemia patients since 61% of the them took statin.

Summary

The all-cause mortality rate decreased over the years (2012-2014: 14%, 2015-2017: 8%, 2018: 5.5%, 2019: 3.8%). For the last two years, the strong decrease is probably due to missing follow-ups. Basically, the finding however corresponds to a lower mean STS score, more CTs and femoral access routes, less patients with NYHA III or IV and less COPD patients over time.

There are 7 continuous variables that were significant in a multiple testing corrected t-test: age, EuroSCORE II, STS score, creatinine, Hb, LVEF and mean gradient (ΔP_m). Both surgery risk scores will not be included in the Cox models since the scores can be computed with the other provided baseline variables, leading to multicollinearity in the regression models.

As for the categorial variables, the multiple testing corrected Fisher's Exact test found 16 significant features: Troponin T, Troponin T hs, atrial fibrillation, history of MI, mitral regurgitation grade 2, mitral regurgitation grade 3, tricuspid regurgitation grade 3, NYHA class IV, peripheral artery disease, COPD, other concomitant procedure, additional CT performed, Marcoumar/ Sintrom, Novel Anticoagulation drugs, Diuretics and the use of solely Aspirin as anticoagulation drug. The CT parameter as well as the femoral access route parameter were not included in the Cox models since they are a clinical standard now but their significance in the statistical testings and in prior Cox models should be emphasized.

In order to obtain a meaningful regression analysis, it was important to look at outliers and at distribution changes. Distribution changes correlating with the event rate are important since it has to be discussed if the distribution changes are causing the change in the event rate. Outliers and distribution changes not relating to the event rate could bias the model learning. There are some outliers for the STS score and the EuroScore II that are most probably wrong but they were not discarded because the scores were not included in the models. The other outliers were kept as well, considering that

they can occur in patients and the values could not be proven wrong with the help of other baseline variables. Concerning distribution changes, there are more male patients throughout the years but the data suggests a higher event rate for males which is why this change should be classified as a simple negative correlation that does not cause the decrease of the event rate. Since the troponin measurement changed from troponin T to troponin T hs and there is no visible significant difference in the distribution of troponin in the event and the no-event population, the significance of troponin in the Fisher's exact test is probably wrong.

Lastly, it should be noted that the medication variables do not correlate with just one disease. It is therefore complicated to interpret their value in a Cox model because the significance could either be due to the treated diseases or to the medication itself.

4.2 Scores

After omitting the values for the STS score, the EuroSCORE II, the additional CT imaging and the access site, several models were computed with different approaches. The first model consists of the variables that were significant in the Benjamini-Hochberg multiple testing corrected statistical tests from Table 4.1. Both troponin variables were discarded for the reasons mentioned above. Additionally, the variable indicating a mitral regurgitation of grade 2 was omitted because it correlates negatively with the variable indicating a mitral regurgitation of grade 3. High correlation between variables in a regression model distorts the results for the coefficients on the one hand. On the other hand, it would be redundant to include a variable that can only be true when the other one is false. The variable for grade 3 was chosen because its p-value was smaller than the one for grade 2 and it intuitively made sense to include a more extreme parameter. The remaining 17 features were used for the 'Benjamini-Hochberg Model'. The features and the statistical measures for the whole model can be seen in Table 4.2. The exponentiated coefficients as well as the standard error, the Wald statistic and the herefrom derived p-value and 95% CI of the coefficients are displayed in Supplementary Table 5.1.

The second model was computed with the approach of selecting the features with the help of LASSO. All variables were put into the *cv.glmnet* function of the *glmnet* package, the resulting 17 variables (see Table 4.2: 'LASSO-on-all Model') were the variables with non-zero coefficients at the value of λ that gave the minimum mean cross-validated error. For the exponentiated coefficients and their statistics, see Supplementary Table 5.2.

Since we wanted to reduce the number of model parameters while preserving a C-index of about 0.74 and preferably minimizing the p-values as well, all 2^{17} possible combinations of the LASSO-selected variables were computed. Cox models were learned for all combinations, the C-index and the average coefficient p-value was extracted. All of the 2^{17} models with a C-index smaller than 0.74 were discarded, the rest was sorted by the average p-value of the Wald statistic of their coefficients. The model with the smallest average p-value (0.052) had 11 variables and a C-index of 0.741 and is shown as 'LASSO-on-all: Smaller Model' (third model) in Table 4.2 (coefficient statistics: Table 4.3).

Model Name	Benjamini-Hochberg Model	LASSO-on-all Model	LASSO-on-all: Smaller Model	LASSO-on-intermediate/low-risk Model	Random Forest Model
Features	1. Age 2. Creatinine 3. Hb 4. LVEF 5. ΔPm 6. Atrial fibrillation 7. History of MI 8. Mitral regurg. grade 3 9. Tricuspid regurg. grade 3 10. NYHA IV 11. Peripheral artery disease 12. COPD 13. Other concomitant procedure 14. Marcoumar/ Sintrom 15. Novel Anticoagulation drugs 16. Diuretics 17. Only Aspirin as anticoag.	1. Age 2. Gender 3. Hb 4. LVEF 5. ΔPm 6. Atrial fibrillation 7. Previous BAV 8. Mitral regurg. grade 3 9. Tricuspid regurg. grade 3 10. Atrio-ventricular block 11. Peripheral artery disease 12. COPD 13. Statin medication 14. Marcoumar/ Sintrom 15. Insulin medication 16. Diuretics 17. Only Aspirin as anticoag.	1. Age 2. Gender 3. Hb 4. ΔPm 5. Mitral regurg. grade 3 6. Atrio-ventricular block 7. Peripheral artery disease 8. COPD 9. Statin medication 10. Diuretics 11. Only Aspirin as anticoag.	1. Gender 2. Hb 3. CCS class 1 4. Tricuspid regurg. grade 3 5. Peripheral artery disease 6. COPD 7. Diuretics	1. Age 2. Gender 3. Creatinine 4. Hb 5. LVEF 6. ΔPm 7. Atrial fibrillation 8. Creatinine Kinase 9. Mitral regurg. grade 3 10. Peripheral artery disease 11. COPD 12. Diuretics 13. Steroid medication
C-index	0.727	0.742	0.741	0.727	0.731
10-fold CV	0.706	0.74	0.731	0.693	0.724
Mean permut. test C-index	0.604	0.601	0.581	0.574	0.588
Wald test	78.5, p-value 7e-10	95.59, p-value 6e-13	85.17, p-value 1e-13	50.92, p-value 1e-08	79.52, p-value 1e-11
LR test	76.2, p-value 2e-09	94.62, p-value 9e-13	86.13, p-value 1e-13	50.71, p-value 1e-08	80.15, p-value 1e-11
AIC	1504	1485.9	1482.1	1097.7	1492.1
BIC	1552	1533.5	1513.1	1115.5	1528.7

Table 4.2: Overview of the developed Cox models. For each model, the concordance index, the 10-fold cross-validated (CV) mean iAUC, the mean concordance index of 100 permutation tests each, the Wald test statistic and its p-value, the Likelihood ratio (LR) test and its p-value and Akaike's and Bayesian Information Criterion (AIC and BIC), are provided. Overall, the 'LASSO-on-all: Smaller model' performed best.

The fourth model was computed only on the population that had a STS score of 8 or lower, classifying the patients as intermediate and low surgery risk patients. This was done because most existing TAVI risk scores were computed on a high or extreme surgery risk population, making it questionable whether or not the score variables are applicable for this population as well (see Section 2.3.2 for more details). Again, the LASSO method was used on all available variables and the λ value with the minimum mean cross-validated error yielded the 7 features used in the ‘LASSO-on-intermediate/low-risk Model’ (see Table 4.2 and Supplementary Table 5.3).

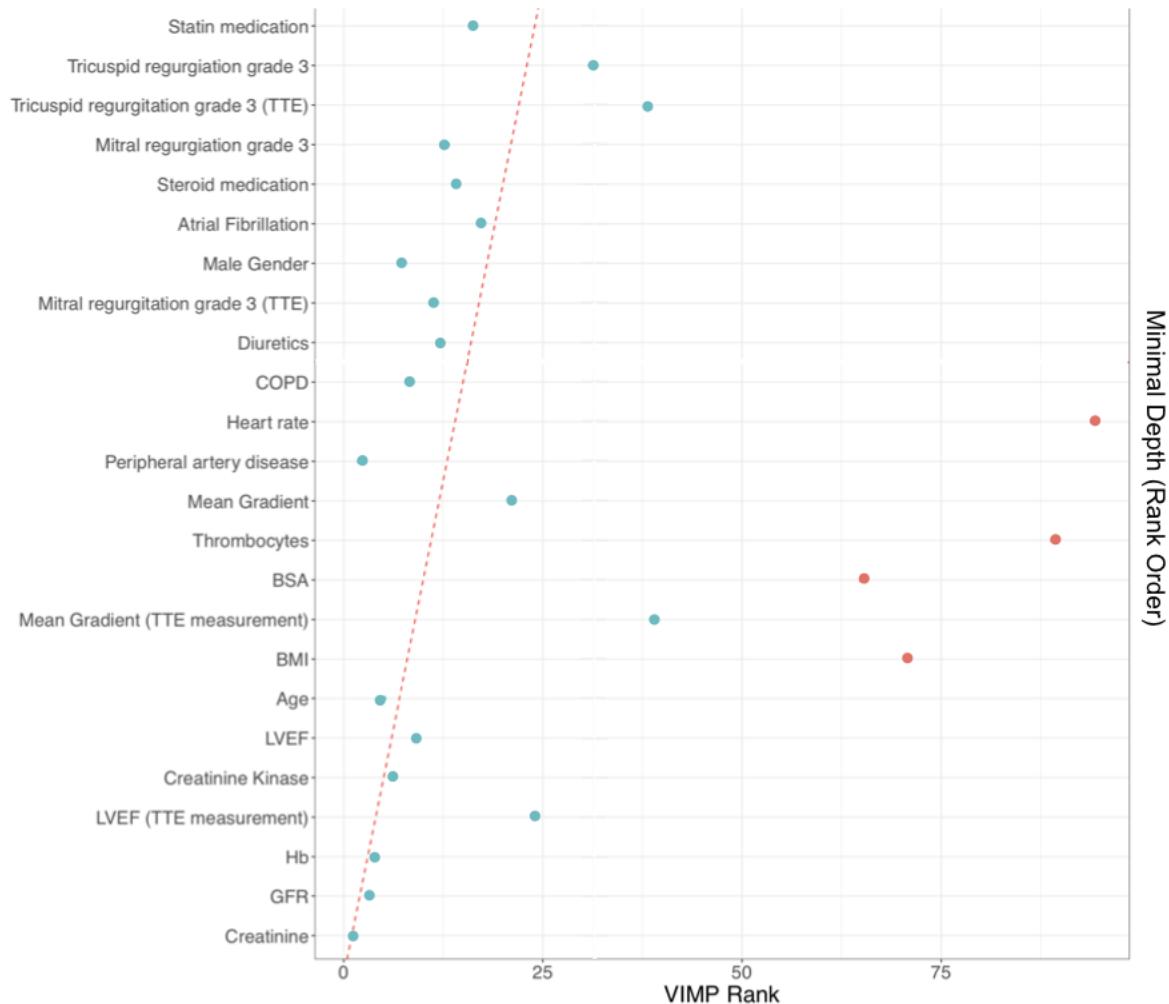


Figure 4.3: Comparison of the variable importance (VIMP) versus the minimal depth measure. The closer the points are to the diagonal line, the more are the two measures in agreement.

The fifth and last model was computed by first learning a random survival forest with 1000 trees (see the Out-of-Bag error rate in Supplementary Figure 5.6), determining variable importance (VIMP) and minimal depth quality measures of the variables and then picking the top features that were best in both kind of measures. The comparison of these two measures can be seen in Figure 4.3. GFR was not chosen because it highly correlated with creatinine. The TTE measurements were not chosen because firstly, they are part of the ‘normal’ parameter together with the angiography and TEE measurement (e.g. LVEF is composed of the TTE measurement, the TEE mea-

surements and the angiography measurements). Secondly, the VIMP and the minimal depth measurement were more in agreement regarding the ‘normal’ parameters. BMI did not have a high VIMP rank but only a high minimal depth rank which is why it was not included as well. The same holds true for BSA, thrombocytes and heart rate. Mitral regurgitation grade 3 was the last variable included in the model. With these 13 features, the Cox model ‘Random Forest Model’ (see Table 4.2) was calculated. The exponentiated coefficients and their statistics are shown in Supplementary table 5.4.

4.2.1 Comparison of the scores

Analyzing the C-indices of all five models, it becomes apparent that the ‘LASSO-on-all Model’ has the best C-index with 0.742, followed by the ‘LASSO-on-all: Smaller Model’ with 0.741. The ‘Random Forest Model’ comes next with 0.731 and the ‘Benjamini-Hochberg Model’ as well as the ‘LASSO-on-intermediate/low-risk Model’ had a C-index of 0.727.

Since the population was not equally balanced regarding events, the 10-fold cross-validation was done with test sets consisting of one tenth of the event population and one tenth of the no-event population in each of the folds. The value in Table 4.2 is the mean of the iAUC from the fold test sets (package *survAUC*). Since the iAUC and the C-index both describe the model fit on the same scale with similar measures, they are perfectly comparable. Therefore, the ‘LASSO-on-all Model’ is the most stable model with just a decrease of 0.002 from a C-index of 0.742 to a mean iAUC of 0.74. The ‘LASSO-on-all: Smaller Model’ has the second highest cross-validated iAUC (0.731, decrease of 0.01). The ‘Random Forest Model’ has the second lowest decrease from the C-index to the mean iAUC (only 0.007), the ‘Benjamini-Hochberg Model’ loses 0.021 and the ‘LASSO-on-intermediate/low-risk Model’ had the worst mean iAUC (0.639). This is probably due to the low number of variables. With less variables, there are less possibilities to explain the events and therefore on the one hand less overfitting but on the other hand the risk of worse performance.

