Predicting Clinical Outcomes in Cardiovascular Diseases

METHODOLOGICAL ADVANCEMENTS AND APPLICATIONS

Sara J. Baart

Predicting Clinical Outcomes in Cardiovascular Diseases

METHODOLOGICAL ADVANCEMENTS

AND APPLICATIONS

Sara J. Baart

Predicting Clinical Outcomes in Cardiovascular Diseases

methodological advancements and applications

Sara Johanna Baart

Sara Baart, 2019
ISBN: 978-94-6323-804-5
Cover Design: Richard van Willegen
Printed by Gildeprint

Predicting Clinical Outcomes in Cardiovascular Diseases

methodological advancements and applications

Het voorspellen van klinische uitkomsten in hart- en vaatziekten

methodologische vorderingen en toepassingen

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof. dr. R. C. M. E. Engels

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op
8 oktober 2019 om 09:30

door

Sara Johanna Baart
geboren te Amsterdam

Erasmus University Rotterdam

Promotiecommissie

Promotoren:	Prof. dr. ir. H. Boersma
	Prof. dr. D. Rizopoulos
Overige leden:	Prof. dr. E. W. Steyerberg
	Prof. dr. ir. Y. T. van der Schouw
	Prof. dr. J. J. M. Takkenberg
Co-promotor:	Dr. I. Kardys

The research described in this thesis was supported by a grant of the Dutch Heart Foundation (2013T083).

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

Financial support by Cardialysis for the publication of this thesis is gratefully acknowledged.

In liefdevolle herinnering aan mijn oma

Contents

Appendix

Introduction

Predicting cardiovascular disease

Cardiovascular disease (CVD) is a class of disorders that affects the heart and vessels. [1] Although cardiovascular mortality has decreased drastically in the last forty years, it is still the leading cause of death in Western countries, and in the Netherlands 1 in 4 deaths are due to CVD. [1–3] The decrease in mortality is caused by improvements in both prevention and treatment of CVD. It is estimated that 90% of cardiovascular diseases are preventable, and many factors that lead to CVD are modifiable. [4] Modifiable risk factors include smoking, high blood pressure, high cholesterol, obesity, stress, physical activity and diet. Examples of non-modifiable factors that can lead to CVD are gender, ethnicity and a family history of cardiovascular disease. The development of CVD is often caused by a combination or different factors.

Before attempts can be made to (further) reduce the burden of CVD, the target population at increased risk needs to be defined. CVD prediction models have proven useful in this respect, as these not only reveal which factors contribute to the risk of developing CVD, but they also enable quantification of these contributions. Prediction models have also been developed for patients with established CVD, to estimate their risk of future CVD events, such as, for example, cardiovascular death or hospitalizations. The risks estimated by these models can ultimately be used to make a decision regarding starting or intensifying treatment. Prediction models often include biological markers, or so-called 'biomarkers', as risk factors. Formally defined is a biomarker "[a] characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention", [5] and can be seen as a measurement taken from a person that gives information on his/her health status. Blood pressure is an example of a biomarker that is often used in CVD prediction models. Some biomarkers can be measured in the blood, such as cardiac Troponins and NT-proBNP, among numerous others. Ideally, blood biomarkers are sensitive and specific tracers of dynamic pathophysiological (disease) processes, and can thus be used for purposes of diagnosis or prognostication.

Improvements in predicting CVD

In different aspects of the prediction of CVD improvements can be made.

First, in recent years, gender has gained particular attention as a factor associated with CVD. For a long time the field of CVD research has focused on men, and biological differences between men and women have been insufficiently addressed. However, there are distinct differences in CVD between men and women, both in the causes and in the manifestation of the disease. In women, the symptoms of a heart attack are often less pronounced, problems occur more often in the smallest vessels compared to the big arteries, and the disease occurs on average 7-10 years later. [6] Moreover, several pregnancy related and reproductive disorders in women have been associated with subsequent development of CVD. [7–9] These are factors that occur exclusively in women. It is yet unclear whether the current prediction models suffice for risk stratification in the female population, or that the models need to be updated.

Second, most prediction models are static models, meaning that the model aims to predict the incidence of CVD or mortality over a certain period of time based on one assessment of the status of a person. The time horizons for prediction vary per model, with some models aiming to estimate risk over ten years or even a lifetime risk. One of the most used prediction models is the Framingham Risk Score, of which different versions exist. The version developed by D'Agostino et al. (2008) has a prediction horizon of 10 years. [10] In 2009 a Framingham model was made to predict a 30-year risk of CVD. [11] However, over such a long period of time, someone's health status will not remain the same. People can adjust their lifestyle to increase their physical activity and decrease their cholesterol levels, for example. Therefore, risk prediction models can potentially be improved if repeated measurements over time are obtained and changes incorporated in the model.

Modelling strategies

Hierarchical models

Repeated measures within patients pose extra methodological challenges, however. In studies with repeated measurements, observations are clustered within a patient. As a consequence the observations are not independent of each other, an assumption made in most standard modelling techniques. Statistical models are available to deal with this issue by taking the clustering of the observations into account. A framework that is often used to analyze such data are mixed effects models. The general idea is that these models estimate both *fixed effects*, which are the mean population effects, and *random effects*, which are cluster specific effects. For notation let $y_i(t)$ denote a continuous repeated or longitudinal measurement for patient i at time t, for example blood pressure that is measured during visits to the treating physician. The mixed effects model for y is of the form

$$y_i(t) = x_i^\top(t)\beta + z_i^\top(t)b_i + \varepsilon_i(t), \tag{1}$$

where β is the vector of parameters for the fixed effects and b_i the vector of random effects for cluster (patient) i. In the mixed effects models, the random effects b_i are usually assumed to follow a normal distribution with mean 0 and covariance matrix D. The design vector for the fixed effects is denoted by $x_i(t)$ and the design vector for the random effects by $z_i(t)$. Non-linear evolutions can be modelled by introducing more complex modelling structures in the design vectors of the fixed and random effects, such as quadratic terms or splines. The error terms are denoted by $\varepsilon_i(t)$ and are also assumed to be normally distributed with mean 0 and variance σ^2.

Problems in mixed models occur when there is missing data due to dropout of patients. Patients that experience an event during the follow up period, as well as those who are too sick to visit the physician will cause missing scheduled observations, which may lead to biased effect estimates. We call this type of missing data 'missing not at random' (MNAR), and an important feature of this type of missingness is that

its mechanism depends on unobserved data. When the missingness depends only on data that are observed, we call this 'missing at random' (MAR) and subsequent mixed effects models will provide unbiased estimates.

The mixed modeling framework can also be used in studies where patients themselves are grouped; patients can belong to different hospitals and/or different countries. We can expect patients from the same country, for example, to be correlated with each other, and through a random effect for the grouping variable this can be taken into account in the model. Nested random effects can also be added to mixed effects models, when higher level hierarchy occurs in the study. These nested random effects can be necessary when a study includes patients from different hospitals in different countries.

Joint models

The information gained by repeatedly measuring characteristics of a patient can provide prognostic value for the event of interest. A potential way to incorporate this information is by adding the longitudinal marker as a time-dependent covariate in the model for the event, such as a time-dependent version of the Cox proportional hazards model [12, 13]. This model handles the covariate as being constant between two measurements and is in general suitable for covariates that are *exogenous*. A variable is exogenous when its value at time t can be known somewhere before t, such as which nurse will treat the patient at a specific visit. Biomarkers, on the other hand, are *endogenous* variables and will not stay constant between two measurements, making the time-dependent Cox model an unsuitable model for these longitudinal outcomes.

This issue, as well as the above-described MNAR problem for the mixed effects models, can both be solved by using the joint modeling framework for longitudinal and time-to-event data. [14–16] In this framework, a mixed effects model as described above is combined with a model for a time-to-event (or survival) outcome. Both models are estimated jointly to relate the value of the (modelled) longitudinal outcome at each point in time to the hazard of the event. Because the dropout process is

now modelled explicitly, the longitudinal trajectories estimated by the mixed effects submodel are unbiased. A graphical representation of the joint model can be found in Figure 1.

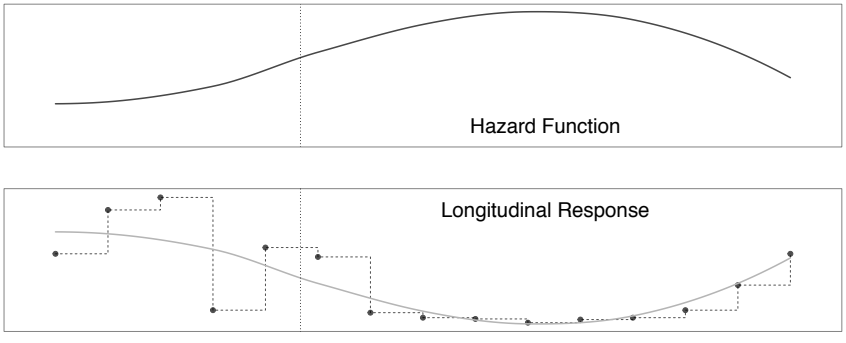


Figure 1: Graphical representation of the joint model

The points in the lower panel of Figure 1 represent the repeatedly measured risk factor such as blood pressure. The estimated profile, obtained by the mixed effects submodel, is displayed with the light grey line. This line gives an estimated value for the biomarker at each point in time and not only at the measured time points. In the upper panel, the value of the light grey line is linked to the hazard of the event. We can see that the hazard of the event increases as the biomarker value decreases. The formula of the joint model is as follows

$$\begin{cases} y_i(t) = m_i(t) + \varepsilon_i(t) \\ \qquad\quad = x_i^\top(t)\beta + z_i^\top(t)b_i + \varepsilon_i(t) \\ \\ h_i(t) = h_0(t)\exp\{\gamma^\top w_i + \alpha m_i(t)\}. \end{cases} \tag{2}$$

Now, $m_i(t)$ is the estimated biomarker value at time t for patient i and corresponds to the light grey line in Figure 1. For the time-to-event outcome let T_i^* denote the time of the event. This is often not measured for the full cohort, because studies usually end before all the patients reach the end point of interest and we call these patients

censored, with C_i being the censoring time and $T_i = \min(T_i^*, C_i)$ the observed time. Additionally, for each patient the event indicator δ_i is given as 1 if $T_i^* \leq C_i$ and 0 otherwise. The hazard for the survival outcome (T_i^*, δ_i) is modelled by $h_i(t)$ with a proportional hazards model and is represented by the line in the upper panel of Figure 1. The two outcomes are linked through the association parameter α.

Bayesian analyses

The joint models estimated in this thesis will be fitted using Bayesian analysis. The Bayesian modeling framework is a way of estimating statistical models which differs from the classical way of analyzing data, referred to as the *frequentist* approach. In Bayesian models the parameters estimated in a model are viewed as random variables that follow a distribution, whereas in the frequentist framework parameters have a fixed value. The Bayesian method combines a prior belief about the parameter (prior distribution) with a likelihood estimated from the current data to obtain a posterior distribution around the parameter of interest following Bayes' theorem

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)}. \tag{3}$$

Here, θ represents the parameter of interest and y the data. $p(y \mid \theta)$ is the likelihood calculated from the data, and $p(\theta)$ the prior belief about the distribution of the parameter. Lastly, $p(y)$ is the marginal probability of the y, independent of θ. Often, the posterior distribution $(p(\theta \mid y))$ is hard to obtain analytically. Accordingly, Monte Carlo Markov Chain (MCMC) [17] sampling methods have been developed and will be used in this thesis to obtain estimates of the posterior distribution. Inference on the parameters can be done on the resulting samples which come from the posterior distribution.

Two-phase Sample Designs and Joint models

Sometimes, the assessment of biomarkers can be expensive. This holds special importance for studies where blood biomarkers are measured with high frequency and

not one, but multiple biomarkers are of interest. To avoid these longitudinal marker studies from becoming too expensive, a so-called two-phase sampling design can be applied to the available patient cohort. The sampling of the patient cohort from the larger population can be viewed as the first phase, and in the second phase a subset of the measurements in the cohort is taken. One type of a two-phase sampling design is the case-cohort design. [18] In this design all patients experiencing the study end point are selected, and only a random sample of the patients that did not reach the study end point. The biomarker values are ascertained only for the patients that are selected. As a consequence, the patients with the study end point are over-represented compared to the full cohort. Models estimated on the subset will give a misspecification of the baseline hazard. This in turn leads to biased estimates of the model parameters and biased estimates of the survival probabilities. New methodology is needed to obtain valid estimates for the joint modeling framework in a case-cohort design.

Relative Conditional Survival models

Another way to model patient outcomes in a dynamic matter is by calculating *conditional* and *relative conditional* survival estimates. These methods, popular in oncology research, aim to provide additional information on the prognosis of a patient by incorporating time a patient has already survived after a certain treatment or diagnosis into the prognosis, and by comparing prognosis of the patients to that of someone of the same age and gender in the general population. [19–22] Often, the initial period after a treatment, such as an operation, is most dangerous for a patient and if he survives the first crucial period, his risk of dying can change radically. Accordingly, the estimated risk of mortality can be updated by incorporating the fact that the patient is still alive at this point. Additionally, mortality rates often include deaths due to other causes than the disease of interest. The proportion of mortality that can actually be attributed to the disease of interest, can be calculated with relative survival. Overall survival is compared to survival rates from someone of the same age and gender in the general population. These rates can be especially informative for

older patients, because they are more likely to die from other causes than just the disease of interest. When both methods are combined, relative conditional survival can demonstrate at which point in time the patient's mortality is the same as the general population and the patient is, in a statistical way of thinking, *cured*. [19]

Research Questions

This thesis aims to answer several questions relating to above-described aspects of clinical outcome prediction in cardiovascular disease and these form the four parts of the thesis:

- How do we obtain unbiased results when estimating joint models in a case-cohort design?
- Can we obtain additional insights into the prognosis of cardiovascular patients by calculating relative conditional survival?
- Concerning the gender aspect of predicting CVD; which models predicting CVD in women exist, are female-specific risk factors included, and how well do they perform?
- Can we improve outcome prediction in cardiovascular patient populations by applying hierarchical modelling techniques?