Figure 4.4 shows the result of the 100 permutation tests for each model. For all five models, the concordance and the cross-validated iAUC are significantly better than the permutation test’s concordances. We can therefore be sure that the parameters from the developed Cox models indeed explain the relationship between the event rate and the parameters. The figure also shows that a higher number of variables in the Cox model leads to higher concordances in the permutation tests. This is due to the effect that the higher the number of variables, the more variation in the data can be explained regardless of the true relationship between the event rate and the baseline variables.

Comparing the cross-validated results to the validated C-indices from the existing scores (see Table 2.3), we can see that even the ‘LASSO-on-intermediate/low-risk Model’ would still have the third best 1-year all-cause mortality prediction performance. We can additionally see that our random permutation tests nearly achieved a similar performance to the Predictor of Poor Outcomes score for 6 months: This score had a concordance of 0.64 with 10 variables which is almost within the range of the upper whisker of the boxplot for the ‘LASSO-on-intermediate/low risk Model’ (11 variables).

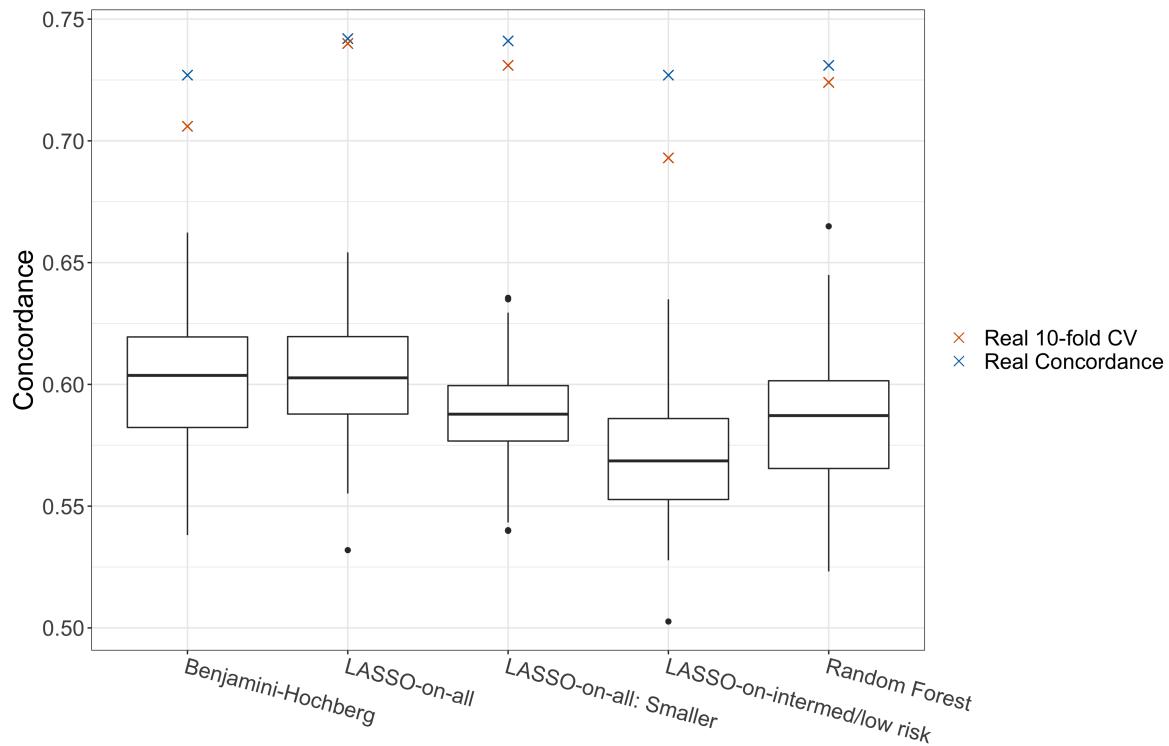


Figure 4.4: Comparison of the concordances of the 100 permutation tests and the real C-index/cross-validated iAUCs. All real concordances and cross-validated iAUCs are visibly significantly better than the random permutation tests. The higher the number of variables of a model, the higher the C-indices of the permutation tests.

If we look at all published scores except for NIS TAVR which predicted in-hospital mortality with post-procedural parameters (for more see Section 2.3.2), the ‘LASSO-on-all Model’, the ‘LASSO-on-all: Smaller Model’ and the ‘Random Forest Model’ are only outperformed by the CoreValve scores (Validated C-index for 1-year all-cause mortality: 0.79, for 30-day all-cause mortality: 0.75). As mentioned in Section 2.3.2, the variables of the CoreValve scores are mainly frailty related (home oxygen use, assisted living). Furthermore, a decreased albumin level can be observed for many forms of illness including liver failure, infections, hepatitis, diabetes, cancer and other chronic diseases. Therefore, the strong prediction power may be an effect of the accumulation of these diseases. Additionally, the Charlson Comorbidity Index and the STS score used in the 1-year mortality score consist of 18 and over 150 variables respectively, making the model prone to overfitting and hard to interpret. In contrast, the ‘LASSO-on-all Model’, the ‘LASSO-on-all: Smaller Model’ and the ‘Random Forest Model’ do not contain scores. Only the medication variables are rather harder to interpret since they can treat multiple underlying diseases.

In order to break these underlying correlations to diseases down, Pearson’s correlation coefficient was computed for the medication variables from the Cox models to all other used variables of the baseline table. Figure 4.5 shows all variables that either showed an absolute correlation bigger than 0.2 to the medication variables or that were mentioned before as indications for the drug treatment. Aspirin medication was added in order to show the differences to ‘Only Aspirin as anticoagulation drug’.

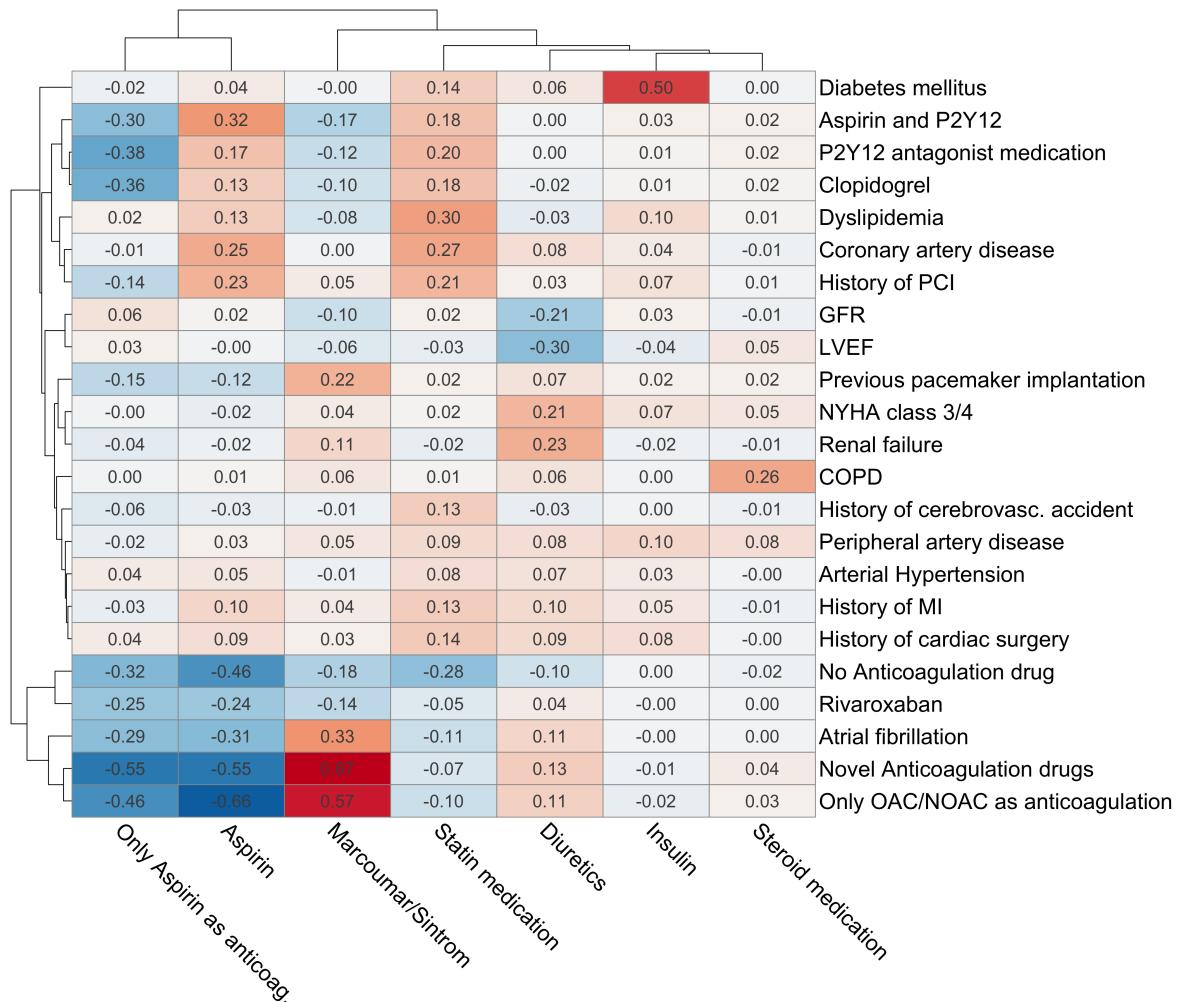


Figure 4.5: Pearson correlation coefficients for the medication variables included in the Cox models and Aspirin for comparison with ‘Only Aspirin as anticoag.’.

Steroid medication clearly has the biggest correlation to COPD which is conform to the known application of steroids (see section 2.1.2).

Insulin is strongly positively correlated to diabetes mellitus which also makes sense intuitively.

In the previous chapter, it was hypothesized that diuretics medication is mainly linked to renal failure and heart failure. This is supported by the correlation coefficients as diuretics are positively correlated with renal failure and NYHA class 3/4 and negatively correlated with a high LVEF and a high GFR. There are no other strong correlations for diuretics. In particular, there is no strong correlation to arterial hypertension which can also be treated with diuretics.

The statin medication correlates positively with dyslipidemia, coronary artery disease, history of PCI and P2Y12 antagonist medication. This corresponds to the fact that statin is used in patients with dyslipidemia or coronary artery disease (see Section 2.1.2) and that coronary artery disease patients are often treated with a PCI [28]. As for P2Y12 medication and Aspirin + P2Y12 anticoagulation therapy, it has been suggested that statin and these medications reduce the probability for postoperative cardiac events and increases the chances of survival in combination [80].

Marcoumar and Sintrom correlate strongly positively with novel anticoagulation drugs (OAC/NOAC), OAC/NOAC as alternative anticoagulation treatment, atrial fibrillation and previous pacemaker implantations. It has been described in Section 2.1.2 that Marcoumar and Sintrom as well as other anticoagulation drugs are used to treat atrial fibrillation. Since the correlations are positive, it seems likely that Marcoumar or Sintrom are often used in combination with OAC/NOAC drugs but not with P2Y12, Aspirin + P2Y12, Rivaroxaban or Clopidogrel since there are negative correlations.

‘Only Aspirin as anticoagulation drug’ is solely negatively correlated with the depicted variables, particularly with the other alternative anticoagulation treatments and atrial fibrillation. The former is plausible because the categories are mutually exclusive. The latter was already mentioned in the previous section but we cannot see a strong connection to peripheral artery disease, strokes, myocardial infarctions or cardiac surgeries as hypothesized. What we can see is a positive correlation for Aspirin and coronary artery disease and history of PCI but this correlation does not exist for ‘only Aspirin as anticoagulation drug’.

Summing up, the steroid medication parameter can only indicate steroid-treated COPD and the insulin parameter can only indicate insulin-dependent diabetes mellitus. Diuretics function as a indicator for renal or heart failure, statin either implies statin-treated dyslipidemia, stain-treated coronary artery disease or a combined statin and P2Y12 or Aspirin + P2Y12 treatment, respectively. A positive Marcoumar and Sintrom medication suggests atrial fibrillation or a previous pacemaker implantation. The residual impact of the parameter and the ‘Only Aspirin as anticoagulation drug’ parameter can be attributed to the medication itself. Of course, this is also possible for the other medication treatments but since most of the underlying diseases were also significant in the conducted statistical tests, this does not seem too probable.

If we continue to compare the five models from Table 4.2, we can see that the Wald statistic p-values are the best for the ‘LASSO-on-all: Smaller Model’, closely followed by the ‘LASSO-on-all Model’ (1e-13 and 6e-13). The same holds true for the p-values of the Likelihood ratio test (1e-13 and 9e-13) and the Akaike’s Information Criterion (the smaller the better but only comparable for the same population size, therefore not for the ‘LASSO-on-intermediate/low-risk Model’). Bayes Information Criterion also ranks the ‘LASSO-on-all: Smaller Model’ but the ‘Random Forest Model’ second instead of the ‘LASSO-on-all Model’. Therefore, only the ‘LASSO-on-all: Smaller Model’ will be discussed in full detail in Section 4.2.2.

4.2.2 LASSO-on-all: Smaller Model

The *coxph* function provided centered linear predictors for each patient, representing the predicted individual risk. The predictors for the ‘LASSO-on-all: Smaller Model’ ranged from 2.77 (maximal hazard of experiencing the event) to -2.8 (minimal hazard), zero represents a neutral risk classification. The distribution of the predictors is displayed in Supplementary Figure 5.7. With the help of these linear predictors, Kaplan-Meier curves were calculated for different risk groups defined by the predictors. On the one hand, these curves served as a sanity check for the success of the regression, on the other hand they should help to classify a new patient into a risk group. Patients with a centered linear predictor ≥ 0.5 were classified as high hazard

patients, patients with a centered linear predictor ≤ -0.5 as low hazard patients. The rest was regarded as intermediate hazard patients. The Kaplan-Meier curves are displayed in Figure 4.6. The curves are clearly distinct from each other, the log-rank test (calculated with `survdiff` from the `survival` package) has a p-value of 5e-15. For comparison, the process was repeated for the STS score (high hazard: >8 , low hazard, ≤ 4) and the EuroSCORE II (same classification). The Kaplan-Meier curves can be seen in Supplementary Figures 5.8 and 5.9. The curves are not as distinct as the curves in Figure 4.6, the log-rank test yielded p-values of 1e-05 for the STS score and 2e-06 for the EuroSCORE II which is still very significant but not as good as the results for the presented Cox model. Especially the intermediate risk patients are not really distinguishable from the low risk patients in both scores.

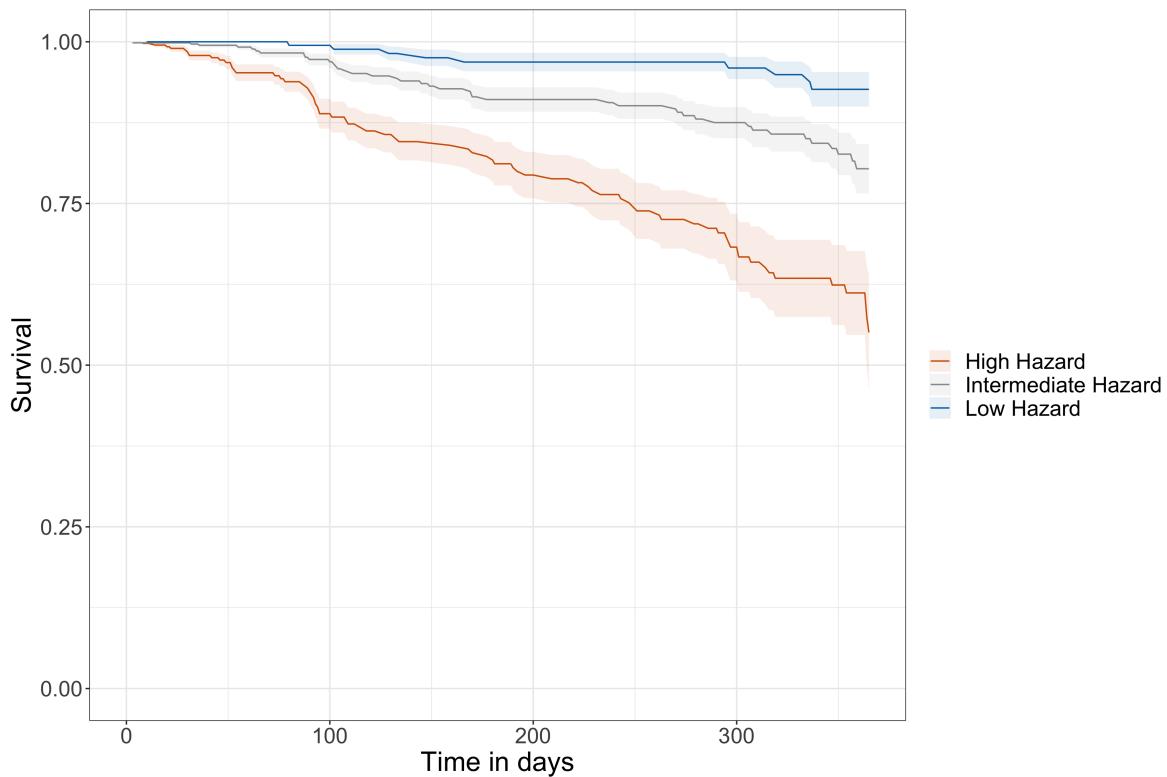


Figure 4.6: Kaplan-Meier curves for the ‘LASSO-on-all: Smaller Model’ - survival predictions with the associated standard error. The log-rank test yielded a p-value of 5e-15. High hazard patients had centered linear predictors ≥ 0.5 , low hazard patients had centered linear predictors ≤ 0.5 .

Furthermore, Pearson’s correlation coefficient was computed pairwise for all the variables in the Cox model in order to make sure that the features are not redundant (see Figure 4.7). There was no strong correlation (> 0.5 or < -0.5) between the variables, the highest absolute value was 0.19 between gender and Hb and mitral regurgitation grade 3 and diuretics, respectively. The former makes sense because Hb is indeed gender specific, men normally have a higher Hb than women [27]. The latter is also not surprising because, as described in the section above, diuretics medication is indicated by heart problems which a mitral regurgitation grade of 3 certainly is. Nevertheless, the risks linked to low Hb levels cannot be represented by gender and vice versa. This is also true for the diuretics variable which represents the risk of renal and heart prob-

lems, simultaneously. Therefore, it cannot be replaced by an indicator of solely mitral regurgitation of grade 3. Just as well, mitral regurgitation of grade 3 may not always be treated with diuretics but can still be an important risk factor.

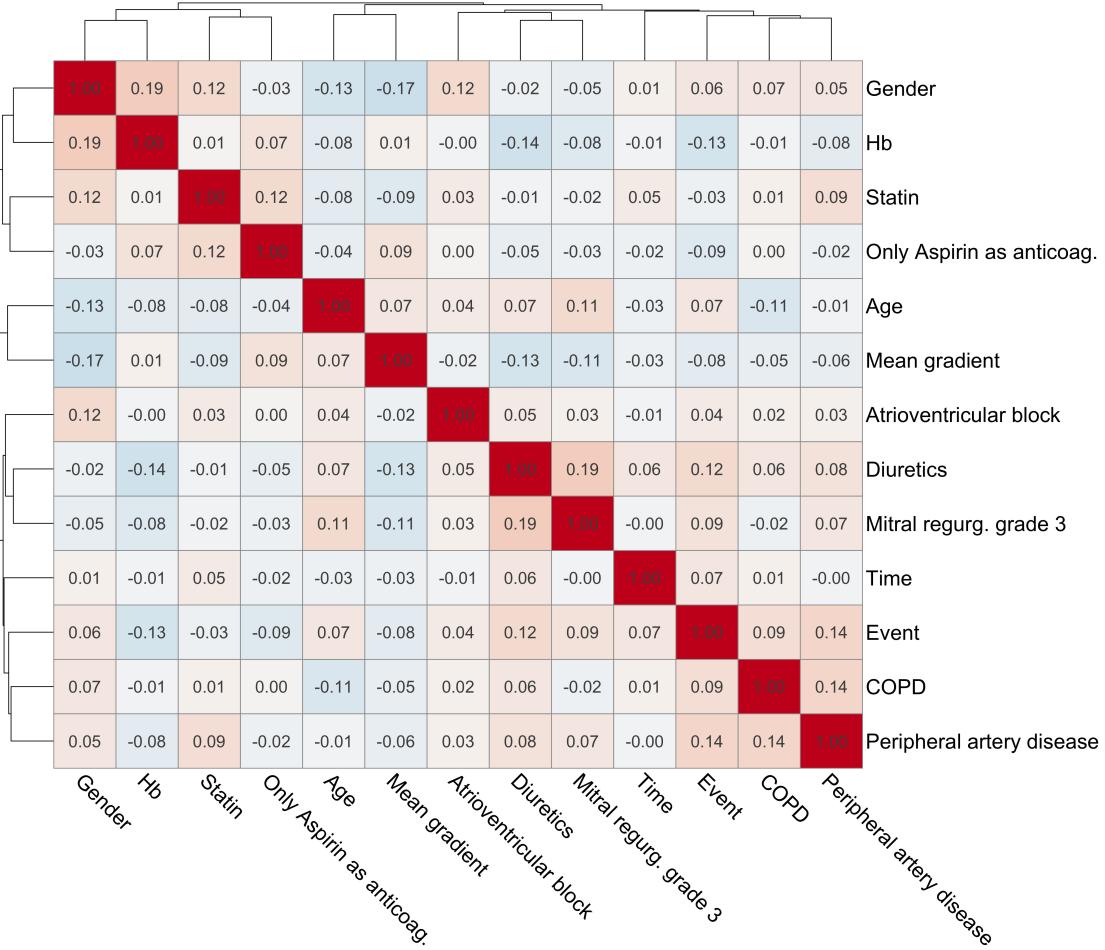


Figure 4.7: Pearson's correlation coefficients for the 'LASSO-on-all: Smaller Model'. There are no strong correlations between the variables. The highest correlation coefficients are 0.19 for Gender and Hb and for mitral regurgitation grade 3 and diuretics, respectively.

Table 4.3 displays the exponentiated coefficients, the standard error of the coefficients, their Wald statistic, its p-value and the 95% CI for the exponentiated coefficient. From the exponentiated coefficients, it can be concluded that lower age, female gender, higher Hb and higher ΔPm have a positive influence on the outcome of the TAVI patients. Also, the absence of mitral regurgitation of grade 3 as well as the absence of atrioventricular block, peripheral artery disease, COPD and diuretic medication and the presence of statin medication and only Aspirin as anticoagulation drug have a positive influence on the survival. Hb and peripheral artery disease had the smallest p-value, corresponding to the fact that they correlate the most with the event parameter (see Figure 4.7: 0.14 for peripheral artery disease, -0.13 for Hb).

If we recall the Cox regression equation

$$h(t) = h_0(t) \cdot \exp(\beta_1 x_1 + \dots + \beta_k x_k) = h_0(t) \cdot \exp(\beta_1 x_1) \cdot \dots \cdot \exp(\beta_k x_k) \quad (4.1)$$

and let x_1 be a continuous parameter like age, Hb or ΔPm , we can see that an increment of x_1 by 1 raises the hazard by $\exp(\beta_1)$ and therefore $(1 - \exp(\beta_1)) \cdot 100\%$:

$$h(t) = h_0(t) \cdot \exp(\beta_1(x_1 + 1)) \cdot \dots \cdot \exp(\beta_k x_k) = \exp(\beta_1 x_1) \exp(\beta_1) \cdot \dots \cdot \exp(\beta_k x_k) \quad (4.2)$$

If x_1 is a boolean parameter like mitral regurgitation grade 3, atrio-ventricular block, peripheral artery disease, COPD, statin medication, diuretics or ‘only Aspirin as anti-coagulation drug’, the hazard will be raised by $(1 - \exp(\beta_1)) \cdot 100\%$ when the boolean parameter is 1:

$$h(t) = h_0(t) \cdot \exp(\beta_1 \cdot 1) \cdot \dots \cdot \exp(\beta_k x_k) \quad (4.3)$$

The parameters of the Cox model and their credibility will be discussed below one by one.

	Exp (Coef-ficient)	standard error	Wald statistic of the co-efficient	p-value of the Wald statistic	95% CI
Age	1.04	0.014	2.86	0.0043	1.01-1.07
Male Gender	1.82	0.197	3.05	0.0023	1.24-2.68
Hb	0.98	0.005	-3.72	0.0002	0.97-0.99
ΔPm	0.99	0.006	-1.37	0.1711	0.98-1.0
Mitral regurg. grade 3	1.49	0.204	1.96	0.0504	1.0-2.22
Atrio-ventricular block	1.43	0.265	1.35	0.1769	0.85-2.4
Peripheral artery disease	2.15	0.204	3.75	0.0002	1.44-3.21
COPD	1.81	0.225	2.65	0.008	1.17-2.82
Statin medication	0.66	0.191	-2.19	0.0282	0.45-0.96
Diuretics	1.61	0.222	2.16	0.0307	1.05-2.49
Only Aspirin as anticoag.	0.7	0.215	-1.65	0.0996	0.46-1.07

Table 4.3: Cox model for the ‘LASSO-on-all: Smaller Model’. Seven of the eleven variables have a p-value smaller than 0.05. From the exponentiated coefficients, it can be concluded that lower age, female gender, higher Hb and higher ΔPm have a positive influence on the outcome of the TAVI patients. Also, the absence of mitral regurgitation of grade 3 as well as the absence of atrio-ventricular block, peripheral artery disease, COPD and diuretic medication and the presence of statin medication and only Aspirin as anticoagulation drug have a positive influence on the survival.

Age

The exponentiated coefficient of the age feature tells us that with an increase of age by 1, the hazard for the patient rises by 4% when all other parameters stay equal. Because of equation 4.2, an age difference of 10 years between two patients A and B, where A is 10 years older than patient B and all other parameters are equal, would lead to a $h_A(t)$ that is 1.04^{10} times higher (48%) than $h_B(t)$. While this seems probable

for patients between 70 and 100, which most of the TAVI patients are, it can certainly not be extrapolated to younger patients. Nevertheless, nothing could be adapted here, because the Cox model is a linear model. The Wald statistic (coefficient/ standard error, see Section 2.2.1 for more details) yields a very significant p-value of 0.0043, 1.0 is not part of the small 95% CI. The significant p-value from the t-test (0.027) and Supplementary Figure 5.4 support the significance of the age variable. Additionally, age was part of the TAVI₂, the C₄CAPRI, the UK-TAVI, the STS/ACC TVT Registry, the TARIS and the FRANCE-2 score and mentioned in several TAVI mortality biomarker studies [1, 40]. All this evidence and statistical results lead to the conclusion that elevated age is indeed a significant parameter for TAVI mortality prediction. Part of the reason for this is probably that with elevated age the survival probability generally decreases while degenerative processes increase, making it more difficult to cope with the stress of such a procedure.

Male Gender

Equation 4.3 lets us infer that males have a 82% higher risk of experiencing the event than females, the Wald statistic p-value is very significant (0.0023). The 95% CI ranges from 1.24-2.68, signifying that the exponentiated coefficient is in this interval with a probability of 95%. Since this interval is very broad (24%-168% higher risk of dying), we can only safely state that, according to the results, the survival probability of males is definitely smaller than for females. The curves from Figure 4.2 suggest a high variance in this probability as well: While in 2012, 2015, 2016 and 2018 over 70% of the event group were male while overall about 53% of the population of those years were male, in 2013 these proportions were 54% versus 51% and in 2017 and 2019 even 37% and 50% versus 66% and 62%.