Thesis outline

The outline of this thesis is as follows. In **Part I** and **Chapter 1** we face an important problem when estimating joint models. A longitudinal biomarker study has been performed using a case-cohort design. In this chapter we investigate how to obtain unbiased results in this scenario, both in parameter estimates and in the predictive accuracy of the model. **Part II** investigates the impact of relative conditional survival methods, popular in oncology research, on prognosis in two cardiovascular patient populations. First we investigate patient survival after percutaneous coronary intervention (PCI) in **Chapter 2** and in **Chapter 3** we investigate prognosis in

patients with heart failure. **Part III** focuses on CVD outcomes in women. In **Chapter 4** we report the results of a systematic review performed on all cardiovascular prediction models in women published so far. We aim to provide a complete overview of existing models, and to present advice on which models should best be used when predicting cardiovascular risk in practice. In **Chapter 5** we aim to model pregnancy outcomes in women with structural heart disease. We face methodological challenges because the patients are clustered within hospitals within different countries. By employing a three-level cluster model these problems can be addressed in a correct way. In **Part IV** various hierarchical modelling techniques are applied to a wide range of clinical cardiovascular disease problems where patient characteristics were measured repeatedly over time. In most cases the aim was to relate the repeated biomarkers to an event of interest (**Chapters 6** to **9**).

References

1. World Health Organization. *Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016* Report (2018).
2. De Boer, A., van Dis, I., Visseren, F. L. J., Vaartjes, I. & Bots, M. *Hart- en vaatziekten in Nederland 2018* Report (Hartstichting, 2018).
3. Mensah, G. A. *et al.* Decline in Cardiovascular Mortality: Possible Causes and Implications. *Circ Res* **120,** 366–380 (2017).
4. Yusuf, S. *et al.* Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* **364,** 937–52 (2004).
5. Biomarkers Definitions Working, Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* **69,** 89–95 (2001).
6. EUGenMed Cardiovascular Clinical Study Group *et al.* Gender in cardiovascular diseases: impact on clinical manifestations, management, and outcomes. *Eur Heart J* **37,** 24–34 (2016).
7. Chen, C. W., Jaffe, I. Z. & Karumanchi, S. A. Pre-eclampsia and cardiovascular disease. *Cardiovasc Res* **101,** 579–86 (2014).
8. Garovic, V. D. *et al.* Hypertension in pregnancy as a risk factor for cardiovascular disease later in life. *Journal of Hypertension* **28,** 826–833 (2010).
9. Van Lennep, J. E. R., Heida, K. Y., Bots, M. L., Hoek, A. & Dutch, C. Cardiovascular disease risk in women with premature ovarian insufficiency: A systematic review and meta-analysis. *European Journal of Preventive Cardiology* **23,** 178–186 (2016).

10. D'Agostino, R. B. *et al.* General cardiovascular risk profile for use in primary care - The Framingham Heart Study. *Circulation* **117,** 743–753 (2008).

11. Pencina, M. J., D'Agostino, R. B., Larson, M. G., Massaro, J. M. & Vasan, R. S. Predicting the 30-Year Risk of Cardiovascular Disease The Framingham Heart Study. *Circulation* **119,** 3078–U61 (2009).

12. Cox, D. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* **59,** 55–95 (1972).

13. Andersen, P. K. & Gill, R. D. Cox Regression-Model for Counting-Processes - a Large Sample Study. *Annals of Statistics* **10,** 1100–1120 (1982).

14. Tsiatis, A. A. & Davidian, M. Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica* **14,** 809–834 (2004).

15. Wulfsohn, M. S. & Tsiatis, A. A. A joint model for survival and longitudinal data measured with error. *Biometrics* **53,** 330–339 (1997).

16. Rizopoulos, D. *Joint Models for Longitudinal and Time-to-Event Data with Applications in R* (Boca Raton, 2012).

17. Gelfand, A. E. & Smith, A. F. M. Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association* **85,** 398–409 (1990).

18. Prentice, R. L. A Case-Cohort Design for Epidemiologic Cohort Studies and Disease Prevention Trials. *Biometrika* **73,** 1–11 (1986).

19. Baade, P. D., Youlden, D. R. & Chambers, S. K. When do I know I am cured? Using conditional estimates to provide better information about cancer survival prospects. *Med J Aust* **194,** 73–7 (2011).

20. Janssen-Heijnen, M. L. *et al.* Prognosis for long-term survivors of cancer. *Ann Oncol* **18,** 1408–13 (2007).

21. Shack, L., Bryant, H., Lockwood, G. & Ellison, L. F. Conditional relative survival: a different perspective to measuring cancer outcomes. *Cancer Epidemiol* **37,** 446–8 (2013).

22. Nelson, C. P., Lambert, P. C., Squire, I. B. & Jones, D. R. Relative survival: what can cardiovascular disease learn from cancer? *Eur Heart J* **29,** 941–7 (2008).

Part I: Joint models in a two-phase sampling design

Chapter 1

Joint models for longitudinal and time-to-event data in a case-cohort design

Baart SJ, Boersma H, Rizopoulos D

Statistics in Medicine 2019; 38(12):2269-2281

Abstract

Studies with longitudinal measurements are common in clinical research. Particular interest lies in studies where the repeated measurements are used to predict a time-to-event outcome, such as mortality, in a dynamic manner. If event rates in a study are low, however, and most information is to be expected from the patients experiencing the study endpoint, it may be more cost efficient to only use a subset of the data. One way of achieving this is by applying a case-cohort design, which selects all cases and only a random sample of the non-cases. In the standard way of analyzing data in a case-cohort design, the non-cases who were not selected are completely excluded from analysis, however the overrepresentation of the cases will lead to bias. We propose to include survival information of all patients from the cohort in the analysis. We approach the fact that we do not have longitudinal information for a subset of the patients as a missing data problem and argue that the missingness mechanism is MAR. Hence results obtained from an appropriate model, such as a joint model, should remain valid. Simulations indicate that our method performs similar to fitting the model on a full cohort, both in terms of parameters estimates and predictions of survival probabilities. Estimating the model on the classical version of the case-cohort design shows clear bias and worse performance of the predictions. The procedure is further illustrated in data from a biomarker study on acute coronary syndrome patients, BIOMArCS.

1 Introduction

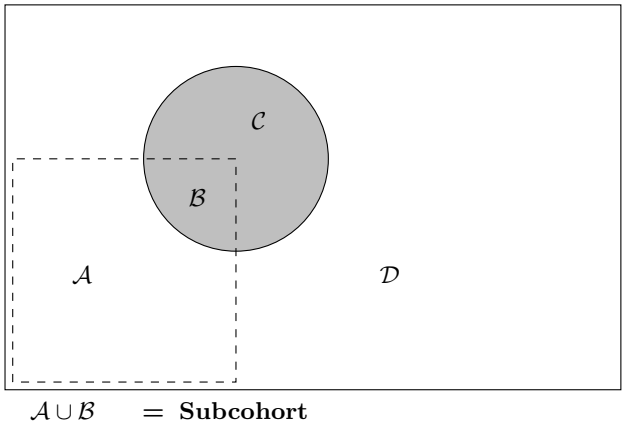
Longitudinal measurements are becoming increasingly popular in clinical research, particularly in studies where patients are followed up to an event of interest. By repeatedly collecting and analyzing measurements on patients, their progress is monitored more closely and temporal trends in the disease progress can be estimated, leading to improved prediction of outcomes. [1] In these kinds of studies two types of outcomes are collected; the longitudinal outcome (often a biomarker) and the time-to-event outcome, e.g. death. When interest lies in using temporal patterns of the longitudinal response to estimate the event of interest, both outcomes can be modeled together by using the joint modeling approach. [2] To increase prediction even further, instead of one biomarker a set of multiple markers can be measured.

The motivation for the current paper comes from the longitudinal 'BIOMarker study to identify the Acute risk of a Coronary Syndrome' (BIOMArCS), in which acute coronary syndrome (ACS) patients were examined in different medical centers in the Netherlands to study the association between (multiple) biomarkers and a recurrent ACS event (primary endpoint). [3, 4] Multiple biomarkers were identified to be of interest, measured in blood samples taken regularly during one year of follow-up. A downside of collecting multiple biomarkers is the rising costs due to the numerous biomarker measurements, since costs are associated with the ascertainment of each biomarker measured. This can cause such a project to become infeasible in practice. On top of the burden of costs, the BIOMArCS study turned out to have a low event rate, with only 5% of the patients reaching the primary endpoint. This means that the overwhelming majority of biomarker measurements belong to the censored patients where low additional information from the longitudinal patterns is expected. This gave motivation to opt for a case-cohort design, which enables analysis of the

relevant subset of patients, while largely maintaining statistical power.

In the case-cohort design [5] a random sample of patients from the full cohort is taken, defined as the subcohort ($\mathcal{A} \cup \mathcal{B}$ in Figure 1). For every patient in the full cohort the failure status is known. The complete longitudinal biomarker information, however, is only measured in the patients who experienced the study endpoint (the cases) and the random subcohort ($\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$ in Figure 1). The advantage the case-cohort design has over the more popular case-control design is that the same random subcohort can be used to study different end points. The disadvantage, and the main reason why the case-cohort design is not as popular, is that the appropriate analysis becomes more complicated. The case-cohort design is also known (early on) as "case-base design" or "hybrid-retrospective design". [5] These designs were described by Kupper, McMichael, and Spirtas (1975) and Miettinen (1982). [6, 7] Prentice (1986) was the first to introduce the design in an failure-time setting and used a pseudo-likelihood estimation approach to obtain unbiased estimates for the hazard of the event. [5] In this approach cases outside the subcohort are only included in the risk-set right before experiencing the endpoint. Other researchers followed and extended this approach by considering other types of weighting schemes. [8–13]

Motivated by BIOMArCS, the aim of our paper is twofold: first to extend the estimation framework of joint models for longitudinal and survival data in the context of case-cohort designs, and second, to assess how dynamic predictions and their accuracy perform in this setting. As mentioned above, the previously developed strategies for case-cohort designs have been based on pseudo-likelihood ideas. However, in joint models a full specification of the joint distribution of the two outcomes is required, making the use of these approaches complicated. Hence, to appropriately account for the selection bias in the case-cohort design, we approach the fact that we do not have longitudinal information for a subset



$\mathcal{A} \cup \mathcal{B}$ = **Subcohort**

$\mathcal{A}$ = **Non-cases inside subcohort**
$\mathcal{B}$ = **Cases inside subcohort**
$\mathcal{C}$ = **Cases outside subcohort**
$\mathcal{D}$ = **Non-cases outside subcohort**
$\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$ = **Case-cohort design**

Figure 1: A graphical representation of the case-cohort design

of the patients as a missing data problem. This theoretically should provide unbiased estimates if the appropriate models are used, and only requires small modification in the formulation of the likelihood of the model. With regard to our second goal, we focus on how the accuracy of dynamic predictions for the survival outcome is influenced by the case-cohort design. The evaluation is based on standard measures of predictive accuracy, such as the time-varying area under the receiver operating characteristic curves, and time-varying squared prediction errors. The remainder of the paper is organized as follows: Section 2 presents the joint model used throughout this paper. Section 3 describes the general scenario of estimating a joint model, as well as our proposed modification to avoid biased estimates in relation to the case-cohort design. Methods to measure the predictive accuracy of the models will be discussed in Section 4. A simulation study to verify our method is performed in Section 5, whereas

Section 6 shows the application to the real-life BIOMArCS data. Finally in Section 7 results will be discussed and conclusions made.

2 Model specification

We consider here a basic joint model for a continuous longitudinal outcome and a time-to-event outcome. More specifically, let $y_i(t)$ be the longitudinal measurement for the ith patient at time t. The longitudinal outcome $y_i(t)$ is modeled by a mixed effects submodel. The design vector for the fixed effects is denoted by $x_i(t)$ and the design vector for the random effects by $z_i(t)$. The time-to-event outcome is modeled by a proportional hazards submodel. Both submodels are of the form:

$$\begin{cases} y_i(t) = m_i(t) + \varepsilon_i(t) \\ \qquad\quad = x_i^\top(t)\beta + z_i^\top(t)b_i + \varepsilon_i(t) \\ \\ h_i(t) = h_0(t)\exp\{\gamma^\top w_i + \alpha m_i(t)\}. \end{cases} \tag{1}$$

The vector β in the longitudinal submodel denotes the parameters for the fixed effects and b_i the random effects for patient i, which are assumed to follow a normal distribution with mean 0 and variance-covariance matrix D. The error terms are denoted by $\varepsilon_i(t)$ and are also assumed to be normally distributed with mean 0 and variance σ^2. Real-life studies often shown nonlinear trends in the longitudinal patterns, which can be incorporated in the design vectors for the fixed and random effects parts ($x_i(t)$ and $z_i(t)$). Furthermore, let T_i^* be the true event time, C_i the censoring time, and $T_i = min(T_i^*, C_i)$ the observed event time. For each patient the event indicator is given by δ_i, taking the value of 1 when $T_i^* \leq C_i$ and 0 otherwise. Baseline covariates used in the survival

submodel are denoted by w_i. The hazard for the survival outcome (T_i, δ_i) is modeled with a proportional hazards model $h_i(t)$ defined in (1). Here we assume $m_i(t)$ is the true and unobserved value of longitudinal outcome for patient i at time t, modeled by the longitudinal submodel. The baseline hazard is given by $h_0(t)$ and is modeled in a flexible manner by B-splines. Finally, α denotes the association between the longitudinal and time-to-event outcome.