Comparing the result with the findings from the existing scores, gender appears as a predictor in the TAVI₂, C₄CAPRI, UK-TAVI, Predictor of Poor Outcomes and TARIS score. All of these studies except for UK-TAVI linked male gender to a higher hazard, TAVI₂ even had very similar results (exponentiated coefficient 1.81, 95% CI: 1.14 - 2.87). The other studies reported lower hazards (C₄CAPRI: 1.05, Predictor of Poor Outcomes: 1.23 (6 months), 1.22 (1 year), TARIS: 1.11). Together with the fact that gender was part of 4 of the 5 constructed models and only not included in the 'Benjamini-Hochberg Model' because it had a p-value of 0.065, it seems like male gender truly is an important predictor for survival after TAVI.

The reason for this is probably partially due to the fact that males generally have a lower life expectancy than females, including in the countries in which the scores were developed [81]. It has also been suggested that there is a relation between better survival for females after TAVI and LVEF, namely that females with low LVEF survive longer than males [82]. Another study found that not LVEF but the presence of coronary artery disease lead to significant gender-related survival differences [83]. Moreover, it was proposed that the higher survival rate of females can be linked to a faster recovery of the LVEF in women [84] and that males more often have comorbidities like coronary and peripheral artery disease, have myocardial infarctions or previous cardiac surgeries and experience vascular complications after TAVI [85].

Hemoglobin

Lower hemoglobin levels are very significantly (p-value of the Wald statistic: 0.0002) associated with worse outcome. Hemoglobin was part of all five developed models and very significant in the t-test (p-value <0.001). This is in agreement with the findings of the TAVI₂, the TARIS score and Takagi *et al.*, who linked low hemoglobin levels to increased mortality after TAVI in a meta-analysis [45] (see Section 2.1.2). As reasons for this observation, they hypothesized that patients with low hemoglobin levels and anemia more often require blood transfusions and that in older patients anemia is often induced by nutritional deficiency or chronic and inflammatory diseases, increasing the probability of peri- and post-procedural problems [45]. Therefore, we will regard hemoglobin as a powerful predictor for 1-year all-cause mortality in TAVI patients.

Mean transvalvular pressure gradient

The p-value of the Wald statistic of ΔP_m is not significant (0.1711). A removal of the variable from the model did not decrease the C-index and its cross-validation by much (C-index 0.737 vs. 0.741, 10-fold CV: 0.729 vs. 0.731). The AIC and BIC were decreased to 1482 and 1510.2, respectively, making it a better fitting model. Nevertheless, the value was kept for several reasons. Firstly, we did not want the concordance index to drop below 0.74 when decreasing the number of coefficients for the 'LASSO-on-all Model'. Secondly, ΔP_m had a significant p-value in the t-test and was informative in the random forest in the VIMP as well as the minimal depth quality measure. Thirdly, ΔP_m was part of the TAVI₂, the C₄CAPRI, the Predictor of Poor Outcomes and the TARIS score. Lastly, in literature, the outcome of high (>40 mmHg) - versus low-gradient severe aortic stenosis patients after either SAVR or TAVI is broadly discussed, especially in combination with high and low LVEF [1, 79, 86]. Higher ΔP_m was generally associated with a better chance of survival, which corresponds to our exponentiated coefficient (0.99) and the exponentiated coefficients of C₄CAPRI (0.99, 95% CI 0.98-1.0), Predictor of Poor Outcomes (per 10 mmHg: 0.82, 95% CI 0.75-0.89 (6 months), 0.84, 95% CI 0.77-0.9 (1 year)), TARIS (0.99, 95% CI 0.98-1.0) and TAVI₂ (≥ 70 mmHg: 2.46, 95% CI 1.33-4.56). As we can see in the scores, the results become more significant when stratifying ΔP_m or even dichotomizing it. Since this is a bad practice for regression models though and a lot of information is lost this way, this was not done here.

Mitral regurgitation grade 3

Mitral regurgitation grade 3 also has a small p-value (0.0504). The exponentiated coefficient indicates that the risk for such a patient is elevated by 49%. The confidence interval is very broad, suggesting that the risk for a patient with mitral regurgitation grade 3 is between equally high and 122% higher than the risk for a patient without mitral regurgitation grade 3. The feature is part of all presented Cox models except for the 'LASSO-on-intermediate/low-risk Model' but only included in the C₄CAPRI score.

Mitral regurgitation is a common side-effect of aortic stenosis, because the stenosis often leads to an enlarged left ventricle preventing the mitral valve to function properly [28,

87]. Therefore, by removing the stenosis, the mitral regurgitation often improves [87]. However, sometimes the regurgitation stays or even worsens, which has been associated with the presence of atrial fibrillation, because this leads to a dilated left atrium which is also a cause for mitral regurgitation [87]. Why mitral regurgitation worsens the outcome and survival probability of TAVI patients can be explained by considering that mitral regurgitation often occurs together with atrial fibrillation, cerebrovascular accidents, renal impairment and poorer LVEF [28, 87]. All of these factors impair the patient's survival probability. Furthermore, mitral regurgitation, especially of a high grade, is a great stress factor for the heart and often lethal when left untreated [28].

Atrioventricular block

Similar to ΔP_m , atrioventricular block does not have a significant p-value (0.1769) and the Cox model's C-index does not deteriorate much when deleting the parameter (0.738 versus 0.741, 10-fold CV: 0.728 versus 0.731). The feature was not included in any of the presented, existing TAVI mortality scores. Atrioventricular block was more often reported as a result of TAVI, requiring permanent pacemaker implantations (see Section 2.3.2). Nevertheless, it seems probable that atrioventricular block prior to TAVI cannot have a positive or neutral influence on the patient's survival.

Peripheral artery disease

Peripheral artery disease has the most significant p-value in the model (0.0002) and in the multiple testing corrected Fisher's exact test (together with diuretics). Only Hb had a more significant corrected p-value overall. Figure 4.2 visualizes the significance of peripheral artery disease as biomarker in the Swiss-TAVI population very well, too. The exponentiated coefficient suggests that the risk for a peripheral artery disease patient is more than twice as high as the one of a patient without peripheral artery disease. The wide 95% CI (1.44-3.21) even indicates that the risk could be more than three times higher. Peripheral artery disease was part of all five presented Cox models of the thesis. Regarding existing, published scores, only C₄CAPRI includes peripheral vascular disease (to which peripheral artery disease belongs) in a TAVI risk score, though (exponentiated coefficient: 1.31, 95% CI: 0.96-1.79). Since it is often evoked by arteriosclerosis and arteriosclerosis also leads to coronary heart disease and cerebrovascular disease [28, 88], it should truly be an important biomarker for survival after TAVI.

There are also several studies claiming peripheral artery disease to be a considerable peri- and postprocedural risk parameter [88, 89]. The reason for this is that on the one side, peripheral artery disease is the major limitation for a trans-arterial approach. Therefore, many of the patients have to receive their TAVI via a transapical access route [88, 89]. As already pointed out, a non-femoral approach is linked to a significant increase in mortality by literature [20, 21, 38, 39, 57] as well as the findings of this thesis (see Section 4.1). On the other side, the comorbidities of peripheral artery disease (coronary artery disease, cerebrovascular disease, chronic renal failure) are also regarded as parameters increasing the hazard for TAVI patients [21-23, 25, 38, 39, 41, 42, 57, 88, 89].

Chronic obstructive pulmonary disease

COPD has a very significant p-value of 0.008. The exponentiated coefficient implies an increase in mortality risk of 81% for COPD patients in comparison to patients without COPD. According to the 95% CI, this risk varies from 17% to 182%. Like peripheral artery disease, COPD was part of all five presented Cox models but only part of the C₄CAPRI score regarding existing, published scores (exponentiated coefficient: 1.21, 95% CI: 0.87-1.68). In other mortality studies, COPD was also associated with a negative impact on survival after TAVI [39, 42, 90, 91]. Specified explanations for this observation included that COPD patients are often deemed ineligible for SAVR because they have an elevated risk for respiratory problems after general anesthesia. As a result, a great proportion of the TAVI cohorts consists of COPD patients (in this population 12.6%). It was observed that quite a few TAVI patients die due to respiratory failure, which was then linked to COPD [39, 42]. Hence, COPD should be regarded as a significant biomarker, but not just for TAVI patients but also in general, since it is a severe disease, which can lead to respiratory failure.

Statin medication

As previously elaborated, the statin parameter can indicate statin-treated dyslipidemia, stain-treated coronary artery disease or a combined statin and P2Y12 or Aspirin + P2Y12 treatment, respectively. Statin was neither significant in the multiple testing corrected Fisher's exact test, nor was it part of another model (except for 'LASSO-on-all' obviously). Nor was dyslipidemia, coronary artery disease or history of PCI. Supplementary Figure 5.5 also does not show separate curves for dyslipidemia.

There are separate curves for some time points for coronary artery disease, history of PCI and statin. However, while the coronary artery disease and the PCI curves indicate a worse survival for patients with the disease, the curves for statin medication indicate the contrary (2015: no-event group 61%, event group 50%, 2017: no-event group 59%, event group 47%). The curves for P2Y12 medication and Aspirin + P2Y12 treatment indicate a worse survival as well for patients who receive this medication. Analyzing the relationships further, 73% of the no-event group with coronary artery disease were treated with statin and 69% of the event group with coronary artery disease. This does not suggest a strong correlation. For dyslipidemia, 74% of the no-event group and 75% of the event-group with dyslipidemia were treated with statin, again suggesting no correlation. For P2Y12, similar results can be seen (79% statin treated in the no-event group with P2Y12, 75% in the event group) as well as for P2Y12 in combination with Aspirin (82% in the no-event group with P2Y12 + Aspirin, 80% for the event group). Therefore, the only three explanations for the positive influence of statin medication on the patient's survival (exponentiated coefficient: 0.66, 95% CI: 0.45-0.96, p-value: 0.0282) that come to mind are as follows: 1) The significance is simply wrong; 2) The statin medication itself has a positive influence on the patient's survival that could not be detected with a Fisher's exact test or with the help of the Random Forest variable significance measurements; 3) The significance is a result of the combined effects of statin-treated coronary artery disease and statin in combination with either P2Y12 alone or together with Aspirin.

As mentioned above, it was suggested that statin in combination with P2Y12 or Aspirin

+ P2Y12 reduces the probability for postoperative cardiac events and increases the chances of survival in combination [80]. Additionally, studies propose that statin has a positive influence on long-term survival after TAVI because it has an adverse effect on inflammatory processes which can occur after TAVI [92, 93]. Nevertheless, this variable and its significance should be treated with caution.

Diuretics

Diuretics medication has a significant p-value of 0.0307, indicating a hazard increase by 61% for a patient under this treatment (95% CI: 1.05-2.49). The feature was part of every presented Cox model but no published score or mortality study. Previously, we hypothesized that diuretics function as a surrogate marker for renal and heart failure. Indeed, 49% of the no-event group and 73% of the event group were treated with diuretics while either having a GFR < 60 or a NYHA class III or IV. We can therefore safely say, that diuretics in combination with GFR or NYHA class III or IV are a powerful predictor for TAVI mortality. This is in agreement with the fact that GFR or creatinine is part of six published TAVI mortality prediction scores (OBSERVANT, C₄CAPRI, UK-TAVI, STS/ACC TVT Registry, Predictor of Poor Outcomes and TARIS) and NYHA classification or dyspnea is part of four scores (OBSERVANT, C₄CAPRI, STS/ACC TVT Registry, FRANCE-2). Some of the conducted mortality biomarker studies support these findings [1, 40] as well as the significant p-values from our simple multiple testing corrected statistical testing.

Only Aspirin as alternative anticoagulation drug

The p-value indicating that only Aspirin was used as an alternative anticoagulation drug is 0.0996, the Cox model suggests that the parameter has a positive influence on the patient's hazard (exponentiated coefficient: 0.7, 95% CI: 0.46-1.07). The hypothesis proposed in the context of Pearson's correlation coefficient was that this feature's influence can be attributed to the medication itself. The feature was also significant in the Benjamini-Hochberg corrected Fisher's exact test but does not occur in the 'LASSO-on-intermediate/low-risk Model' or the 'Random Forest Model'.

A study comparing Aspirin treatment versus Aspirin + Clopidogrel treatment after TAVI came to the conclusion that treatment with only Aspirin "tended to reduce the occurrence of major adverse events following TAVR [...] [and] reduced the risk for major or life-threatening events while not increasing the risk for MI or stroke" but demanded further investigation [94]. Moreover, we have to take into account that this study investigated the influence of Aspirin treatment after and not prior to TAVI, a parameter which we are not including in our Cox model. We can therefore only hypothesize that those patients that were treated with solely Aspirin as anticoagulation therapy were continued to be treated like this after TAVI. The parameter must therefore be handled with caution.

Summary

When classifying the patients into high, intermediate and low risk groups with the help of the centered linear predictors of the Cox model, the resulting Kaplan-Meier curves

are clearly distinct from each other and yield a significant log-rank test p-value (5e-15). The features are not strongly correlated to each other. There is therefore no redundant variable in the model.

All parameters of the ‘LASSO-on-all: Smaller Model’ except for mean transvalvular pressure gradient, atrioventricular block, statin medication and ‘only Aspirin as alternative anticoagulation drug’ are regarded as powerful predictors for the 1 year all-cause mortality prediction after TAVI. More precisely, elevated age, male gender, low hemoglobin levels, the presence of peripheral artery disease and/or COPD and diuretics medication that coincides with either renal impairment or NYHA class III/IV are with a high probability significant biomarkers for increased mortality after TAVI. Even though ΔP_m was not significant, it was kept since there is an ongoing discussion whether or not low gradient patients have an increased risk regarding their TAVI outcome (which our model would support). Atrioventricular block was also kept as a predictor since it seemed probable that it would have a negative impact on the survival probability of a patient. Regarding statin medication, which was statistically significant in the model, we are currently supporting the hypothesis that statin counteracts inflammatory processes after TAVI and therefore benefits the patient’s survival probability. For the Aspirin medication as anticoagulation therapy, we are siding with the proposal that Aspirin alone is a better treatment than combined anticoagulation treatment. However, in order to interpret our results like that, we must hypothesize for both statin and Aspirin treatment that the patients took their medication in the same manner after the procedure.