3 Estimation

Bayesian estimation in a standard full cohort

In this study the Bayesian framework will be used for estimation. The parameters of the model will be estimated using Markov Chain Monte Carlo (MCMC) methods. The contribution of patient i to the posterior distribution of the joint model is defined as

$$p(\theta, b_i \mid T_i, \delta_i, y_i) \propto p(T_i, \delta_i \mid b_i, \theta) p(y_i \mid b_i, \theta) p(b_i \mid \theta) p(\theta),$$

where θ denotes the vector of all parameters. The contribution of patient i to the likelihood of the survival submodel is written as

$$\begin{aligned} p(T_i, \delta_i \mid b_i, \beta, \theta_t) &= h_i\{T_i \mid \mathcal{M}_i(T_i), \theta_t\}^{\delta_i} S_i\{T_i \mid \mathcal{M}_i(T_i), \theta_t\} \\ &= [h_0(T_i \mid \gamma_s) \exp\{\gamma^\top w_i + \alpha m_i(T_i)\}]^{\delta_i} \times \\ &\quad \exp\left\{ - \int_0^{T_i} h_0(s \mid \gamma_s) \exp\{\gamma^\top w_i + \alpha m_i(s)\} ds \right\}, \end{aligned}$$

where $\theta_t = (\gamma_s, \gamma, \alpha)$ and $m_i(t) = x_i^\top(t)\beta + z_i^\top(t) b_i$. Additionally, $\mathcal{M}_i(T_i)$ denotes the complete history of longitudinal marker for patient i. The contribution of patient i to the likelihood of the longitudinal submodel is given by

$$p(y_i \mid b_i, \theta_y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ - \frac{\sum_{j=1}^{n_i}(y_{ij} - x_{ij}^\top \beta - z_{ij}^\top b_i)^2}{2\sigma^2} \right\},$$

with $\theta_y = (\beta, \sigma)$ and $\theta = (\theta_y^\top, \theta_t^\top)^\top$.

Uninformative normal priors are used for the β, γ and α parameters, as well as the parameters for the B-splines in the baseline hazard (γ_s). For the elements of the variance-covariance matrix of the random effects (D) an inverse Wishart prior is used and a gamma prior is used for the variance of the errors of the longitudinal outcome (σ^2). Initial values for the parameters of the prior distribution are obtained from estimations based on fitting the longitudinal and time-to-event submodels separately. The joint models are analyzed with JAGS software, using Gibbs sampling to execute the MCMC methods.

Bias in a case-cohort design

If a study follows a case-cohort design, estimation with the above mentioned standard likelihood will result in bias, due to the outcome dependent missingness in the data. The bias occurs, because in the case-cohort design only a selection of the censored or non-event patients is used in the analysis, along with all the event patients. As a consequence the event rate in the case-cohort is higher than the event rate in the original full cohort.

In a standard full cohort the observed data is $\mathcal{F}_n = \{y_i, T_i, \delta_i; i = 1, ..., n\}$ and is fully observed for each patient. In the case-cohort design, additionally we have S_i as the indicator for the randomly drawn subcohort with a pre-specified size z (e.g., $z = 1/3$) ($\mathcal{A} \cup \mathcal{B}$ in Figure 1) and CC_i denoting the indicator for

being included in the case-cohort design ($\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$ in Figure 1), whereby

$$CC_i = \begin{cases} 1 & \text{if } \delta_i = 1 \text{ or } S_i = 1, \\ \\ 0 & \text{if } \delta_i = 0 \text{ and } S_i = 0 \end{cases}$$

or $CC_i = \delta_i + (1 - \delta_i)S_i$. The full set of observed data is now $\mathcal{F}_n = \{S_i, CC_i, y_i, T_i, \delta_i; i = 1, \ldots, n\}$. There are four distinct groups a patient in the case-cohort design can belong to as defined in Figure 1. In each group the following data is collected

$$\begin{aligned} \mathcal{A} &= \{S_i = 1, CC_i = 1, y_i^o, T_i, \delta_i = 0\}, \\ \mathcal{B} &= \{S_i = 1, CC_i = 1, y_i^o, T_i, \delta_i = 1\}, \\ \mathcal{C} &= \{S_i = 0, CC_i = 1, y_i^o, T_i, \delta_i = 1\}, \\ \mathcal{D} &= \{S_i = 0, CC_i = 0, y_i^m, T_i, \delta_i = 0\}, \end{aligned}$$

where y_i^o are the observed longitudinal measurements and y_i^m the unascertained longitudinal measurements. In the standard version of the case-cohort design, only patients belonging to $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$ are included in the analysis. CC_i can be seen as selection indicator and the missing data in the case-cohort design (patients in $\mathcal{D}$) can be interpreted as missing due to selection bias. Since these missings depend on unobserved data, the missing data mechanism will be missing not at random (MNAR). The different event rates between the full cohort and the case-cohort design will result in a misspecification of the baseline hazard. This, in turn will lead to bias both in the estimation of the parameters of the model and the estimation of survival probabilities.

Unbiased estimation using survival information from entire cohort

The bias caused by the outcome-dependent missings can be circumvented by utilizing the survival information of the entire cohort, which has to be available due to the nature of the case-cohort design, as argued by Dong and colleagues. [14] Since the random subcohort ($\mathcal{A} \cup \mathcal{B}$) is supplemented with the remaining cases outside the random subcohort ($\mathcal{C}$), it follows that the patients left out are all event-free and therefore censored patients ($\mathcal{D}$).

If all survival information is used in the analysis, the missing data only comes from missing longitudinal measurements in $\mathcal{D}$. In this case these missing values are missing depending on observed information (survival status) and are therefore missing at random (MAR). The probability that the longitudinal response is missing, which is the same as the probability that the patient belongs to group $\mathcal{D}_i$, can be written as

$$p(\mathcal{D}_i \mid \delta_i, y_i^o, y_i^m, \psi) = p(\mathcal{D}_i \mid \delta_i, \psi), \tag{2}$$

where ψ is the vector of parameters describing the missingness model. In the version of the case-cohort design used throughout this manuscript, this is simply the probability of not being drawn by the random subcohort ($p = 1 - z$). To obtain unbiased estimates for the joint model we have to estimate the full distribution of all processes, including $\mathcal{D}_i$. When the complete survival information is taken into account (so patients in $\mathcal{D}$ are included in the analysis), the full distribution can be decomposed as

$$p(T_i, \delta_i, y_i^o, \mathcal{D}_i \mid b_i, \theta, \psi) = \int p(T_i, \delta_i, y_i^o, y_i^m \mid b_i, \theta) \times p(\mathcal{D}_i \mid b_i, \delta_i, y_i^o, y_i^m, \psi) dy_i^m.$$

Under (2) this becomes

$$p(T_i, \delta_i, y_i^o, \mathcal{D}_i \mid b_i, \theta, \psi) = p(T_i, \delta_i, y_i^o \mid b_i, \theta) \times p(\mathcal{D}_i \mid b_i, \delta_i, \psi). \tag{3}$$

Because of the decomposition, the distribution of CC_i does not depend on y_i^m but only on observed data δ_i. Additionally, since ψ and δ are distinct, the missing data caused by $\mathcal{D}_i$ is ignorable and analysis on the observed data gives unbiased results. This decomposition does not hold when patients in $\mathcal{D}$ are excluded from the analysis, where as a result $\mathcal{D}_i$ depends on unobserved data.

In the newly proposed version of the case-cohort design, all patients will be included in the analysis, but not all patients supply the same amount of information. The posterior distribution stated earlier, will be different for certain patients. For the patients in the case-cohort design ($CC_i = 1$), all information is available and the posterior distribution remains equal. For the censored patients outside the subcohort ($CC_i = 0$), the longitudinal information is not measured and therefore missing. However, the values are imputed by the model and the posterior distribution of longitudinal submodel is replaced by imputed values (y_i^m). The values are based on the posterior predictive distribution of the missing data, which is

$$p(y_i^m \mid T_i, \delta_i = 0, \mathcal{F}_n) = \int p(y_i^m \mid T_i, \delta_i = 0, \theta) p(\theta \mid \mathcal{F}_n) d\theta,$$

where the first term of the integral can be expressed as

$$p(y_i^m \mid T_i, \delta_i = 0, \theta) = \int p(y_i^m \mid b_i, \theta) p(b_i \mid T_i, \delta_i = 0, \theta) db_i$$

Based on the observed data and averaged over the posterior distribution of the parameters and random effects estimated by the model, this distribution is

available. For each patient, the missing values of y can be obtained directly, and this occurs during estimation of the model. Aside from the survival information, any available covariate measurements taken on baseline can also be included for these patients. The posterior distribution for all patients in the cohort will therefore be given by

$$p(\theta, b_i, y_i^m \mid T_i, \delta_i, y_i^o) \propto \begin{cases} p(T_i, \delta_i \mid b_i, \theta)\, p(y_i^o \mid b_i, \theta)\, p(b_i \mid \theta)\, p(\theta) & \text{if } CC_i = 1, \\ p(T_i, \delta_i \mid b_i, \theta)\, p(y_i^m \mid b_i, \theta)\, p(b_i \mid \theta)\, p(\theta) & \text{if } CC_i = 0. \end{cases}$$

4 Predictive performance

In clinical studies it is often of interest to use the estimated model to predict survival probabilities for (a) new patient(s). Therefore we need to assess the performance of the model in terms of predictive accuracy of the survival outcome. In general, a joint model fitted on the data sample $\mathcal{F}_n = \{T_i, \delta_i, y_i; i = 1, \ldots, n\}$ is used to make survival predictions for a new patient j, with longitudinal measurements $(\mathcal{Y}_j(t))$ up to time t. The information that the new patient provided longitudinal measurements up to t, is used to postulate that the patient was event free at t and interest lies in events taking place in a medically relevant time interval $(t, t + \Delta t]$. The probability that the patient survives this time window is

$$\pi_j(t + \Delta t \mid t) = \Pr(T_j^* \geq t + \Delta t \mid T_j^* > t, \mathcal{Y}_j(t), \mathcal{F}_n). \tag{4}$$

This probability can be estimated based on the posterior predictive distribution given by

$$\pi_j(t + \Delta t \mid t) = \int P(T_j^* \geq t + \Delta t \mid T_j^* > t, \mathcal{Y}_j(t), \theta) p(\theta \mid \mathcal{F}_n) d\theta,$$

where the first part of the integrand can be rewritten as

$$\begin{aligned} P(T_j^* \geq t + \Delta t \mid T_j^* > t, \mathcal{Y}_j(t), \theta) &= \int P(T_j^* \geq t + \Delta t \mid T_j^* > t, b_j, \theta) \\ &\quad p(b_j \mid T_j^* > t, \mathcal{Y}_j(t), \theta) db_j \\ &= \int \frac{S_j\{t + \Delta t \mid \mathcal{M}_j(t + \Delta t, b_j), \theta\}}{S_j\{t \mid \mathcal{M}_j(t, b_j), \theta\}} \\ &\quad p(b_j \mid T_j^* > t, \mathcal{Y}_j(t), \theta) db_j. \end{aligned}$$

Based on these equations and the posterior distribution of the parameters for the original data $\mathcal{F}_n$ obtained by the MCMC samples, Monte Carlo estimates of $\pi_j(t + \Delta t \mid t)$ can be obtained by a new simulation scheme. More details on this procedure can be found in Rizopoulos. [2, 15]

In this paper we will assess the accuracy of the predictions in terms of discrimination and calibration. A model shows good discrimination if the estimated longitudinal biomarker profile can discriminate well between patients with and without the study endpoint. A model is calibrated well if the estimated longitudinal patterns can predict a future endpoint with high accuracy. In the situation of a case-cohort design, the data used to fit the joint model is $\mathcal{F}_n = \{S_i, CC_i, T_i, \delta_i, y_i; i = 1, ..., n\}$, where for a set of the patients y_i is missing, as discussed earlier. For these patients $\mathcal{Y}_j(t)$ is not observed and therefore the corresponding survival probability in (4) can not be estimated. In this paper the predictive measures will be calculated only on patients from the random subcohort ($S_i = 1$), so the event rate corresponds to the full cohort while no

missing data occurs in the patients. To assess the discrimination of the model, the area under the ROC curve (AUC) can be estimated, using longitudinal information up to time t for a new (set of) patient(s) and then calculate the AUC up to Δt.

With c in $[0, 1]$, a patient is labeled as event-free if $\pi_j(t + \Delta t \mid t) > c$ and as experiencing the endpoint if $\pi_j(t + \Delta t \mid t) \leq c$. The AUC, calculated for a pair of randomly chosen patients $\{i, j\}$ is therefore

$$\text{AUC}(t, \Delta t) = \Pr[\pi_i(t+\Delta t \mid t) < \pi_j(t+\Delta t \mid t) \mid \{T_i^* \in (t, t+\Delta t]\} \cap \{T_j^* > t+\Delta t\}].$$

This means that we would assign a higher survival probability to patient j than to patient i, if patient i experiences the endpoint in the time window $t + \Delta t$ and patient j does not.

However, since T_i^* is not observed for all patients due to censoring, this equation cannot be solved directly. Therefore the estimated AUC is decomposed as

$$\widehat{\text{AUC}}(t, \Delta t) = \widehat{\text{AUC}}_1(t, \Delta t) + \widehat{\text{AUC}}_2(t, \Delta t) + \widehat{\text{AUC}}_3(t, \Delta t) + \widehat{\text{AUC}}_4(t, \Delta t). \quad (5)$$

The first part ($\widehat{\text{AUC}}_1(t, \Delta t)$) refers to the pairs without censoring, so for which the event times can be ordered directly, and the remaining parts refer to the patient pairs where censoring occurs. [15] The full specification of the AUC is given in the supplemental material.

The calibration of the model is measured by the prediction error (PE), where based on all available information of a patient j, the estimated survival probability ($\pi_j(t+\Delta t \mid t)$) is compared to the observed survival ($I(T_j^* > t+\Delta t)$).

The expected prediction error is then as follows

$$\mathrm{PE}(t + \Delta t \mid t) = E[\{I(T_j^* > t + \Delta t) - \pi_j(t + \Delta t \mid t)\}^2].$$

Lower values of PE indicate smaller differences between the observed and predicted survival and therefore a better calibrated model. An appropriate estimator for time-to-event data is proposed by Henderson et al. (2002) [16] and is given in the supplemental material.

For the real life application, an internal validation of the model was applied to evaluate the predictive performance of the model. [17] Since the same data is used for fitting the model and evaluating the performance of the model, optimistic predictions can occur. This holds particular importance when the data set is small. In this paper corrections for the optimism will be done by a bootstrap method developed by Harrell. [18] This method works in several steps.

1. First, fit the model on the data and calculate the apparent predictive measures (here the AUC and PE), denoted by AUC_{app} and PE_{app}.

2. Take a bootstrap sample of the data. Refit the model on the bootstrap sample and calculate the apparent predictive measures, denoted by $\mathrm{AUC}_{b,boot}$ and $\mathrm{PE}_{b,boot}$.

3. Thirdly, calculate the predictive measures on the original data from the model fitted on the bootstrap sample, called $\mathrm{AUC}_{b,orig}$ and $\mathrm{PE}_{b,orig}$.

4. Then, calculate the optimism in this bootstrap sample by $O_{\mathrm{AUC,b}} = \mathrm{AUC}_{b,boot} - \mathrm{AUC}_{b,orig}$ and $O_{\mathrm{PE,b}} = \mathrm{PE}_{b,boot} - \mathrm{PE}_{b,orig}$.

5. Repeat steps 2-4 B times. Harrell recommends to use a B between 100-200.

6. After the optimism is calculated for all B bootstrap samples, correct the apparent predictive measure with each optimism ($\text{AUC}_{cor,b} = \text{AUC}_{app}$ - $\text{O}_{\text{AUC,b}}$ and $\text{PE}_{cor,b} = \text{PE}_{app} + O_{\text{PE,b}}$).

7. In the last step, take the average of all these corrected predictive measures to obtain the for optimism adjusted AUC and PE ($\text{AUC} = B^{-1} \sum_B \text{AUC}_{cor,b}$ and $\text{PE} = B^{-1} \sum_B \text{PE}_{cor,b}$). Additionally the 2.5% and 97.5% percentiles of the bootstrapped samples can be obtained as an indication of the spread of the estimator.

5 Simulation study

Design

A simulation study was carried out to verify that the proposed model results in unbiased estimates and shows good predictive performance. Data sets representing the full-cohort were simulated and from these data sets a case-cohort design was imitated by drawing a random set of patients and supplementing the cases to this. The submodel for the simulated longitudinal outcome is defined as

$$y_i(t) = \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 \text{G}_i + b_{1i} + b_{2i} t + b_{3i} t^2 + \varepsilon_i(t), \tag{6}$$

where the β's define the average population trajectory, the b's subject-specific deviations from this trajectory and are assumed to be normally distributed ($b_i \sim \mathcal{N}(0, D)$). The variance-covariance matrix of the random effects (D) is left unstructured. G is a binary covariate, drawn from a binomial distribution with probability 0.5. A quadratic term for time was added to the fixed and random effects to imitate non-linear trajectories often found in real-life longitudinal

studies. The survival times are generated by

$$h_i(t) = h_0(t) \exp\{\gamma \mathrm{G}_i + \alpha m_i(t)\}. \tag{7}$$

Here $m_i(t)$ is assumed to be the true longitudinal outcome at time t. The baseline hazard $h_0(t)$ was generated with a Weibull distribution with a shape parameter (ϕ) of 2. The scale of the Weibull model is $\exp\{\gamma \mathrm{G}_i + \alpha m_i(t)\}$ and the hazard function can therefore also be written as $h_i(t) = h_0(t) \exp\{\gamma \mathrm{G}_i + \alpha m_i(t)\} = \phi t^{\phi-1} \exp\{\gamma G_i + \alpha m_i(t)\}$. The association parameter α was set equal to 1. The remaining parameter settings were: $\beta_1 = 1$, $\beta_2 = 0.3$, $\beta_3 = 0.1$, $\beta_4 = 0.1$, γ = -2, $\sigma^2 = 1$. Data sets were simulated with 2000 subjects and 25 planned measurements per subject. The mean of the exponential distribution for the censoring mechanism varied and the maximum follow-up time was 15.

Analysis

Two versions of the case-cohort design were generated from the simulated data sets. In the first version, the survival information of all patients was retained and only the biomarker values for the unselected patients were put to missing. The second version (also called the classical case-cohort) only uses information from the patient in the case-cohort design, and completely removes the remaining patients for analyses. The same joint model was fitted on all three data sets, where the results from the full cohort were viewed as the golden standard. Four different scenarios with varying event rates and varying sizes of the random subcohort were simulated 200 times. In scenario 1 the the mean value of censoring time was set at 3.2 and the coefficient of the intercept of the Weibull regression at -7.5, which resulted in an 20% event rate. Here, 1/3 of the cohort was randomly sampled as subcohort. In scenario 2 the event rate was kept at 20%, but now the size of the subcohort was 1/6 of the full cohort.

For scenario 3 and 4 the event rate was set to 5% using a mean censoring time of 2.5 and an intercept coefficient of -9.5. The sizes of the random subcohort in scenario 3 and 4 were 1/3 and 1/6, respectively. For the predictive performances of the models a validation data set was simulated with 1000 subjects using the same scenario as the data on which the model was fitted. Time-dependent AUC and PE were calculated on two intervals during follow up, where the intervals depended on the simulation scenario.

Results

Table 1 shows the characteristics of the simulated data in the four different scenarios. Apart from the number of biomarker measurements, the dimensions of the data sets for the full cohort (FC) and the case-cohort (CCI) are the same. In the classical case-cohort design (CCII) additionally the number of patients and event rate differ from the full cohort. It is clear that a different event rate, together with the size of the drawn subcohort, has a large impact on the size of the remaining case-cohort data set. For scenario 4, the resulting event rate in the classical case-cohort data set is 5 times as high (25%) as it was in the full cohort. The results of the model estimation are shown in Table 2. For each scenario the association parameter (α) is given, along with the bias (the difference between the mean estimate of the simulation and the simulated parameter value) and the coverage rate. The coverage rate is calculated as the percentage of times the true simulated value of α falls in the credible interval of each simulation. For all four scenarios the bias of α in the CCI is small and close to the estimate of α based on the full cohort (the difference between mean α_{FC} and $\alpha_{CCI} \leq 0.023$). This is also the case for the coverage rate, which is similar for the FC and the CCI. The CCII, on the other hand, shows a clear downward bias (mean bias between 0.15-0.35) and low coverage rates between

0% and 13%. For the scenario's with a low event rate, all three models give an underestimation of the true parameter value of α, however the FC and CCI give similar performances compared to CCII. Table 2 additionally shows the estimated parameters of the longitudinal submodel (β's), and the parameter of the survival submodel (γ). These parameters indicate the same results; the estimates for the FC and the CCI are very similar and clear bias is found for the CCII. The bias, percentiles and coverage rates of these parameters can be found in the supplemental material.

The performance of the predictive accuracy of the models is assessed by evaluating the AUC and PE on two different time points during the simulation follow-up. The time points depend on the follow-up time in the data and can therefore differ per scenario. The outcomes are shown for scenario 2 by the boxplots in Figure 2. The boxplots for the other scenarios can be found in the supplemental material. The CCI performs very similar compared to the FC in terms of predictive accuracy, however only slightly worse (as demonstrated by a smaller AUC and a higher PE). The CCII analysis demonstrates a decidedly worse performance in prediction, particularly in terms of calibration. The other scenarios show a similar result, although less pronounced.

An additional simulation study was performed to evaluate the method in smaller data sets ($n = 500$). The results can be found in the supplemental material and are in line with the other simulations.

Table 1: Characteristics of the simulated data sets based on 200 replications of each scenario.

% Events		Scenario	Size subcohort: 1/3			Scenario	Size subcohort: 1/6		
			FC	CCI	CCII		FC	CCI	CCII
20%	patients, n	1	2000	2000	900	2	2000	2000	700
	events, n		400	400	400		400	400	400
	event rate, %		20%	20%	40%		20%	20%	60%
	measurements, n		15,000	7000	7000		19,000	6000	6000
5%	patients, n	3	2000	2000	700	4	1900	1900	400
	events, n		100	100	100		100	100	100
	event rate, %		5%	5%	15%		5%	5%	25%
	measurements, n		11,000	4500	4500		9000	2000	2000

FC, Full cohort; CCI, Case-cohort design, retain all survival information; CCII, Case-cohort design, classical version

Table 2: Results from estimating a joint model on simulated data based on 200 replications per scenario.

% Events		Scenario	Size subcohort: 1/3			Scenario	Size subcohort: 1/6		
			FC	CCI	CCII		FC	CCI	CCII
20%	α	1	0.975	0.971	0.849	2	0.976	0.966	0.799
	bias		-0.025	-0.029	-0.151		-0.024	-0.034	-0.201
	(2.5%-97.5%)		(0.89-1.07)	(0.88-1.07)	(0.76-0.94)		(0.89-1.07)	(0.88-1.06)	(0.71-0.89)
	coverage		92%	91%	13%		92%	88%	4%
	β_1		1.003	0.996	1.087		1.004	0.986	1.139
	β_2		0.319	0.331	0.558		0.324	0.357	0.713
	β_3		0.110	0.104	0.142		0.109	0.097	0.154
	β_4		0.104	0.105	0.092		0.102	0.099	0.092
	γ		-1.979	-1.987	-1.774		-1.979	-1.978	-1.676
5%	α	3	0.856	0.845	0.727	4	0.858	0.835	0.649
	bias		-0.144	-0.155	-0.273		-0.142	-0.165	-0.351
	(2.5%-97.5%)		(0.74-0.99)	(0.72-0.98)	(0.61-0.86)		(0.74-0.99)	(0.71-0.97)	(0.53-0.78)
	coverage		38%	33%	1%		39%	32%	0%
	β_1		1.003	0.993	1.062		1.005	0.990	1.127
	β_2		0.331	0.343	0.474		0.334	0.371	0.638
	β_3		0.108	0.099	0.127		0.106	0.087	0.146
	β_4		0.101	0.103	0.055		0.100	0.107	0.023
	γ		-2.730	-2.760	-2.421		-2.771	-2.806	-2.238

The *bias* indicates the difference between the simulated parameter value and the estimated value by each of the models. The *coverage* is calculated by the percentage of times the true simulated values falls in the credible interval of each simulation.

Simulated values of the parameters: $\alpha = 1$, $\beta_1 = 1$, $\beta_2 = 0.3$, $\beta_3 = 0.1$, $\beta_4 = 0.1$, $\gamma = -2$.