4.3 Limitations

Limitations evoked by the underlying cohort are that even though the cohort size of 1434 patients and the number of events (124) are sufficiently large for statistical analyses, better and more reliable results could be achieved with more data. Input data from 2018 and 2019 was used even though there is a considerable number of missing follow-ups for this time frame. Some changes in the registry population over the years like differing valve types and devices from different producers, changes in anticoagulation medication and a decreasing mortality rate probably due to more experience, more femoral access routes, more local anesthesia procedures and more preprocedural CT images might distort the Cox model’s results. Additionally, the registry did not record frailty-related variables whose importance was emphasized in various studies. It did also not record the calcification status of the aorta and the coronary arteries, which could have improved the score as well.

The score itself is limited by the choice of the LASSO method for selecting the variables. It cannot be guaranteed that there does not exist a better combination of baseline variables of the Swiss-TAVI registry for predicting one-year all-cause mortality after TAVI. There is however evidence that the variables are powerful predictors, since most of them were significant in a multiple testing corrected t- and Fisher’s Exact test and in the random forest. Furthermore, since the cohort consists mostly of elderly patients, it is unsure whether the results can be extrapolated to younger patients. The significance of the parameters ΔP_m and atrioventricular block is also questionable. Another limitation of the score is that other risk scores (STS and EuroSCORE II) as

well as indicators of access routes and preprocedural CT were deliberately not included in the score even though they were very significant in other statistical analyses and prior Cox models.

Moreover, it is uncertain whether the variables in the score are also meaningful for predicting other adverse events like cardiovascular death, myocardial infarctions, strokes, cerebrovascular events, severe bleedings or acute kidney injuries. It was also not tested whether the features predict short-term all-cause mortality or long-term all-cause mortality beyond 1 year. Additionally, because 88% of the population were intermediate or low surgery risk patients, the score is not automatically applicable to high or extreme surgery risk patients. Since many of the features of the score are, however, occurring in published TAVI mortality scores which are mainly derived from high and extreme surgery risk populations, our presented score should be extendable to this population as well.

Finally, it should be mentioned that the score has not been tested on an external cohort.

5 Conclusion

In this section, the workflow and results of this thesis are summarized. Furthermore, conclusions are drawn from the developed mortality score and an outlook for future work on this topic is given.

5.1 Summary and Conclusion

In the scope of this thesis, a score for 1-year all-cause mortality prediction after TAVI was developed using only pre-procedural variables from the Swiss-TAVI registry. Contrary to existing scores, most of the 1434 patients used for the score development were classified as intermediate and low surgery risk patients. The score consists of 11 variables, balancing the risk of a too simple model and an overfitted model.

The score was developed by building five Cox regression models with different approaches and determining the best model by various statistical measures and tests. The first approach consisted of conducting simple statistical tests on the variables from the registry, correcting the results for multiple testing and then including the significant features in a Cox model. The second approach used the LASSO method for feature selection, the third one selected the features by building a random survival forest. The LASSO approach lead to the best results but included 17 variables and not all of these 17 variables were statistically significant. Therefore, the best possible model containing a subset of the LASSO selected variables was computed and is presented as the resulting mortality score of this thesis.

The model was tested and validated to assess its performance, robustness and plausibility in a statistical as well as in a biological sense. It reached a good concordance of 0.741 and proved to be robust in a 10-fold cross-validation where it had a mean iAUC of 0.731, outperforming all published TAVI 30-day and longterm mortality scores but the CoreValve score. Permutation tests showed that the performance of the model could not have been recreated by chance. The score successfully stratified the patients into high, intermediate and low hazard groups by using Kaplan-Meier survival curves and outperformed the classifications of the established surgery risk scores STS and EuroSCORE II. The computation of correlation coefficients for the model parameters showed that none of them were redundantly included in the score. All but two variables were significant in the statistical tests, all parameters were mentioned in the literature to be indicators for the outcome of patients after TAVI.

The score could have been improved further by a bigger data set size, the addition of frailty-related parameters and information about the calcification status of the aorta and the coronary arteries. Other limitations include that there were significant changes

in the distribution of some important registry parameters. In addition, the performance of the score was neither tested for another endpoint nor for an external cohort.

There are two major advantages of this score. First, it was constructed using a fully computational and mathematically meaningful approach for variable selection, allowing for the possibility to find new biomarkers for risk prediction while minimizing the personal bias. The biological plausibility of the discovered features was therefore assessed and verified after the mathematical process and not used as a variable selection method like in many of the published TAVI mortality scores. Second, the cohort consists of mainly intermediate and low surgery risk patients, making it applicable for these patient groups as well, a main limitation of most published scores. We strongly assume that the score will yield good results on high and extreme surgery risk cohorts, too, because the features were found to be powerful predictors in literature for these patient groups.

This score may be helpful to estimate the patient's personal hazard before he undergoes TAVI. It has been shown by various studies that TAVI always leads to better survival chances for the patient than medical therapy which is why this score should not be used for decision making between these two options. Instead, it can help to raise the patient's awareness for possible complications, thus planning more follow-up examinations and maybe preventing all-cause mortality after the procedure.

5.2 Outlook

In future studies, the score should be validated on an external cohort and different outcomes like cardiovascular mortality, long-term mortality, myocardial infarction and other VARC (Valve Academic Research Consortium) defined clinical endpoints. To further improve the score, more data and the assessment of frailty-related parameters as well as the calcification status is necessary. Since it is not guaranteed that the LASSO method is the optimal method for feature selection in Cox regression, different computational approaches have to be pursued.

Bibliography

- [1] S. V. Vasan and D. B. Sawyer, eds. *Encyclopedia of Cardiovascular Research and Medicine*. Vol. 1. Elsevier, 2018.
- [2] J. J. Popma et al. “Transcatheter aortic valve replacement using a self-expanding bioprosthetic in patients with severe aortic stenosis at extreme risk for surgery”. In: *Journal of the American College of Cardiology* 63.19 (2014), pp. 1972–1981.
- [3] G. M. Deeb et al. “Transcatheter Aortic-Valve Implantation for Aortic Stenosis in Patients Who Cannot Undergo Surgery”. In: *Journal of the American College of Cardiology* 67.22 (2016), pp. 2565–2574.
- [4] S. A. Nashef et al. “EuroSCORE II”. In: *Eur J Cardiothorac Surg* 41.4 (2012), pp. 734–744.
- [5] S. M. O’Brien et al. “The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2–isolated valve surgery”. In: *Ann. Thorac. Surg.* 88.1 Suppl (2009), pp. 23–42.
- [6] D. M. Shahian et al. “The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 1-Background, Design Considerations, and Model Development”. In: *Ann. Thorac. Surg.* 105.5 (2018), pp. 1411–1418.
- [7] S. M. O’Brien et al. “The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 2-Statistical Methods and Results”. In: *Ann. Thorac. Surg.* 105.5 (2018), pp. 1419–1428.
- [8] H. Baumgartner et al. “2017 ESC/EACTS Guidelines for the Management of Valvular Heart Disease”. In: *Rev Esp Cardiol (Engl Ed)* 71.2 (2018), p. 110.
- [9] P. Balan et al. “The Society of Thoracic Surgery Risk Score as a predictor of 30-day mortality in transcatheter vs surgical aortic valve replacement: a single-center experience and its implications for the development of a TAVR risk-prediction model”. In: *J Invasive Cardiol* 29.3 (2017), pp. 109–114.
- [10] F. Biancari et al. “Meta-analysis on the performance of the EuroSCORE II and the Society of Thoracic Surgeons Scores in patients undergoing aortic valve replacement”. In: *Journal of cardiothoracic and vascular anesthesia* 28.6 (2014), pp. 1533–1539.
- [11] N. Piazza 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 systemâ”a Bern-Rotterdam Study”. In: *American heart journal* 159.2 (2010), pp. 323–329.

- [12] E. Durand 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". In: *The American journal of cardiology* 111.6 (2013), pp. 891–897.
- [13] Y. Watanabe et al. "Is EuroSCORE II better than EuroSCORE in predicting mortality after transcatheter aortic valve implantation?" In: *Catheterization and Cardiovascular Interventions* 81.6 (2013), pp. 1053–1060.
- [14] Y. Yatsynovich et al. "Comparison of transcatheter aortic valve replacement risk score against currently accepted surgical risk models as predictors of 30-day mortality in transcatheter aortic valve replacement". In: *Journal of interventional cardiology* 30.6 (2017), pp. 595–603.
- [15] A. Halkin 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". In: *International journal of cardiology* 215 (2016), pp. 227–231.
- [16] A. W. Schoenberger et al. "Improvement of risk prediction after transcatheter aortic valve replacement by combining frailty with conventional risk scores". In: *JACC: Cardiovascular Interventions* 11.4 (2018), pp. 395–403.
- [17] J. M. Sinning et al. "Risk scores and biomarkers for the prediction of 1-year outcome after transcatheter aortic valve replacement". In: *American heart journal* 170.4 (2015), pp. 821–829.
- [18] N. Beohar et al. "The relative performance characteristics of the logistic European System for Cardiac Operative Risk Evaluation score and the Society of Thoracic Surgeons score in the Placement of Aortic Transcatheter Valves trial". In: *The Journal of thoracic and cardiovascular surgery* 148.6 (2014), pp. 2830–2837.
- [19] T. Pilgrim et al. "Predicting Mortality After Transcatheter Aortic Valve Replacement: External Validation of the Transcatheter Valve Therapy Registry Model". In: *Circ Cardiovasc Interv* 10.11 (2017).
- [20] B. Iung et al. "Predictive factors of early mortality after transcatheter aortic valve implantation: individual risk assessment using a simple score". In: *Heart* 100.13 (2014), pp. 1016–1023.
- [21] G. P. Martin et al. "Novel United Kingdom prognostic model for 30-day mortality following transcatheter aortic valve implantation". In: *Heart* 104.13 (2018), pp. 1109–1116.
- [22] P. Lantelme et al. "Development of a risk score based on aortic calcification to predict 1-year mortality after transcatheter aortic valve replacement". In: *JACC: Cardiovascular Imaging* 12.1 (2019), pp. 123–132.
- [23] M. Seiffert et al. "Development of a risk score for outcome after transcatheter aortic valve implantation". In: *Clin Res Cardiol* 103.8 (2014), pp. 631–640.
- [24] P. Debonnaire et al. "Value of the "TAVI2-SCORé" versus surgical risk scores for prediction of one year mortality in 511 patients who underwent transcatheter aortic valve implantation". In: *Am. J. Cardiol.* 115.2 (2015), pp. 234–242.

- [25] D. Capodanno et al. “A simple risk tool (the OBSERVANT score) for prediction of 30-day mortality after transcatheter aortic valve replacement”. In: *Am. J. Cardiol.* 113.11 (2014), pp. 1851–1858.
- [26] S. V. Arnold et al. “Predictors of poor outcomes after transcatheter aortic valve replacement: results from the PARTNER (Placement of Aortic Transcatheter Valve) trial”. In: *Circulation* 129.25 (2014), pp. 2682–2690.
- [27] H. C. Lederhuber and V. Lange. *Kardiologie*. Elsevier, 2010.
- [28] R. U. Mletzko and H. Moecke. *Pschyrembel Kardiologie*. Walter de Gruyter, 2012.
- [29] M. B. Leon et al. “Transcatheter or surgical aortic-valve replacement in intermediate-risk patients”. In: *New England Journal of Medicine* 374.17 (2016), pp. 1609–1620.
- [30] Michael J Reardon et al. “Surgical or transcatheter aortic-valve replacement in intermediate-risk patients”. In: *New England journal of medicine* 376.14 (2017), pp. 1321–1331.
- [31] V. H. Thourani et al. “Transcatheter aortic valve replacement versus surgical valve replacement in intermediate-risk patients: a propensity score analysis”. In: *The Lancet* 387.10034 (2016), pp. 2218–2225.
- [32] M. J. Mack et al. “Transcatheter aortic-valve replacement with a balloon-expandable valve in low-risk patients”. In: *New England Journal of Medicine* 380.18 (2019), pp. 1695–1705.
- [33] H. G. H. Thyregod et al. “Transcatheter versus surgical aortic valve replacement in patients with severe aortic valve stenosis: 1-year results from the all-comers NOTION randomized clinical trial”. In: *Journal of the American College of Cardiology* 65.20 (2015), pp. 2184–2194.
- [34] N. T. Kouchoukos et al. *Kirklin/Barratt-Boyes Cardiac Surgery*. 4th ed. Vol. 1. Elsevier Saunders, 2013.
- [35] World Health Organization and International Programme on Chemical Safety. *Biomarkers in risk assessment : validity and validation*. 2001.
- [36] D. F. Hernandez-Suarez et al. “Machine learning prediction models for in-hospital mortality after transcatheter aortic valve replacement”. In: *JACC: Cardiovascular Interventions* 12.14 (2019), pp. 1328–1338.
- [37] J. B. Hermiller et al. “Predicting Early and Late Mortality After Transcatheter Aortic Valve Replacement”. In: *J. Am. Coll. Cardiol.* 68.4 (2016), pp. 343–352.
- [38] G. Schymik et al. “European experience with the second-generation Edwards SAPIEN XT transcatheter heart valve in patients with severe aortic stenosis: 1-year outcomes from the SOURCE XT Registry”. In: *JACC: Cardiovascular Interventions* 8.5 (2015), pp. 657–669.
- [39] N. E. Moat et al. “Long-term outcomes after transcatheter aortic valve implantation in high-risk patients with severe aortic stenosis: the UK TAVI (United Kingdom Transcatheter Aortic Valve Implantation) Registry”. In: *Journal of the American College of Cardiology* 58.20 (2011), pp. 2130–2138.