FC, Full cohort; CCI, Case-cohort design - retain all survival information; CCII: Case-cohort design - classical version

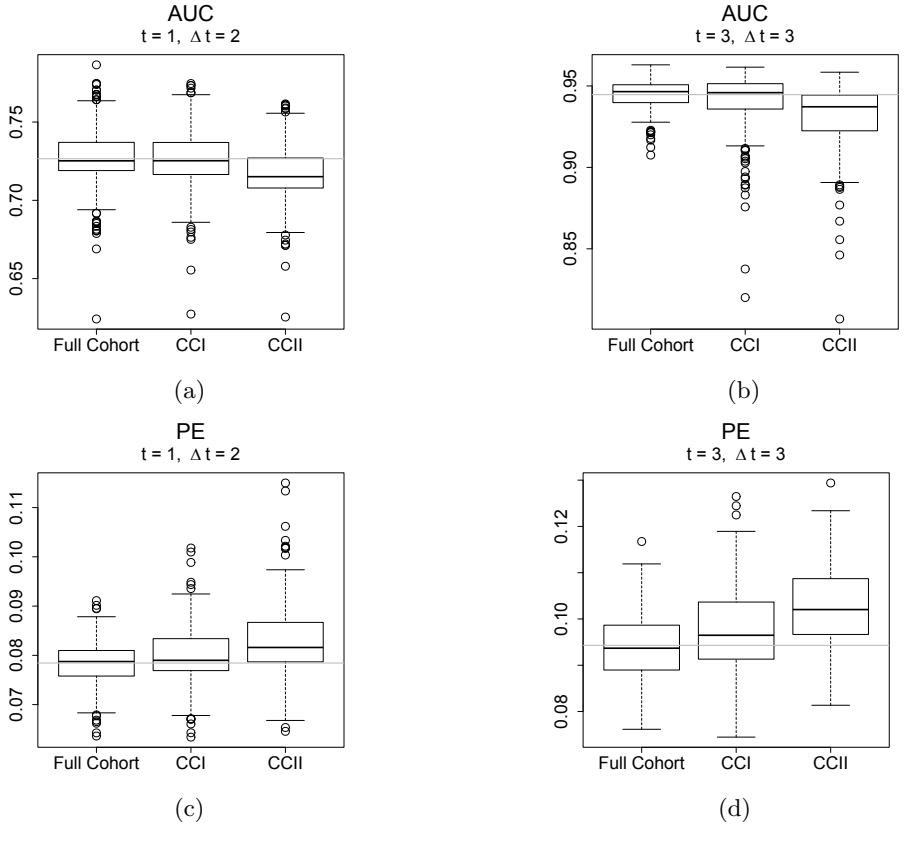


Figure 2: Predictive accuracy measures from scenario 2 (Event rate: 20% - size subcohort: 1/6)

6 Application to BIOMArCS study

Study design

We illustrate the use of our findings on data from the BIOMArCS study. In this multi-center study patients admitted for acute coronary syndrome (ACS) at several Dutch hospitals in the Netherlands were enrolled between January 1, 2008 and September 1, 2014. Patient follow-up ended at September 1, 2015. Patients were followed for the first year after their initial cardiac event. They

were invited back to the hospital on regular occasions, where blood samples were collected. The first blood sample was collected during hospitalization for the index event. Subsequent blood samples were collected every two weeks for the first six months of follow up and once a month during the last six months of follow up. The goal of BIOMArCS was to study the association between longitudinal patterns of multiple biomarkers and the primary endpoint. In total 839 patients were included with a median of 17 blood samples per patient. The primary endpoint was a composite of cardiovascular mortality, non-fatal acute coronary syndrome or unplanned coronary revascularization due to progressive angina pectoris during 1-year follow-up. In total 45 patients were identified as having the primary end point (5.4% of the entire cohort). The low event rate combined with the high number of biomarker measurements, led to the decision to only ascertain biomarker values in a subset of the patients using the case-cohort design. A random sample of 150 patients was selected ($\mathcal{A} \cup \mathcal{B}$ in Figure 1). Of these, 142 patients were event free at the end of follow-up and 8 patients had experienced the primary end point. The subcohort of 150 was supplemented with the remaining 37 event patients outside the subcohort ($\mathcal{C}$ in Figure 1) reaching a total of 187 patients in the case-cohort design.

Analysis BIOMArCS

It is of interest to model how strongly *Cardiac Troponin-I* (TnI), a well established cardiovascular biomarker, [19] is related to the hazard of the primary endpoint. The distribution of TnI is heavily skewed, so a log_2 transformation was applied. On top of that, the TnI values were transformed to z-scores, for potential head-to-head comparison between different biomarkers. Patients showed nonlinear evolutions due to a stabilization period after the index event,

which were modeled by a piecewise linear regression model, with the breakpoint at 30 days. The longitudinal submodel used to fit TnI on the BIOMArCS data is of the form

$$z\text{TnI}_i(t) = \beta_1 + \beta_2 t + \beta_3 (t-30)_+ + \beta_4 \text{Sex}_i + b_{1i} + b_{2i} t + b_{3i}(t-30)_+ + \varepsilon_i(t), \tag{8}$$

where $(\cdot)_+$ denotes $(A)_+ = A$ if $A > 0$ and 0 elsewhere. Sex is a covariate that denotes the gender (1 = female and 2 = male) of the patient. The variance-covariance matrix of the random effects (D) is left unstructured. The survival submodel is given by

$$h_i(t) = h_0(t) \exp\{\gamma \text{Sex}_i + \alpha m_i(t)\}. \tag{9}$$

The baseline hazard $h_0(t)$ is modeled with cubic B-splines, with 5 knots placed based on the percentiles of the observed event times (67, 338, 359, 368 and 382 days). Since the full cohort is unknown in the BIOMArCS data, for this application we can only estimate and compare the two versions of the case-cohort design. The predictive performance of the models is again assessed by calculating the area under the ROC curve (AUC) and prediction error (PE). These measures are calculated on a subset of the data that consists only of the random subcohort ($S_i = 1$), because in this subcohort the event rate is equal to the event rate in the full cohort and longitudinal measurements are available for all patients. A downside of using this subset of the data is that the random subcohort only has 8 endpoints, which can lead to unstable estimates of the predictive accuracy. For the calculation of the AUC and PE, longitudinal information from the first 60 days was used to calculate the respective diagnostic measurements at time 100 ($\Delta t = 40$ days). This interval was chosen by the

distribution of the event times of the 8 events in the BIOMArCS subcohort. To account for the fact that these validation measures are estimated on the same data set as the model was developed, they are corrected with Harrell's optimism measure using the bootstrap method. [18]

Results BIOMArCS

Applying a case-cohort design to the BIOMArCS data has a large consequence on the number of patients used in the analyses. In the full cohort and therefore also in newly proposed version of the case-cohort design (again denoted by CCI), there were 839 patients, where the classical case-cohort design (denoted by CCII) only uses 187 patients. This also leads to a substantial difference in event rate which is 24% in CCII, compared to 5% in CCI. Both versions of the case-cohort design use 1492 TnI measurements and additionally in CCI there is a large number of missing TnI values (9829) corresponding to the unascertained TnI measurements from the patients outside the case-cohort design. The results from the model estimates are presented in Table 3. The parameter estimates are very similar for both models. The α parameter, denoting the association between the longitudinal marker TnI and the composite endpoint, is 0.30 (95% credible interval: 0.10 - 0.50) and 0.33 (95% credible interval: 0.14 - 0.53) for the new and classical case-cohort design respectively. The remaining parameters are also very similar. The predictive accuracy measures, corrected for optimism, are presented in the last part of Table 3. CCI performs slightly better in predicting new events by showing larger AUC (0.551 vs 0.533) and smaller PE (0.014 vs 0.017).

Table 3: Results from estimating a joint model for repeated TnI values and the combined study endpoint on two versions of the case-cohort design in the BIOMArCS data.

		CCI		CCII	
Longitudinal submodel		*Mean*	*95% CI*	*Mean*	*95% CI*
β_1	Intercept	8.87	(7.98, 9.66)	8.98	(8.26, 9.78)
β_2	Slope ($t < 30$ days)	-6.35	(-7.15, -5.56)	-6.34	(-7.07, -5.63)
β_3	Δ Slope($t < 30$, $t \geq 30$)	-6.77	(-7.55, -5.97)	-6.76	(-7.46, -6.08)
β_4	Sex	0.54	(0.15, 0.93)	0.48	(0.11, 0.88)
Survival submodel		*Mean*	*95% CI*	*Mean*	*95% CI*
α	Association	0.30	(0.10, 0.50)	0.33	(0.14, 0.53)
γ	Sex, survival	-0.43	(-1.04, 0.21)	-0.44	(-1.07, 0.15)
Predictive accurary		*Estimate*	*(2.5% - 97.5%)*	*Estimate*	*(2.5% - 97.5%)*
AUC	$t = 60, \Delta t = 40$	0.551	(0.420 - 0.695)	0.533	(0.438 - 0.633)
PE	$t = 60, \Delta t = 40$	0.014	(0.007 - 0.031)	0.017	(0.011 - 0.032)

β_3 indicates the difference between the slope estimates before and after 30 days. The coefficient for the slope after 30 days is given by $(\beta_2 + \beta_3)$.
The AUC and PE are calculated using longitudinal measurements up to $t = 60$ (days) to predict events in (60, 100]. The measures are corrected with Harrell's optimism and shown with the 2.5% and 97.5% confidence limits.
CC, Case-cohort design - retain all survival information; CCII, Case-cohort design - classical version; CI, Credible Interval; AUC, Area under the ROC curve; PE, Prediction error

7 Discussion

Longitudinal studies following patients over time are becoming increasingly more popular in clinical research, since they can incorporate dynamic patterns reflecting disease progress and thus improve prediction of events. If longitudinal studies are extended further to include multiple markers, different aspects of the disease can be modeled, which in turn leads to additional improvement of the model. A severe downturn is the increasing costs associated with ascertaining large numbers of biomarker measurements. To ensure practical use of these studies, new methods are necessary so that unbiased results and optimal efficiency

are warranted when only utilizing a subset of the measurements. A case-cohort design can help in cost reduction, by measuring all patients who experienced the study endpoint and only a subset of the patients without the endpoint. However, the overrepresentation of the cases causes bias, interpreted as selection bias, in estimation of the model parameters and when predictions for a new patient are made. By incorporating survival information of all patients, the problem is solved and models will show unbiased estimations. The simulation study we performed, showed that by incorporating all survival information, the case-cohort design performs very similar to the full cohort in terms of unbiased estimation and predictive accuracy. When the classical case-cohort is applied for comparison, in general, the model will show biased estimates and worse predictive accuracy.

The difference in estimates between the two versions of the case-cohort design however, was not found in the real-life application. Possibly, this is due to the smaller size of association parameter in the BIOMArCS study (0.3), compared to value of the parameter in the simulated data (which was 1). The difference in event rate also had a modest impact on predicting new events as shown by the corrected predictive accuracy methods. The newly proposed version of the case-cohort design performed slightly better in terms of discrimination and calibration than the classical case-cohort design. It should be noted however, that, although corrected for optimism, these measures were calculated on a subset of the data with only eight events (the random subcohort). New methods are necessary to incorporate the complete survival information in these functions in a similar manner as we incorporated them in the model estimation.

The findings throughout this paper combined, we can conclude that for studies with large amounts of longitudinal measurements, costs can be saved while results remain reliable, by applying a case-cohort design and incorporating

the survival information from the complete cohort in the models. This work can be extended to find the optimal selection of longitudinal measurements taken while retaining unbiased estimates and high values of predictive accuracy and developing new methods to efficiently estimate the predictive accuracy.

References

1. Van Vark, L. C. *et al.* Prognostic Value of Serial ST2 Measurements in Patients With Acute Heart Failure. *J Am Coll Cardiol* **70,** 2378–2388 (2017).
2. Rizopoulos, D. *Joint models for longitudinal and time-to-event data: With applications in R* (CRC Press, 2012).
3. Oemrawsingh, R. M. *et al.* Cohort profile of BIOMArCS: the BIOMarker study to identify the Acute risk of a Coronary Syndrome-a prospective multicentre biomarker study conducted in the Netherlands. *BMJ Open* **6,** e012929 (2016).
4. Oemrawsingh, R. M. *et al.* High-Frequency Biomarker Measurements of Troponin, NT-proBNP, and C-Reactive Protein for Prediction of New Coronary Events After Acute Coronary Syndrome. *Circulation* **139,** 134–136 (2019).
5. Prentice, R. L. A Case-Cohort Design for Epidemiologic Cohort Studies and Disease Prevention Trials. *Biometrika* **73,** 1–11. ISSN: 0006-3444 (1986).
6. Kupper L. L.; McMichael, A. J. & Spirtas, R. A Hybrid Epidemiologic Study Design Useful in Estimating Relative Risk. *J Am Stat Assoc* **70,** 524–528 (1975).
7. Miettinen, O. Design options in epidemiologic research. An update. *Scand J Work Environ Health* **8 Suppl 1,** 7–14 (1982).
8. Barlow, W. E. Robust Variance-Estimation for the Case-Cohort Design. *Biometrics* **50,** 1064–1072 (1994).
9. Barlow, W. E., Ichikawa, L., Rosner, D. & Izumi, S. Analysis of case-cohort designs. *J Clin Epidemiol* **52,** 1165–1172 (1999).
10. Kalbfleisch, J. D. & Lawless, J. F. Likelihood analysis of multi-state models for disease incidence and mortality. *Stat Med* **7,** 149–60 (1988).
11. Lin, D. Y. & Ying, Z. Cox Regression with Incomplete Covariate Measurements. *J Am Stat Assoc* **88,** 1341–1349 (1993).
12. Self, S. G. & Prentice, R. L. Asymptotic-Distribution Theory and Efficiency Results for Case Cohort Studies. *Ann Stat* **16,** 64–81 (1988).
13. Nan, B., Yu, M. G. & Kalbfleisch, J. D. Censored linear regression for case-cohort studies. *Biometrika* **93,** 747–762 (2006).
14. Dong, X., Kong, L. & Wahed, A. S. Accelerated failure time model for case-cohort design with longitudinal covariates subject to measurement error and detection limits. *Stat Med* **35,** 1327–39 (2016).

15. Rizopoulos, D., Molenberghs, G. & Lesaffre, E. M. E. H. Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biom J* **59,** 1261–1276 (2017).

16. Henderson R.; Diggle, P. & Dobson, A. Identification and efficacy of longitudinal markers for survival. *Biostatistics* **3,** 33–50 (2002).

17. Harrell, F. E. *Regression Modeling Strategies* (Springer, 2001).

18. Harrell, F. E. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* **15,** 361–87. ISSN: 0277-6715 (1996).

19. NICE. *Myocardial infarction (acute): Early rule out using high-sensitivity troponin tests (Elecsys Troponin T high-sensitive, ARCHITECT STAT High Sensitive Troponin-I and AccuTnI+3 assays)* Web Page. Accessed December 12, 2017. 2014.