- [40] M. Seiffert et al. “Predictors and outcomes after transcatheter aortic valve implantation using different approaches according to the valve academic research consortium definitions”. In: *Catheterization and Cardiovascular Interventions* 82.4 (2013), pp. 640–652.
- [41] A. Duncan et al. “Long-term outcomes after transcatheter aortic valve replacement in high-risk patients with severe aortic stenosis: the UK Transcatheter Aortic Valve Implantation Registry”. In: *JACC: Cardiovascular Interventions* 8.5 (2015), pp. 645–653.
- [42] J. Rodés-Cabau et al. “Long-term outcomes after transcatheter aortic valve implantation: insights on prognostic factors and valve durability from the Canadian multicenter experience”. In: *Journal of the American College of Cardiology* 60.19 (2012), pp. 1864–1875.
- [43] D. Silverberg et al. “The association between congestive heart failure and chronic renal disease”. In: *Current opinion in nephrology and hypertension* 13.2 (2004), pp. 163–170.
- [44] Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. “KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease”. In: *Kidney Int Suppl* 3.1 (2013), pp. 1–150.
- [45] H. Takagi et al. “Meta-analysis of impact of anemia and hemoglobin level on survival after transcatheter aortic valve implantation”. In: *The American journal of cardiology* 123.2 (2019), pp. 306–314.
- [46] A. Gatta, A. Verardo, and M. Bolognesi. “Hypoalbuminemia”. In: *Internal and Emergency Medicine* 7.3 (2012), pp. 193–199.
- [47] R. N. Forthofer, E. S. Lee, and M. Hernandez. *Biostatistics: A Guide to Design, Analysis, and Discovery*. Elsevier, 2007.
- [48] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. 2nd ed. Springer, 2009.
- [49] D. R. Cox. “Regression Models and Life-Tables”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 34.2 (1972), 187–220.
- [50] H. Zou and T. Hastie. “Regularization and variable selection via the elastic net”. In: *Journal of the royal statistical society: series B (statistical methodology)* 67.2 (2005), pp. 301–320.
- [51] G. James et al. “Linear Model Selection and Regularization”. In: *An Introduction to Statistical Learning: with Applications in R*. New York, NY: Springer New York, 2013, pp. 203–264.
- [52] R. Tibshirani. “The lasso method for variable selection in the Cox model”. In: *Stat Med* 16.4 (1997), pp. 385–395.
- [53] J. Friedman, T. Hastie, and R. Tibshirani. “Regularization Paths for Generalized Linear Models via Coordinate Descent”. In: *Journal of Statistical Software* 33.1 (2010), pp. 1–22.
- [54] N. Simon et al. “Regularization Paths for Cox’s Proportional Hazards Model via Coordinate Descent”. In: *Journal of Statistical Software* 39.5 (2011), pp. 1–13.

- [55] R. R. Hocking. “A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression”. In: *Biometrics* 32.1 (1976), pp. 1–49.
- [56] F. E. Harrell Jr, K. L. Lee, and D. B. Mark. “Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors”. In: *Statistics in medicine* 15.4 (1996), pp. 361–387.
- [57] F. H. Edwards et al. “Development and validation of a risk prediction model for in-hospital mortality after transcatheter aortic valve replacement”. In: *JAMA cardiology* 1.1 (2016), pp. 46–52.
- [58] C. R. Smith et al. “Transcatheter versus surgical aortic-valve replacement in high-risk patients”. In: *New England Journal of Medicine* 364.23 (2011), pp. 2187–2198.
- [59] M. J. Mack et al. “5-year outcomes of transcatheter aortic valve replacement or surgical aortic valve replacement for high surgical risk patients with aortic stenosis (PARTNER 1): a randomised controlled trial”. In: *The Lancet* 385.9986 (2015), pp. 2477–2484.
- [60] D. H. Adams et al. “Transcatheter aortic-valve replacement with a self-expanding prosthesis”. In: *New England Journal of Medicine* 370.19 (2014), pp. 1790–1798.
- [61] M. B. Leon et al. “3-year outcomes in high-risk patients who underwent surgical or transcatheter aortic valve replacement”. In: *Journal of the American College of Cardiology* 67.22 (2016), pp. 2565–2574.
- [62] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2019.
- [63] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc. Boston, MA, 2016.
- [64] *SWISS TAVI Registry: Prospective, National, Multi-Center Registry of Patients Undergoing Transcatheter Aortic Valve Implantation*. 2011. URL: <https://clinicaltrials.gov/ct2/show/NCT01368250> (visited on 02/01/2020).
- [65] P. M. Wenaweser et al. “Short-term clinical outcomes among patients undergoing transcatheter aortic valve implantation in Switzerland: the Swiss TAVI registry”. In: *EuroIntervention* 10.8 (2014), pp. 982–989.
- [66] Boston scientific corporation. *Form 8-K Current Report Pursuant To Section 13 or 15(d) of the securities exchange act of 1934*. 2017. URL: <https://www.sec.gov/Archives/edgar/data/885725/000088572517000004/form8-kproductannouncement.htm> (visited on 02/03/2020).
- [67] Cardiovascular News. *TAVI company Direct Flow Medical closes after failing to secure funding*. 2017. URL: <https://cardiovascularnews.com/tavi-company-direct-flow-medical-closes-after-failing-to-secure-funding/> (visited on 02/03/2020).
- [68] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2019.
- [69] T. M. Therneau. *A Package for Survival Analysis in S*. version 2.38. 2015.

- [70] T. M. Therneau and P. M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. New York: Springer, 2000. ISBN: 0-387-98784-3.
- [71] S. Potapov, W. Adler, and M. Schmid. *survAUC: Estimators of prediction accuracy for time-to-event data*. R package version 1.0-5. 2012.
- [72] R. Kolde. *pheatmap: Pretty Heatmaps*. R package version 1.0.12. 2019.
- [73] J. Ehrlinger. *ggRandomForests: Visually Exploring Random Forests*. R package version 2.0.1. 2016.
- [74] X. Chen and H. Ishwaran. “Random forests for genomic data analysis”. In: *Genomics* 99.6 (2012), pp. 323–329.
- [75] T. Ali et al. “The changing pattern of referral in acute kidney injury”. In: *QJM: An International Journal of Medicine* 104.6 (2011), pp. 497–503.
- [76] X. Bao et al. “Low serum creatinine predicts risk for type 2 diabetes”. In: *Diabetes/metabolism research and reviews* 34.6 (2018), e3011.
- [77] R. C. Skoda. “Thrombocytosis”. In: *ASH Education Program Book* 2009.1 (2009), pp. 159–167.
- [78] G. N. Kouvelos et al. “Postoperative levels of cardiac troponin versus CK-MB and high-sensitivity C-reactive protein for the prediction of 1-year cardiovascular outcome in patients undergoing vascular surgery”. In: *Coronary artery disease* 22.6 (2011), pp. 428–434.
- [79] T. Taniguchi et al. “High- Versus Low-Gradient Severe Aortic Stenosis: Demographics, Clinical Outcomes, and Effects of the Initial Aortic Valve Replacement Strategy on Long-Term Prognosis”. In: *Circ Cardiovasc Interv* 10.5 (2017).
- [80] A. M. Steely et al. “Underutilization of antiplatelet and statin therapy after postoperative myocardial infarction following vascular surgery”. In: *Journal of vascular surgery* 67.1 (2018), pp. 279–286.
- [81] *Life Expectancy of the World Population*. URL: <https://www.worldometers.info/demographics/life-expectancy/> (visited on 03/01/2020).
- [82] T. Bauer et al. “Left ventricular function determines the survival benefit for women over men after transcatheter aortic valve implantation (TAVI)”. In: *EuroIntervention* 13.04 (2017), pp. 467–474.
- [83] A. Erez et al. “Factors affecting survival in men versus women following transcatheter aortic valve implantation”. In: *The American journal of cardiology* 113.4 (2014), pp. 701–705.
- [84] M. A. Sherif et al. “Effect of gender differences on 1-year mortality after transcatheter aortic valve implantation for severe aortic stenosis: results from a multicenter real-world registry”. In: *Clinical Research in Cardiology* 103.8 (2014), pp. 613–620.
- [85] V. Stangl et al. “Influence of sex on outcome following transcatheter aortic valve implantation (TAVI): systematic review and meta-analysis”. In: *Journal of interventional cardiology* 27.6 (2014), pp. 531–539.

- [86] M. A. Clavel, J. Magne, and P. Pibarot. “Low-gradient aortic stenosis”. In: *Eur. Heart J.* 37.34 (2016), pp. 2645–2657.
- [87] K. Boerlage-van Dijk et al. “Mitral regurgitation prior to transcatheter aortic valve implantation influences survival but not symptoms”. In: *International journal of cardiology* 204 (2016), pp. 95–100.
- [88] J. M. Sinning et al. “The impact of peripheral arterial disease on early outcome after transcatheter aortic valve implantation: results from the German Transcatheter Aortic Valve Interventions Registry”. In: *American heart journal* 164.1 (2012), pp. 102–110.
- [89] D. Ueshima et al. “The impact of pre-existing peripheral artery disease on transcatheter aortic valve implantation outcomes: A systematic review and meta-analysis”. In: *Catheterization and Cardiovascular Interventions* (2019).
- [90] Y. B. Liao et al. “The relationship between chronic obstructive pulmonary disease and transcatheter aortic valve implantation: A systematic review and meta-analysis”. In: *Catheterization and Cardiovascular Interventions* 87.S1 (2016), pp. 570–578.
- [91] M. Mok et al. “Chronic obstructive pulmonary disease in patients undergoing transcatheter aortic valve implantation: insights on clinical outcomes, prognostic markers, and functional status changes”. In: *JACC: Cardiovascular Interventions* 6.10 (2013), pp. 1072–1084.
- [92] I. Merdler et al. “Effect of statin therapy and long-term mortality following transcatheter aortic valve implantation”. In: *The American journal of cardiology* 123.12 (2019), pp. 1978–1982.
- [93] H. Takagi et al. “Meta-analysis for impact of statin on mortality after transcatheter aortic valve implantation”. In: *The American journal of cardiology* 124.6 (2019), pp. 920–925.
- [94] J. Rodés-Cabau et al. “Aspirin versus aspirin plus clopidogrel as antithrombotic treatment following transcatheter aortic valve replacement with a balloon-expandable valve: the ARTE (Aspirin Versus Aspirin+ Clopidogrel Following Transcatheter Aortic Valve Implantation) randomized clinical trial”. In: *JACC: Cardiovascular Interventions* 10.13 (2017), pp. 1357–1365.

Appendix

Supplementary Data

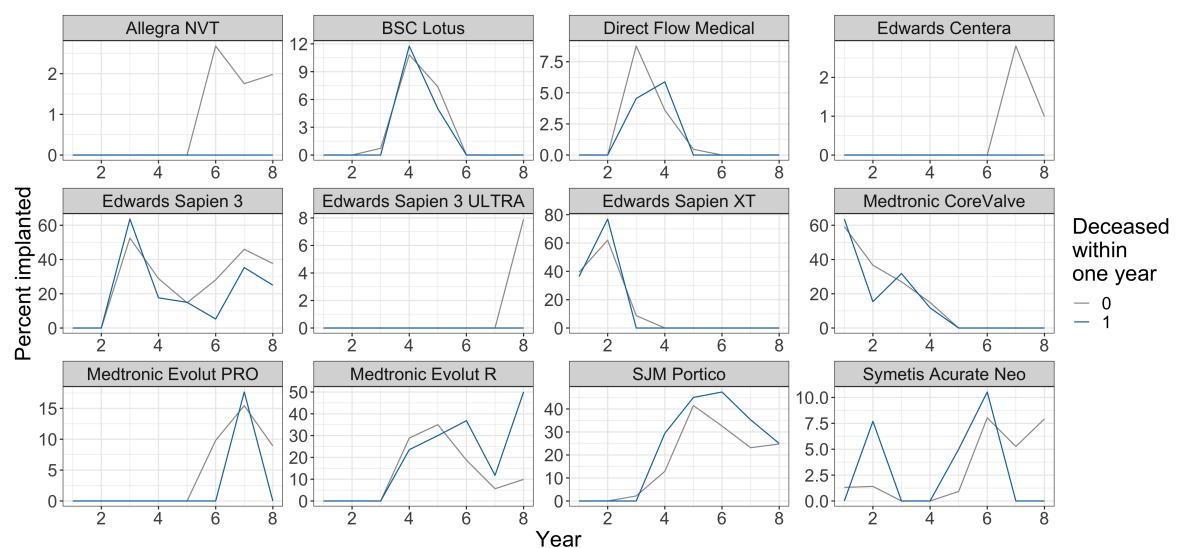


Figure 5.1: Percentage of devices implanted, stratified by event. The event is defined as all cause mortality.

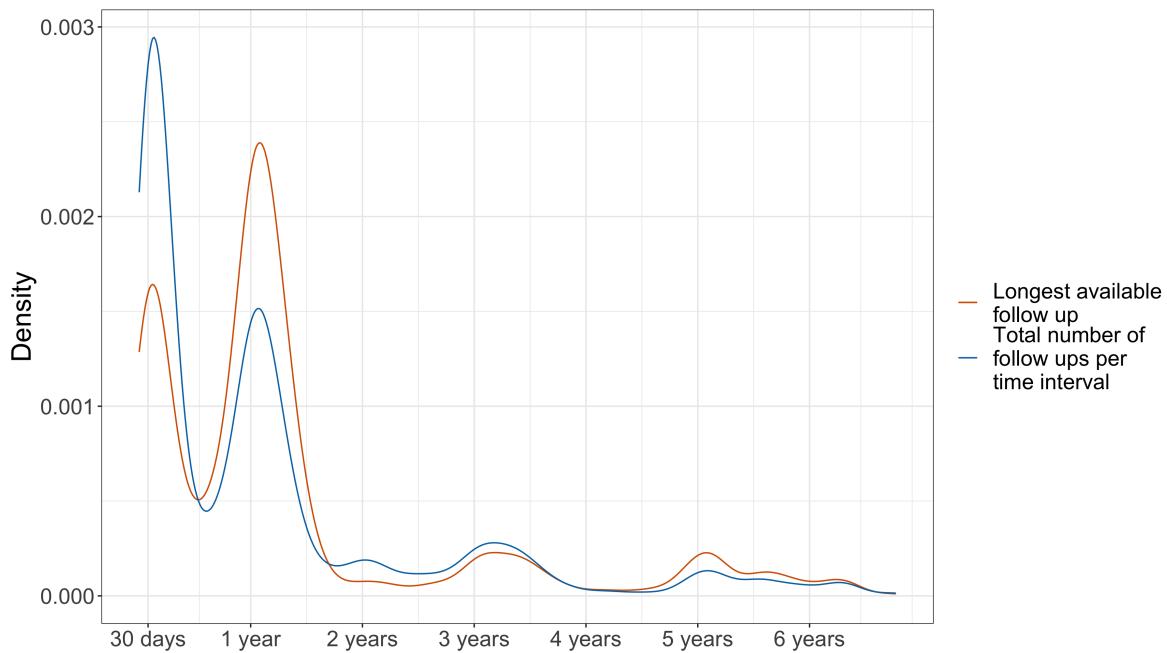


Figure 5.2: Density plot of the distribution of all available follow-up time points and of the longest follow-up available for each patient. Most of the follow-ups were made 30 days after the procedure, most of the patients were just followed up for one year. The time points are calculated as difference between the procedure date and the follow-up date.

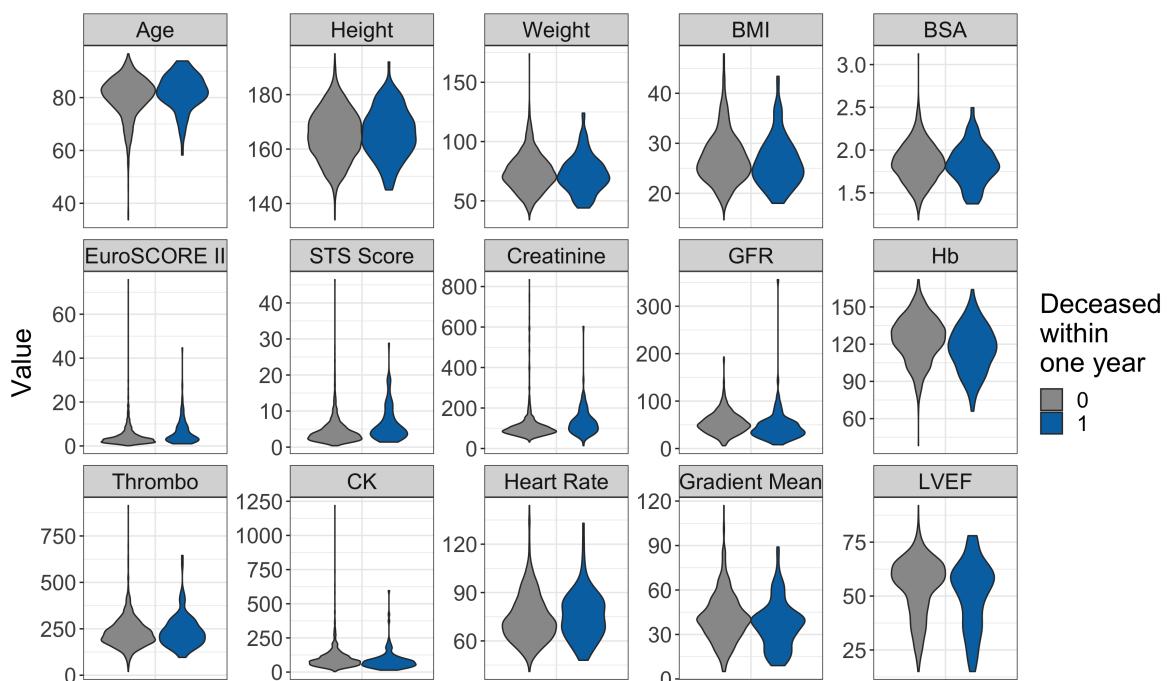


Figure 5.3: Violin Plots for the distribution of the continuous variables, stratified by event.

Appendix

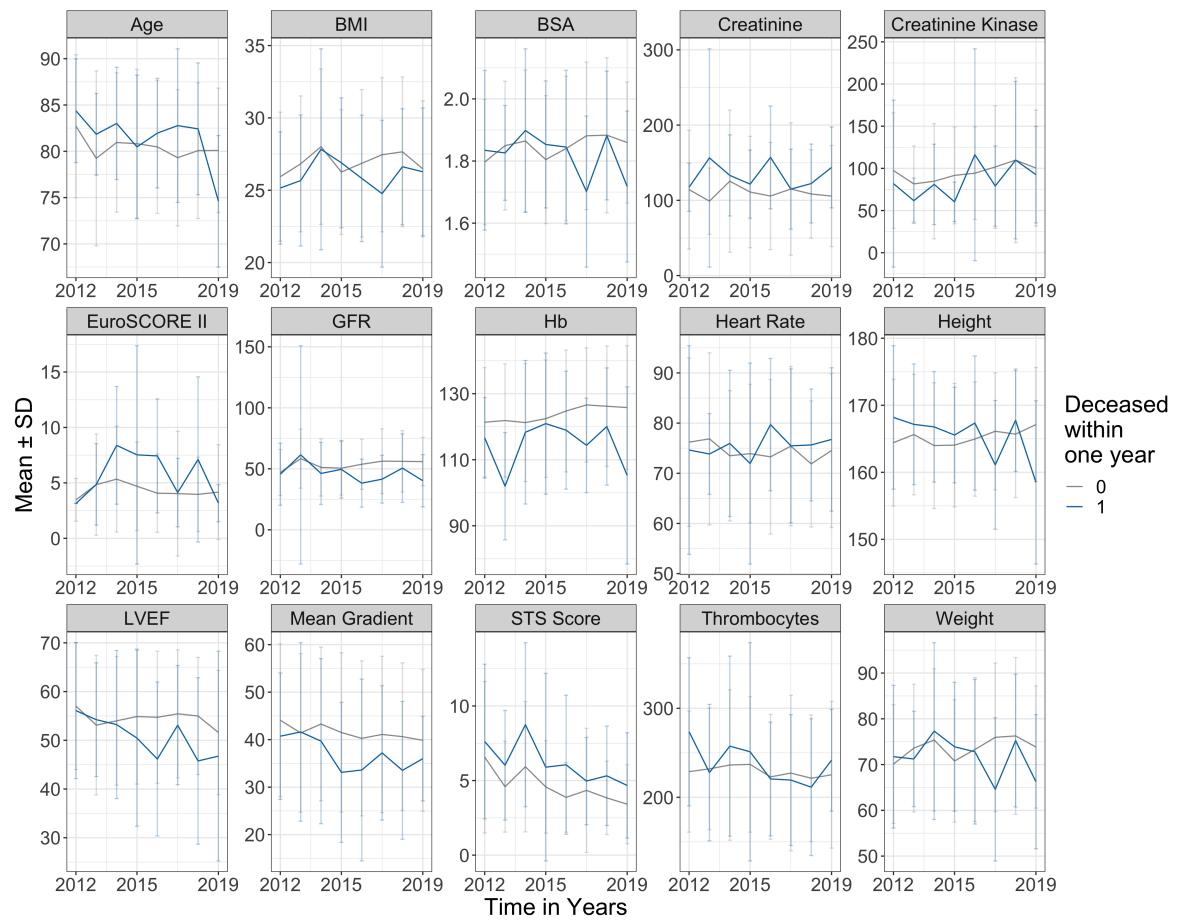


Figure 5.4: Distribution of the continuous variables over the years, stratified by event.

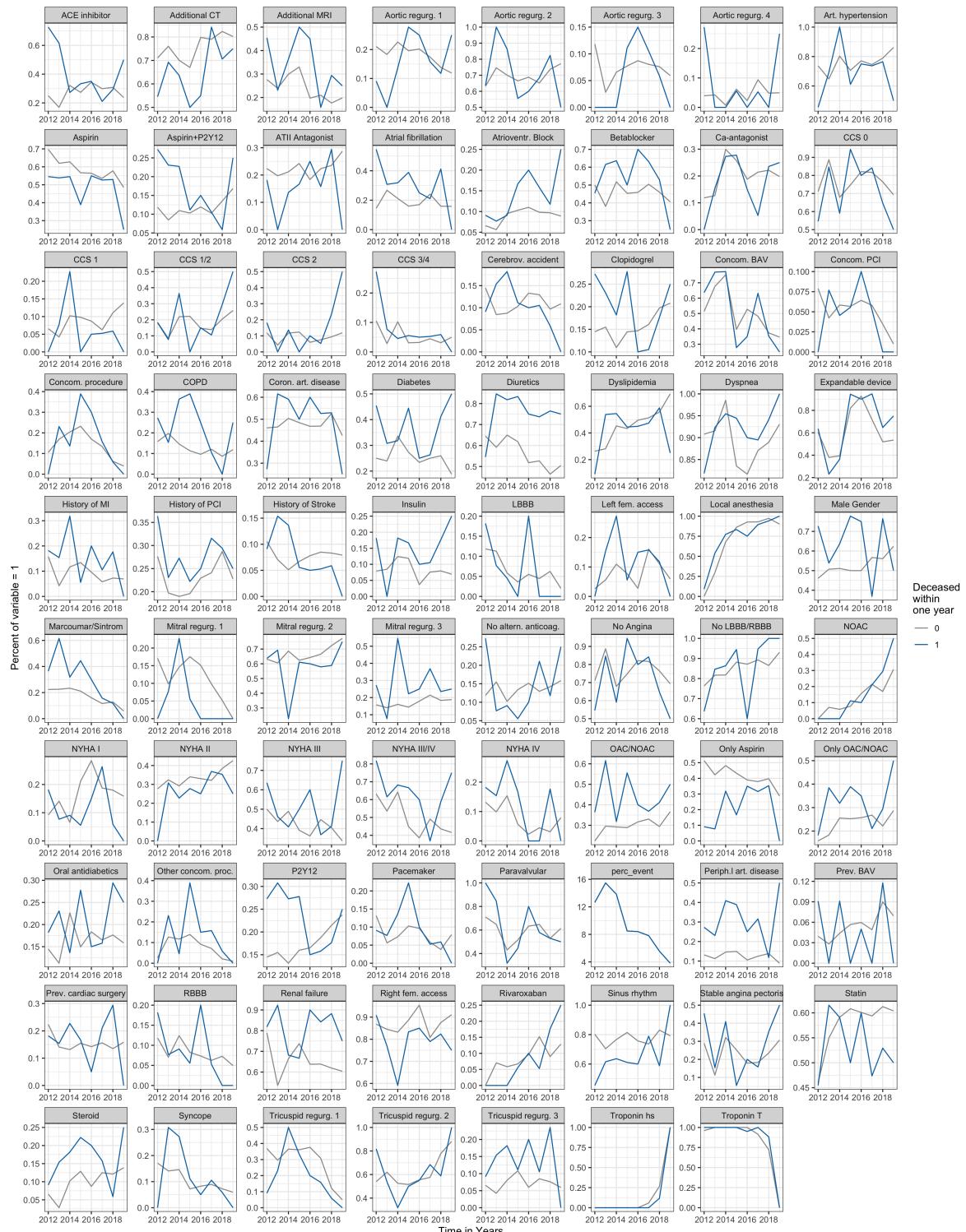


Figure 5.5: Distribution of the boolean-coded variables, stratified by event.