Supplemental Material

S1. Full specification of the AUC and PE

Area under the ROC curve

The estimated AUC can be decomposed as

$$\widehat{\text{AUC}}(t, \Delta t) = \widehat{\text{AUC}}_1(t, \Delta t) + \widehat{\text{AUC}}_2(t, \Delta t) + \widehat{\text{AUC}}_3(t, \Delta t) + \widehat{\text{AUC}}_4(t, \Delta t).$$

Here AUC_1 refers to the patients pairs whose survival times can be ordered directly and is given by

$$\widehat{\text{AUC}}_1(t, \Delta t) = \frac{\sum_{i=1}^{n} \sum_{j=1; j \neq i}^{n} I\{\hat{\pi}_i(t + \Delta t \mid t) < \hat{\pi}_j(t + \Delta t \mid t)\} \times I\{\Omega_{ij}^{(1)}(t)\}}{\sum_{i=1}^{n} \sum_{j=1; j \neq i}^{n} I\{\Omega_{ij}^{(1)}(t)\}},$$

with $I(\cdot)$ as the indicator function and

$$\Omega_{ij}^{(1)}(t) = [\{T_i \in (t, t + \Delta t]\} \cap \{\delta_i = 1\} \cap \{S_i = 1\}] \cap [\{T_j > t + \Delta t\} \cap \{S_j = 1\}],$$

indicates that the event times are not censored, both patients belong to the randomly drawn subcohort ($S_i = 1$), $i, j = 1, ..., n$ and $i \neq j$.

$\text{AUC}_2(t, \Delta t)$, $\text{AUC}_3(t, \Delta t)$, $\text{AUC}_4(t, \Delta t)$ refer to the patient pairs where censoring occurs. Their corresponding indicator functions $I\{\Omega_{ij}^{(m)}(t)\}$ are

$$\Omega_{ij}^{(2)}(t) = [\{T_i \in (t, t + \Delta t]\} \cap \{\delta_i = 0\} \cap \{S_i = 1\}] \cap [\{T_j > t + \Delta t\} \cap \{S_j = 1\}],$$

for the pairs where i is a censored patient and j experiences an event,

$$\begin{aligned}\Omega_{ij}^{(3)}(t) =&[\{T_i \in (t, t + \Delta t]\} \cap \{\delta_i = 1\} \cap \{S_i = 1\}] \cap [\{T_i < T_j \leq t + \Delta t\} \cap \\ &\{\delta_j = 0\} \cap \{S_j = 1\}],\end{aligned}$$

for the pairs where i is a patient that experiences an event and j is censored, and finally

$$\begin{aligned}\Omega_{ij}^{(4)}(t) =&[\{T_i \in (t, t + \Delta t]\} \cap \{\delta_i = 0\} \cap \{S_i = 1\}] \cap [\{T_i < T_j \leq t + \Delta t\} \cap \\ &\{\delta_j = 0\} \cap \{S_j = 1\}],\end{aligned}$$

for the pairs where both i and j are censored patients.

$\widehat{\text{AUC}}_m(t, \Delta t)$ can be estimated by

$$\widehat{\text{AUC}}_m(t, \Delta t) = \frac{\sum_{i=1}^{n} \sum_{j=1; j \neq i}^{n} I\{\hat{\pi}_i(t + \Delta t \mid t) < \hat{\pi}_j(t + \Delta t \mid t)\} \times I\{\Omega_{ij}^{(m)}(t)\} \times \hat{\nu}_{ij}^{(m)}}{\sum_{i=1}^{n} \sum_{j=1; j \neq i}^{n} I\{\Omega_{ij}^{(m)}(t)\} \times \hat{\nu}_{ij}^{(m)}},$$

with $m = 2, 3, 4$. For the pairs where censoring occurs, we use $\hat{\nu}_{ij}^{(m)}$ as weighting functions for the probability that the patients would have been comparable (i.e. without censoring), with $\hat{\nu}_{ij}^{(2)} = 1 - \hat{\pi}_i(t + \Delta t \mid T_i)$, $\hat{\nu}_{ij}^{(3)} = 1 - \hat{\pi}_j(t + \Delta t \mid T_j)$ and $\hat{\nu}_{ij}^{(4)} = \{1 - \hat{\pi}_i(t + \Delta t \mid T_i)\} \times \hat{\pi}_j(t + \Delta t \mid T_j)$.

Prediction error

The calibration is measured by the prediction error (PE), where low values of PE show a well-calibrated model. The expected prediction error is as follows:

$$\mathrm{PE}(t+\Delta t \mid t) = E[\{I(T_j^* > t+\Delta t) - \pi_j(t+\Delta t \mid t)\}^2].$$

An appropriate estimator for time-to-event data is

$$\begin{aligned}\widehat{\mathrm{PE}}(t+\Delta t \mid t) =& \{n(t)\}^{-1} \sum_{j:T_j \geq t} \Big\{ I(T_j \geq t+\Delta t)\{1-\hat{\pi}_j(t+\Delta t \mid t)\}^2 \\ &+ \delta_j I(T_j < t+\Delta t)\{0-\hat{\pi}_j(t+\Delta t \mid t)\}^2 \\ &+ (1-\delta_j) I(T_j < t+\Delta t) \\ &\times \big[\hat{\pi}_j(t+\Delta t \mid T_j)\{1-\hat{\pi}_j(t+\Delta t \mid t)\}^2 \\ &+ \{1-\hat{\pi}_j(t+\Delta t \mid T_j)\} \times \{0-\hat{\pi}_j(t+\Delta t \mid t)\}^2\big] \Big\}.\end{aligned}$$

In this equation $n(t)$ denotes the number of patients still at risk at time t and the remaining parts sum over three types of situations. The first and second terms correspond to the patients that were still event free after $t+\Delta t$ and the patient that experienced the event between t and Δt, respectively. The third term refers to the patients that were censored in the interval $[t, t+\Delta t]$.

S2. Extensive results from the simulation study

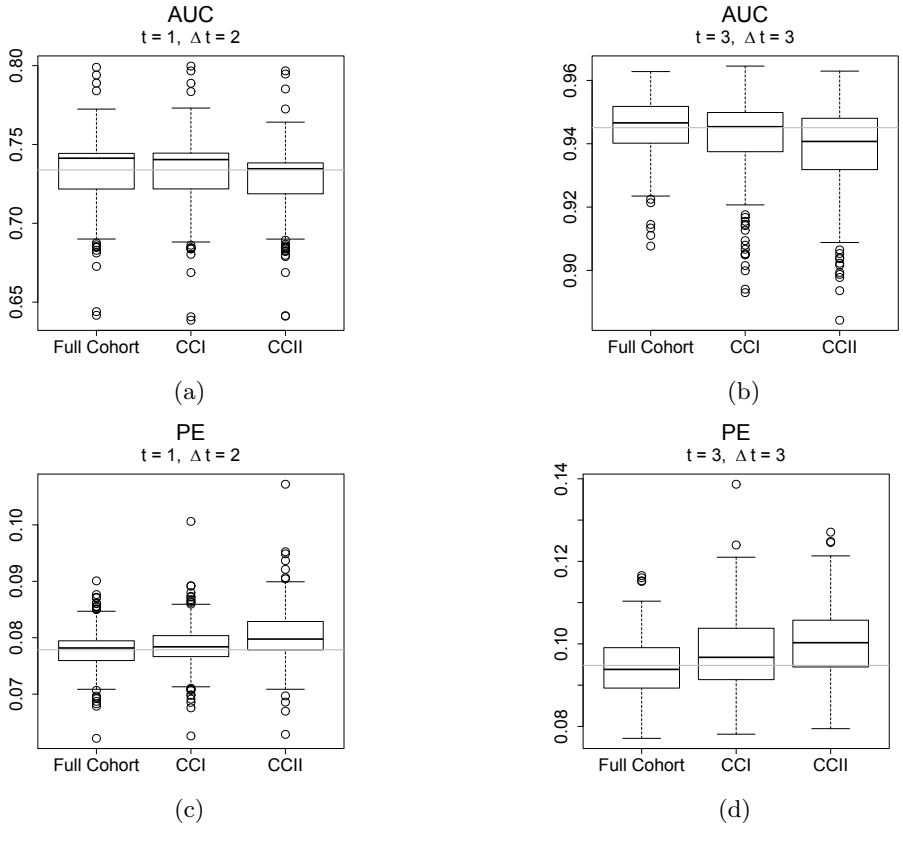
Supplemental Table 1. Results from estimating a joint model on simulated data based on 200 replications per scenario.

% Events		Scenario	Size subcohort: 1/3			Scenario	Size subcohort: 1/6		
			FC	CCI	CCII		FC	CCI	CCII
20%	α	1	0.975	0.971	0.849	2	0.976	0.966	0.799
	bias		-0.025	-0.029	-0.151		-0.024	-0.034	-0.201
	(2.5% - 97.5%)		(0.89 - 1.07)	(0.88 - 1.07)	(0.76 - 0.94)		(0.89 - 1.07)	(0.88 - 1.06)	(0.71 - 0.89)
	coverage		92%	91%	13%		92%	88%	4%
	β_1		1.003	0.996	1.087		1.004	0.986	1.139
	bias		0.003	-0.004	0.087		0.004	-0.014	0.139
	(2.5% - 97.5%)		(0.92 - 1.08)	(0.89 - 1.10)	(0.98 - 1.19)		(0.93 - 1.08)	(0.86 - 1.11)	(1.02 - 1.26)
	coverage		93%	92%	62%		97%	96%	36%
	β_2		0.319	0.331	0.558		0.325	0.357	0.713
	bias		0.019	0.031	0.258		0.025	0.057	0.413
	(2.5% - 97.5%)		(0.25 - 0.39)	(0.23 - 0.43)	(0.45 - 0.66)		(0.25 - 0.40)	(0.24 - 0.47)	(0.60 - 0.83)
	coverage		93%	91%	1%		89%	83%	0%
	β_3		0.110	0.104	0.142		0.109	0.097	0.154
	bias		0.01	0.004	0.042		0.009	-0.003	0.054
	(2.5% - 97.5%)		(0.09 - 0.13)	(0.08 - 0.13)	(0.12 - 0.17)		(0.09 - 0.13)	(0.07 - 0.12)	(0.12 - 0.19)
	coverage		82%	94%	12%		84%	98%	8%
	β_4		0.104	0.105	0.092		0.102	0.099	0.092
	bias		0.004	0.005	-0.008		0.002	-0.001	-0.008
	(2.5% - 97.5%)		(0.00 - 0.21)	(-0.04 - 0.25)	(-0.06 - 0.24)		(-0.10 - 0.21)	(-0.07 - 0.27)	(-0.08 - 0.27)
	coverage		95%	93%	92%		96%	93%	96%
	γ		-1.979	-1.987	-1.774		-1.979	-1.978	-1.676
	bias		0.021	0.013	0.226		0.021	0.022	0.324
	(2.5% - 97.5%)		(-2.27 - -1.7)	(-2.29 - -1.7)	(-2.06 - -1.50)		(-2.27 - -1.70)	(-2.28 - -1.68)	(-1.96 - -1.40)
	coverage		95%	94%	65%		93%	95%	42%

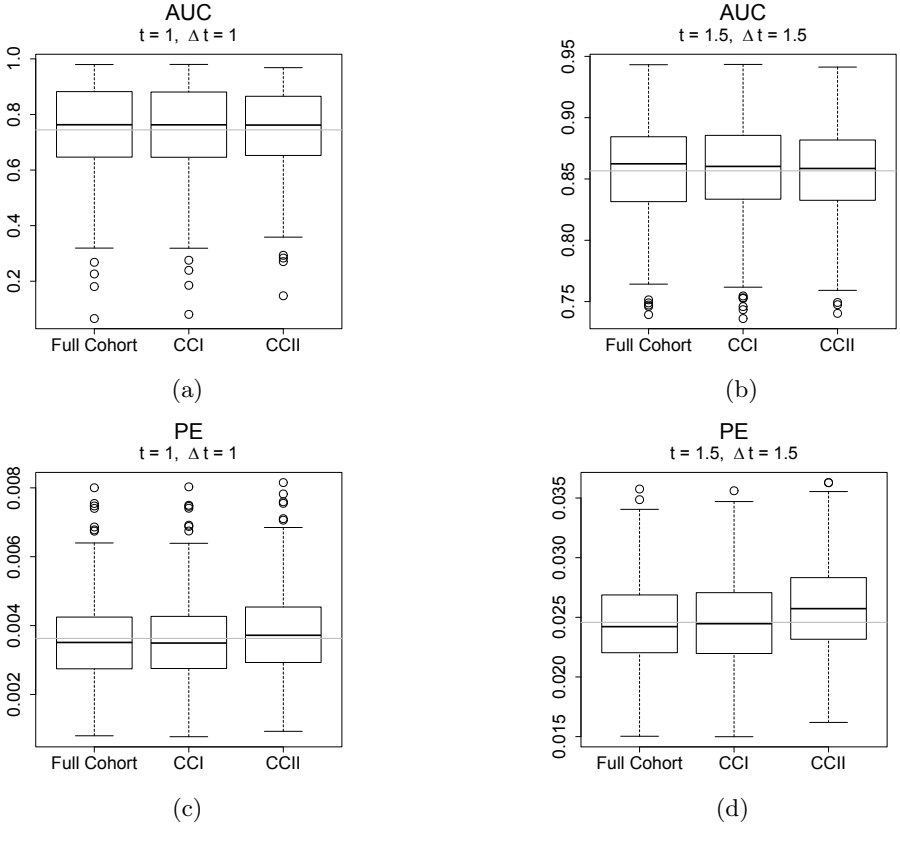
5%	α	3	0.856	0.845	0.727	4	0.858	0.835	0.649
	bias		-0.144	-0.155	-0.273		-0.142	-0.165	-0.351
	(2.5% - 97.5%)		(0.74 - 0.99)	(0.72 - 0.98)	(0.61 - 0.86)		(0.74 - 0.99)	(0.71 - 0.97)	(0.53 - 0.78)
	coverage		38%	33%	1%		39%	32%	0%
	β_1		1.003	0.993	1.062		1.005	0.990	1.127
	bias		0.003	-0.007	0.062		0.005	-0.010	0.127
	(2.5% - 97.5%)		(0.92 - 1.08)	(0.87 - 1.12)	(0.93 - 1.19)		(0.93 - 1.09)	(0.82 - 1.16)	(0.96 - 1.29)
	coverage		96%	95%	82%		96%	90%	64%
	β_2		0.331	0.343	0.474		0.334	0.371	0.638
	bias		0.031	0.042	0.174		0.034	0.071	0.339
	(2.5% - 97.5%)		(0.25 - 0.41)	(0.21 - 0.47)	(0.34 - 0.61)		(0.25 - 0.41)	(0.19 - 0.55)	(0.46 - 0.82)
	coverage		88%	91%	29%		88%	87%	4%
	β_3		0.108	0.099	0.127		0.106	0.087	0.146
	bias		0.008	-0.001	0.027		0.006	-0.013	0.046
	(2.5% - 97.5%)		(0.09 - 0.13)	(0.06 - 0.13)	(0.09 - 0.16)		(0.08 - 0.13)	(0.04 - 0.13)	(0.10 - 0.20)
	coverage		90%	99%	70%		93%	92%	58%
	β_4		0.101	0.103	0.055		0.100	0.107	0.023
	bias		0.001	0.003	-0.045		0.00	0.007	-0.077
	(2.5% - 97.5%)		(-0.01 - 0.21)	(-0.07 - 0.28)	(-0.12 - 0.24)		(-0.01 - 0.21)	(-0.12 - 0.34)	(-0.22 - 0.26)
	coverage		94%	94%	91%		95%	95%	88%
	γ		-2.730	-2.760	-2.421		-2.771	-2.806	-2.238
	bias		-0.73	-0.76	-0.421		-0.771	-0.806	-0.238
	(2.5% - 97.5%)		(-3.36 - -2.15)	(-3.44 - -2.13)	(-3.06 - -1.83)		(-3.40 - -2.18)	(-3.51 - -2.15)	(-2.89 - -1.63)
	coverage		26%	32%	76%		24%	31%	93%