Appendix

	Exp (Coef-ficient)	standard error	Wald statistic of the co-efficient	p-value of the Wald statistic	95% CI
Age	1.04	0.014	2.51	0.012	1.01-1.07
Creatinine	1.0	0.001	0.07	0.9422	1.0-1.0
Hb	0.98	0.005	-3.14	0.0017	0.97-0.99
LVEF	0.99	0.008	-1.46	0.145	0.97-1.0
ΔP_m	1.0	0.007	-0.7	0.4818	0.98-1.01
Atrial fibrilla-tion	1.25	0.236	0.93	0.3519	0.78-1.98
History of MI	1.13	0.249	0.48	0.6318	0.69-1.84
Mitral regurg. grade 3	1.31	0.218	1.23	0.2196	0.85-2.0
Tricuspid re-gurg. grade 3	1.23	0.274	0.76	0.4472	0.72-2.11
NYHA class IV	1.15	0.286	0.48	0.6284	0.66-2.01
Peripheral artery disease	2.01	0.209	3.33	0.0009	1.33-3.03
COPD	1.79	0.23	2.53	0.0116	1.14-2.81
Other con-comitant procedure	1.12	0.273	0.42	0.6775	0.66-1.91
Marcoumar/ Sintrom	1.26	0.315	0.74	0.4576	0.68-2.34
Novel anti-coagulation drugs	0.76	0.32	-0.84	0.4019	0.41-1.43
Diuretics	1.42	0.229	1.53	0.1261	0.91-2.22
Only Aspirin as anticoag.	0.67	0.245	-1.61	0.1066	0.42-1.09

Table 5.1: Cox model for the ‘Benjamini-Hochberg Model’

	Exp (Coef-ficient)	standard error	Wald statistic of the coefficient	p-value of the Wald statistic	95% CI
Age	1.04	0.014	2.57	0.0102	1.01-1.07
Gender	1.82	0.201	2.99	0.0028	1.23-2.7
COPD	1.83	0.229	2.65	0.008	1.17-2.87
Previous BAV	0.45	0.436	-1.81	0.0710	0.19-1.07
Peripheral artery disease	2.17	0.209	3.71	0.0002	1.44-3.27
Marcoumar/ Sintrom	1.06	0.227	0.25	0.8027	0.68-1.65
Statin medication	0.63	0.195	-2.38	0.0175	0.43-0.92
Diuretics	1.48	0.231	1.69	0.0913	0.94-2.32
Insulin	1.54	0.274	1.57	0.1153	0.9-2.63
Hb	0.98	0.005	-4.03	0.0001	0.97-0.99
Atrio-ventricular block	1.4	0.277	1.22	0.2226	0.81-2.42
ΔP_m	0.99	0.007	-0.94	0.3473	0.98-1.01
LVEF	0.99	0.008	-0.87	0.3838	0.98-1.01
Only Aspirin as anticoag.	0.8	0.235	-0.97	0.3334	0.5-1.26
Atrial fibrillation	1.21	0.219	0.88	0.3807	0.79-1.86
Mitral regurgitation grade 3	1.37	0.218	1.45	0.1473	0.89-2.1
Tricuspid regurgitation grade 3	1.24	0.277	0.77	0.442	0.72-2.13

Table 5.2: Cox model for the ‘LASSO-on-all Model’

Appendix

	Exp (Coef-ficient)	standard error	Wald statistic of the co-efficient	p-value of the Wald statistic	95% CI
Gender	1.85	0.229	2.68	0.0074	1.18-2.89
COPD	1.67	0.272	1.88	0.0598	0.98-2.84
Peripheral artery disease	1.66	0.251	2.03	0.0425	1.02-2.72
Diuretics	1.73	0.238	2.29	0.0218	1.08-2.76
Hb	0.98	0.006	-3.29	0.001	0.97-0.99
CCS class 1	0.22	0.717	-2.1	0.036	0.05-0.91
Tricuspid re-gurg. grade 3	2.14	0.304	2.5	0.0125	1.18-3.88

Table 5.3: Cox model for the ‘LASSO-on-intermediate/ low risk Model’

	Exp (Coef-ficient)	standard error	Wald statistic of the co-efficient	p-value of the Wald statistic	95% CI
Age	1.04	0.014	2.98	0.0029	1.01-1.07
Gender	1.73	0.203	2.72	0.0065	1.17-2.58
Creatinine	1.0	0.001	-0.29	0.7739	1.0-1.0
Hb	0.98	0.005	-3.79	0.0001	0.97-0.99
Creatinine ki-nase	1.0	0.002	-0.48	0.6323	1.0-1.0
LVEF	0.99	0.008	-1.1	0.273	0.98-1.01
ΔPm	0.99	0.007	-0.83	0.4072	0.98-1.01
Peripheral artery disease	2.04	0.207	3.45	0.0006	1.36-3.07
COPD	1.73	0.236	2.33	0.0198	1.09-2.75
Diuretics	1.5	0.23	1.76	0.0778	0.96-2.35
Atrial fibrilla-tion	1.31	0.199	1.38	0.1688	0.89-1.94
Steroid medi-cation	1.29	0.262	0.97	0.3342	0.77-2.15
Mitral regurg. grade 3	1.47	0.211	1.82	0.0695	0.97-2.22

Table 5.4: Cox model for the ‘Random Forest Model’

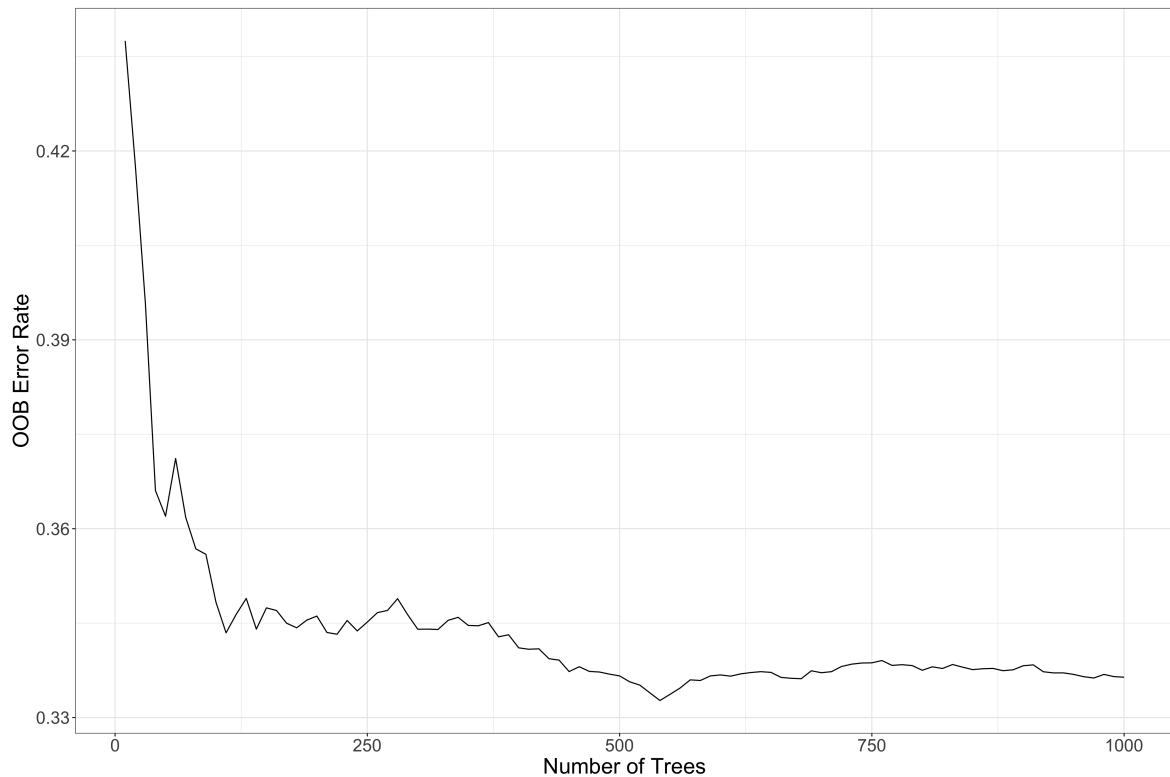


Figure 5.6: Out of Bag error rate for the number of trees.

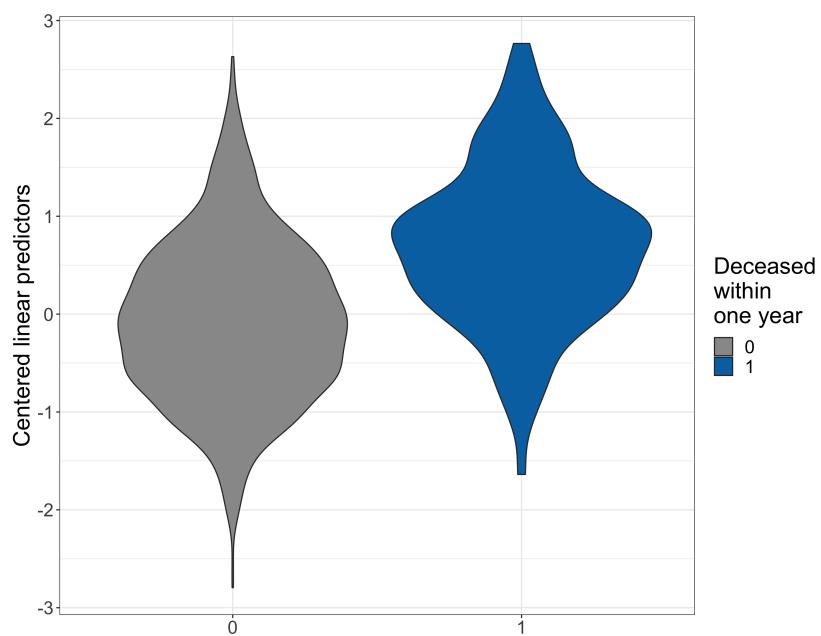


Figure 5.7: Distribution of the centered linear predictors for the 'LASSO-on-all: Smaller model'.

Appendix

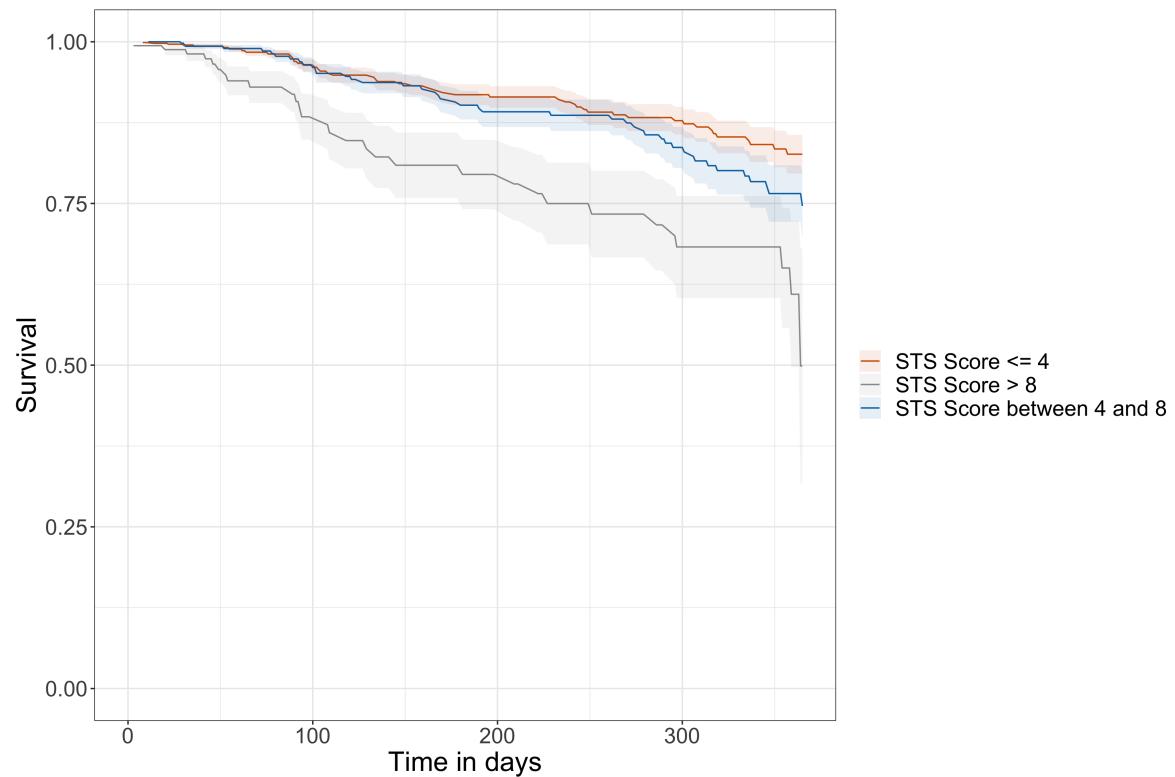


Figure 5.8: Kaplan-Meier curves STS score

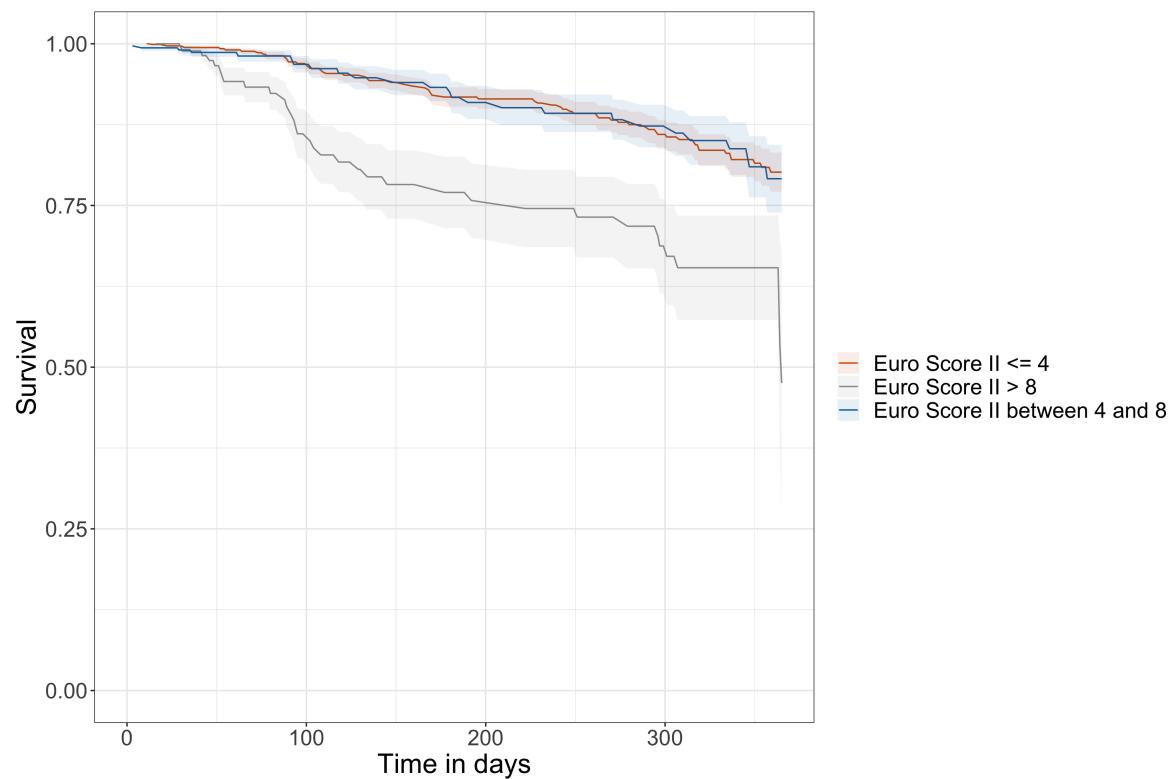


Figure 5.9: Kaplan-Meier curves Euro II score