The *bias* indicates the difference between the simulated parameter value and the estimated value by each of the models. The *coverage* is calculated by the percentage of times the true simulated values falls in the credible interval of each simulation.
Simulated values of the parameters: $\alpha = 1$, $\beta_1 = 1$, $\beta_2 = 0.3$, $\beta_3 = 0.1$, $\beta_4 = 0.1$, γ = -2.
FC, Full cohort; CCI, Case-cohort design - retain all survival information; CCII: Case-cohort design - classical version

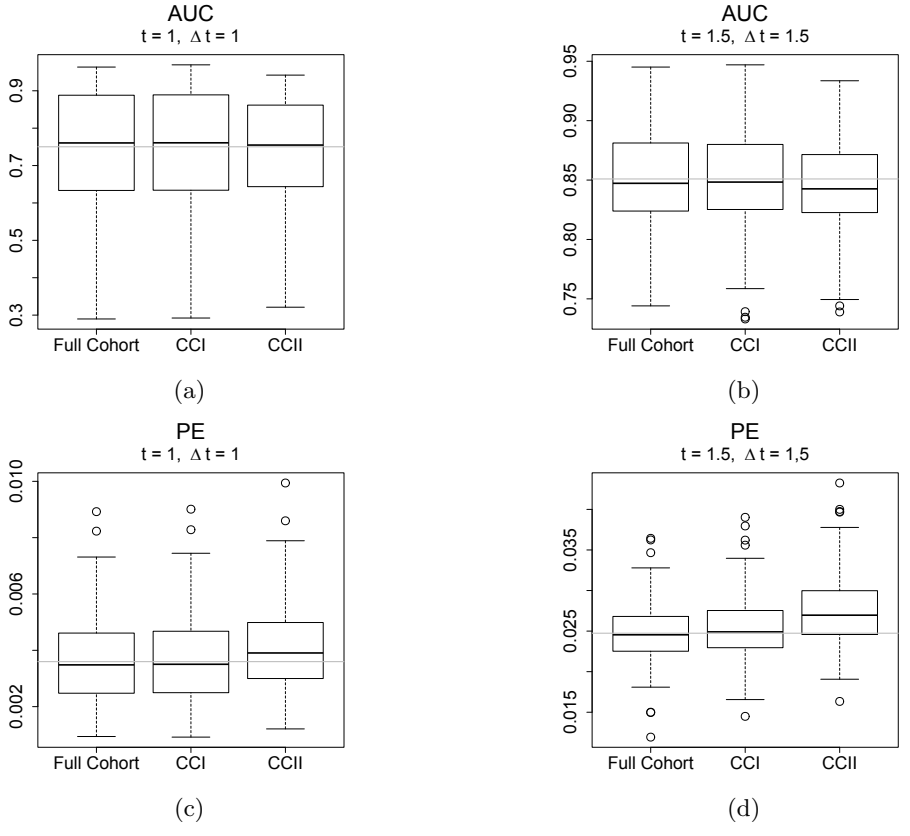
S3. Boxplots for simulation results



Supplemental Figure 3: Predictive accuracy measures from scenario 1



Supplemental Figure 4: Predictive accuracy measures from scenario 3



Supplemental Figure 5: Predictive accuracy measures from scenario 4

S4. Results from a simulation study with 500 simulated subjects.

We have performed an additional simulation study, to evaluate our method in data sets with less subjects. We simulated data sets with 500 subjects, and event rate of 25% and imitated a case-cohort design with a subcohort size of 1/3 of the full cohort. Supplemental table 2 and supplemental figure 4 show the results of this simulation. All results are in line with the previous simulations, where the newly proposed version of the case-cohort performs similar to the full cohort and the standard version of the case-cohort design performs less well. The differences, however are less pronounced in these simulations.

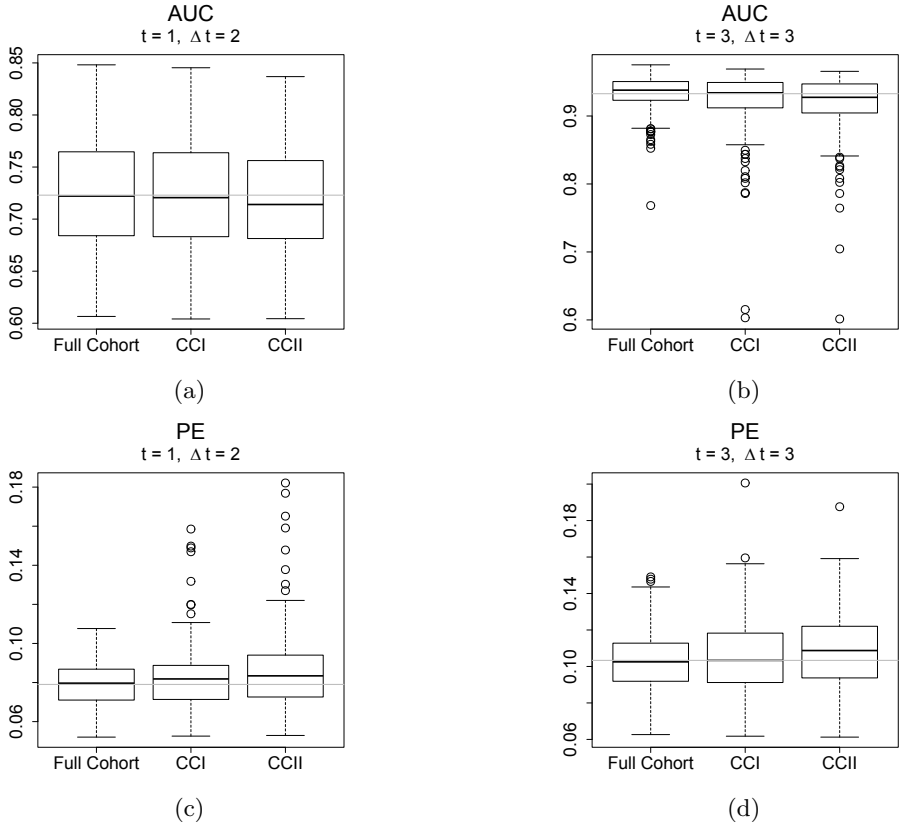
Supplemental Table 2. Results from estimating a joint model on simulated data based on 200 replications per scenario - with n = 500, ER = 25% and size of CC = 1/3

	FC	CCI	CCII
Summary simulated data			
patients, n	500	500	250
events, n	125	125	125
event rate, %	25%	25%	50%
measurements, n	2500	1350	1350
Results simulations			
α	0.905	0.894	0.793
bias	-0.095	-0.106	-0.207
(2.5% - 97.5%)	(0.76 - 1.07)	(0.74 - 1.07)	(0.64 - 0.96)
coverage	78%	77%	35%
β_1	1.006	0.982	1.085
bias	0.006	-0.018	0.085
(2.5% - 97.5%)	(0.85 - 1.16)	(0.76 - 1.20)	(0.87 - 1.30)
coverage	92%	92%	89%
β_2	0.305	0.348	0.540
bias	0.005	0.048	0.240
(2.5% - 97.5%)	(0.16 - 0.45)	(0.14 - 0.55)	(0.33 - 0.75)
coverage	96%	92%	38%
β_3	0.115	0.099	0.155
bias	0.015	-0.001	0.055
(2.5% - 97.5%)	(0.08 - 0.15)	(0.05 - 0.15)	(0.10 - 0.22)
coverage	93%	95%	57%
β_4	0.104	0.123	0.108
bias	0.004	0.023	0.008
(2.5% - 97.5%)	(-0.11 - 0.32)	(-0.18 - 0.42)	(-0.20 - 0.41)
coverage	97%	95%	94%
γ_1	-1.920	-1.939	-1.726
bias	0.08	0.061	0.274
(2.5% - 97.5%)	(-2.48 - -1.40)	(-2.53 - -1.38)	(-2.30 - -1.19)
coverage	94%	95%	83%

The *bias* indicates the difference between the simulated parameter value and the estimated value by each of the models. The *coverage* is calculated by the percentage of times the true simulated values falls in the credible interval of each simulation.

Simulated values of the parameters: $\alpha = 1$, $\beta_1 = 1$, $\beta_2 = 0.3$, $\beta_3 = 0.1$, $\beta_4 = 0.1$, $\gamma = -2$.

FC, Full cohort; CCI, Case-cohort design - retain all survival information; CCII: Case-cohort design - classical version



Supplemental Figure 6: Predictive accuracy measures of estimated joint models on simulated data based on 200 replications per scenario - with n = 500, ER = 25% and size of CC = 1/3

S5. Code

The code for simulating data from the simulation study and performing the analyses can be found at: https://github.com/SaraBaart/JM-CaseCohort

Part II: Relative conditional survival in patients with cardiovascular disease

Chapter 2

Impact of relative conditional survival estimates on patient prognosis after percutaneous coronary intervention

Baart SJ, van Domburg RT, Janssen-Heijnen MLG, Deckers JW, Akkerhuis KM, Daemen J, van Geuns RJ, Boersma H, Kardys I

Circ Cardiovasc Qual Outcomes 2017; 10-6

Abstract

Background: Some aspects of prognosis are not reflected by cumulative survival estimates. These aspects include information on the time already survived by the patient, and the patient's survival compared to the general population. Conditional survival (i.e., conditional on having survived a certain period of time already) and relative conditional survival (i.e. compared to the general population) do incorporate these aspects. We investigated these measures of prognosis in patients undergoing percutaneous coronary intervention (PCI).

Methods and results: We studied 17903 consecutive patients undergoing PCI between 2000-2014. Cumulative survival was estimated for patients with ST-elevation myocardial infarction (STEMI; n=5996, 853 deaths), non-STEMI (NSTEMI; n=5371, 901 deaths) and stable angina pectoris (SAP; n=6536, 965 deaths), in four age categories. One-year conditional and relative conditional survival up to 10 years post-PCI were calculated. The results demonstrated that 1-year cumulative survival for STEMI patients aged $\geq$ 76 years was 83%. One-year conditional survival, conditional on surviving the first month, was 92% in this group, and relative conditional survival (relative to the general population) was 99%. In younger age categories, and in NSTEMI and SAP patients, similar patterns were found, albeit less pronounced. Five-year relative conditional rendered similar results.

Conclusion: Relative conditional survival provides a comprehensive picture of patient prognosis, particularly for older STEMI patients. Although as expected, their cumulative survival is low, once they survive the first month after PCI, their prognosis is comparable to that of the general population. Therefore, relative conditional survival estimates provide an important, meaningful addition when discussing prognosis with patients.

Introduction

Patients diagnosed with coronary artery disease often ask their treating cardiologist about their prognosis. [1] Currently, prognosis is usually presented in the form of % risk of fatal events up to a certain time-point or, alternatively, cumulative probability of survival up to that time-point (estimated by the SCORE for example). [2] To account for differences in patient characteristics, the probabilities are usually stratified on factors such as age and gender. [2]

However, cumulative survival probabilities fail to account for several aspects of prognosis. Firstly, after surviving up to a certain point in time, a patient's prognosis may change. For example, after experiencing an acute myocardial infarction, adverse events are more likely to occur during the first month of follow-up. [3] Patients who survive this crucial period, may have higher survival probabilities for the rest of the follow-up period. This issue may be addressed by calculating conditional survival (also known as landmark analysis). [4] Conditional survival estimates the survival probability from a certain time-point on wards, including only patients who were still alive at that time-point. Therefore, it enables dynamic modelling of prognosis.

Secondly, cumulative survival includes death due to other reasons than the condition under investigation, and may thus pose an overly pessimistic perspective on the effect of the disease on survival. This carries particular importance in elderly patients and during longer-term follow-up. Calculating the survival probability of a patient relative to the survival probability in the general population (relative survival) may aid interpretation of cumulative survival. Relative survival probabilities are calculated by taking the ratio of the estimated survival in a certain patient cohort and the survival probability in the general population (expected survival) with the same age and sex. [5] Of note is that difficulties in interpretation, and over-estimation of relative survival,

may occur when a high proportion of deaths in the general population is due to the disease of interest. [6] Hinchliffe et al. have proposed an adjustment of the expected survival for such cases. [6]

The two methods discussed above - conditional and relative survival - can be combined to calculate the relative conditional survival, i.e. the relative survival probability after surviving a certain time-period. Relative conditional survival can demonstrate to a patient at what moment in time his prognosis becomes similar to that of an otherwise comparable person that did not have the disease. [7] In the field of oncology these types of survival probabilities are already being used. [7–9] In the cardiovascular field, use of these survival probabilities is less common. A few studies have examined conditional survival (or landmark analysis) and relative survival, but they examined these two entities separately. [10–13] Moreover, these studies focused on patients with myocardial infarction; currently, no data are available on patients with stable coronary artery disease.

In the current study, we estimated the relative conditional survival of 17903 patients undergoing percutaneous coronary intervention (PCI) for ST-elevation myocardial infarction (STEMI), non-STEMI (NSTEMI) or stable angina pectoris (SAP). Herewith, our study is the first to investigate whether incorporating information on a patient's survival up to a certain time point, as well as incorporating information on survival of the general population, provides additional insights into a patient's prognosis.

Methods

Study Population

A total of 17903 consecutive patients undergoing PCI with stent placement between January 2000 and July 2014 at Erasmus MC, Rotterdam, the Netherlands were included in this study. Baseline data collection was performed prospectively and included age, gender, and indication for the index PCI (STEMI, NSTEMI or SAP). The preferred stent type changed during the study period: bare metal stents (BMS) were used until April 2002, sirolimus-eluting stents (SES) between April 2002 and March 2003, paclitaxel-eluting stents (PES) between March 2003 and March 2007, and everolimus-eluting stents (EES) between March 2007 and July 2014. [14, 15] The preferred stent was almost exclusively used in all patients within these subsequent periods, except for (the small number of) patients who participated in trials comparing different stents. Patient management was in accordance with the applicable guidelines of the European Society of Cardiology, which changed over time. [16]

The primary endpoint was all-cause mortality. Patients were actively followed up on this endpoint by periodically reviewing hospital medical records and municipal civil registries. The latest follow-up was performed in July 2015. Patients lost to follow-up were censored at the date they were last known to be alive at the municipal civil registries or at the hospitals.

Ethics

This was an observational study. For the purpose of this study patients were not subjected to acts, neither was any mode of behaviour imposed, otherwise than as part of their regular treatment. Therefore, according to Dutch law, written informed consent for a patient to be enrolled in this study was not required.

This study was conducted according to the Privacy Policy of the Erasmus MC, and according to the Erasmus MC regulations for the appropriate use of data in patient oriented research.

Statistical Analysis

For the analyses, the patients were stratified on indication for PCI (STEMI, NSTEMI and SAP) as well as age (22-55, 56-65, 66-75, and 76-95 years). Cumulative survival (S) was calculated using the Kaplan-Meier method and standard errors for survival were based on Greenwood's formula. [5] Greenwood's formula allows standard errors to be calculated in a similar manner for cumulative survival, conditional survival, and relative conditional survival, and was applied as such. Of note is that in previous papers, cumulative survival has often been termed "observed" survival to contrast it with relative survival. [5, 10] Survival probabilities with standard errors $\leq 5\%$ were considered reliable estimates as was done previously. [8]

Conditional survival probabilities were calculated by taking the ratio of cumulative survival at a certain time point and cumulative survival at an earlier time point. Specifically, the x-year conditional survival, conditioned on having survived y years, was calculated by dividing the survival at $y + x$ years by the survival at y years (Equation 1):

$$CS(x \mid y) = \frac{S(y+x)}{S(y)} \tag{1}$$

The difference between the survival at y and $y+x$ years results in the x-year survival, conditional on surviving y years. [17] For example, the 1-year survival probability conditional on surviving five years is calculated as the ratio of the cumulative survival at six years divided by the cumulative survival at five years: $CS(1 \mid 5) = \frac{S(5+1)}{S(5)}$.

To calculate relative survival, the expected survival is needed. One-year survival probabilities for the general population were retrieved from Statistics Netherlands ('Centraal Bureau voor de Statistiek'; CBS) on 23 February 2017 and were stratified on calendar year, age and gender. [18] In order to account for the high incidence of coronary artery disease in the general population, the adjustment suggested by Hinchliffe et al. [6] was applied here and adjusted expected survival was obtained (ES^*):

$$ES^* = ES + \alpha(1 - ES) \tag{2}$$

Where ES are the survival probabilities directly obtained from Statistics Netherlands and α is the proportion of deaths due to the specific disease of interest. Approximations for α were made based on cause-specific deaths from Statistics Netherlands and disease prevalence estimates from literature. [19] This resulted in a specific α for each indication, year, age and gender. Relative survival at a certain time point was then calculated as the ratio of the cumulative survival (S) and the adjusted expected survival (ES^*) [10] using the Ederer II method. [5] Relative survival at y years is defined as:

$$RS(y) = \frac{S(y)}{ES^*(y)} \tag{3}$$

We then combined both methods (conditional survival and relative survival) and calculated the relative survival at x conditional on y, which we call relative conditional survival, as follows:

$$RCS(x \mid y) = \frac{RS(y + x)}{RS(y)} \tag{4}$$

For example, the 1-year relative survival probability conditional on five years is calculated as the ratio of the cumulative relative survival at six years divided

by the cumulative relative survival at five years: $RCS(1 \mid 5) = \frac{S(5+1)}{ES^*(5+1)} / \frac{S(5)}{ES^*(5)}$. It should be noted that relative conditional probabilities may exceed 100%. Such situations may occur when the number of events is lower than expected compared to the general population during a certain follow-up period.

In this paper, 1-year (relative) conditional survival probabilities were calculated - conditional on having already survived a certain period of time ($CS(1 \mid y)$ and $RCS(1 \mid y)$). However, in case a patient has already survived a substantial time period (f.e. 5 or 10 years), it might not be clinically relevant to estimate only the short term – i.e. 1-year – survival probabilities. Therefore, the analyses were repeated calculating 5-year, instead of 1-year, survival probabilities ($CS(5 \mid y)$ and $RCS(5 \mid y)$).

In order to allow for correction for multiple variables at the same time, we subsequently applied regression models to our relative survival data. In such models, the hazard for a patient is split into the expected hazard and the excess hazard due to the disease of interest. Several methods are available, depending on the type of data and software available. In this study, we used the Poisson model. [20] For the interpretation of excess hazard to be valid, the assumption that the proportion of death due to the disease in the general population is negligible is of great importance. In this study, this was ensured by adjusting the expected survival. We included age, gender, follow-up time, indication for catheterization, and the interactions between the latter two variables in the model.

Continuous normal variables are presented as mean and standard deviation and non-normal variables are presented as median and interquartile range (IQR). Analyses were performed with SAS version 9.4. Relative survival was calculated with a publicly available SAS syntax and macro [21] that uses the Ederer II method. [5] The graphs were made with R version 3.3.2.

Results

Mean age was 63 years and 72% were men. Indication for PCI was STEMI in 33%, NSTEMI in 30%, and SAP in 37% of the patients. Median survival time in months was 42 (interquartile range 22-83 months) (Table 1). Follow-up information was complete until 31 July 2015 for 98% of the patients.

Table 1: Baseline characteristics

Characteristics	Total n = 17903 (100%)	STEMI n = 5996 (33%)	NSTEMI n = 5371 (30%)	SAP n = 6536 (37%)
Age (years), mean ± SD	63 ± 12	61 ± 13	64 ± 12	64 ± 11
Male, n (%)	12887 (72%)	4488 (75%)	3726 (69%)	4673 (72%)
Survival time (months), median (IQR)	42 (22-83)	35 (19-64)	39 (19-75)	54 (31-91)
Diabetes mellitus, n (%)	3216 (18%)	690 (12%)	166 (22%)	1360 (21%)
Hypertension, n (%)	8560 (48%)	2222 (37%)	2850 (53%)	3488 (53%)
Hypercholesterolaemia, n (%)	9699 (54%)	2178 (36%)	3216 (60%)	4305 (66%)
Active smoking, n (%)	4964 (28%)	2446 (41%)	1299 (24%)	1219 (19%)
Renal failure, n (%)	981 (5%)	102 (2%)	423 (8%)	456 (7%)
Family history of coronary heart disease, n (%)	5872 (33%)	1776 (30%)	1795 (33%)	2301 (35%)
Prior MI, n (%)	4158 (23%)	561 (9%)	1745 (32%)	1852 (28%)
Prior PCI, n (%)	1971 (11%)	341 (6%)	636 (12%)	994 (15%)
Prior CABG, n (%)	1424 (8%)	144 (2%)	527 (10%)	753 (12%)

SD, standard deviation; IQR, interquartile range; STEMI, ST-elevation myocardial infarction; NSTEMI, non STEMI; SAP, stable angina pectoris; MI, myocardial infarction; PCI, percutaneous coronary intervention; CABG, coronary artery bypass grafting

Figure 1 shows Kaplan Meier curves with 95% confidence intervals for cumulative survival. Corresponding survival probabilities for certain time points during follow-up are given in Table 2. For all three indications for PCI, clear

differences were present in cumulative survival between the age groups. As expected, the oldest patients had the lowest survival probabilities, and the probabilities diverged during follow-up with patients aged 76-95 years showing the largest decrease ($p < 0.01$ between all age groups at ten years) (Figure 1, Table 2).

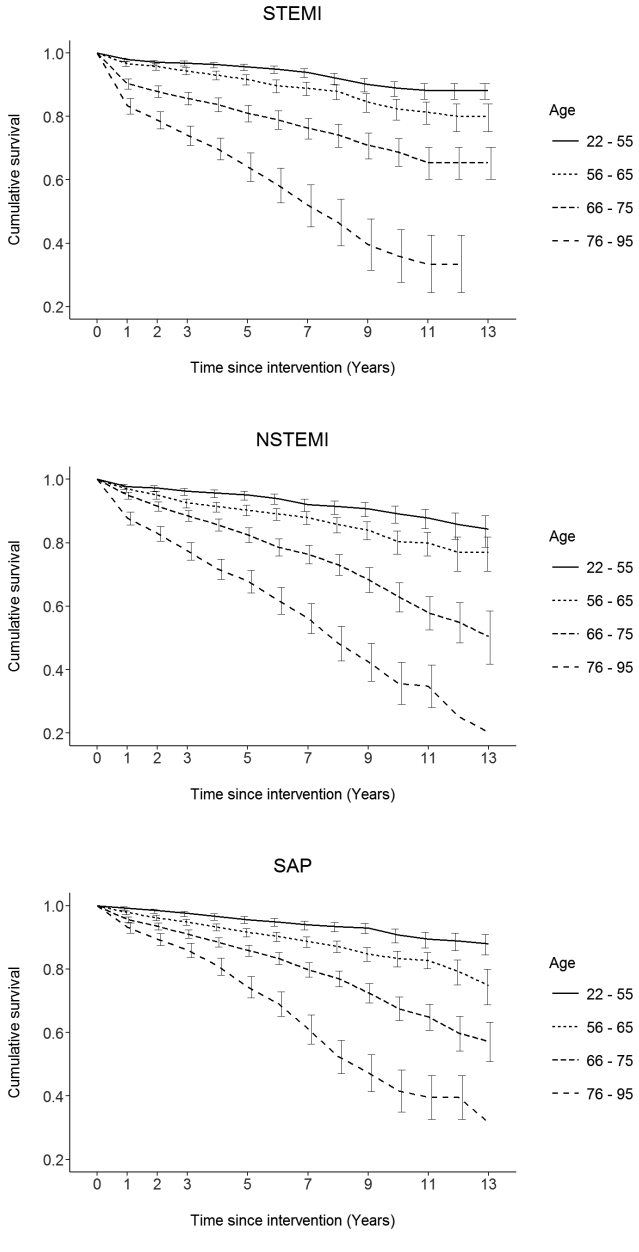


Figure 1: Cumulative survival stratified by indication for percutaneous coronary intervention and age. NSTEMI indicates non–ST-segment–elevation myocardial infarction; SAP, stable angina pectoris; and STEMI, ST-segment–elevation myocardial infarction.

Table 2: Cumulative survival probabilities stratified by indication for PCI and age

Indication for PCI	Age, y	Alive at start	Cumulative Deaths at					Survival at (95% Confidence Interval)				
			1 y	1 y, 1 mo	2 y	5 y	10 y	1 y	1 y, 1 mo	2 y	5 y	10 y
STEMI	22-55	2071	43	45	60	76	114	98% (97%-98%)	98% (97%-98%)	97% (96%-98%)	96% (94%-97%)	89% (86%-91%)
	56-65	1661	55	57	70	107	144	97% (96%-97%)	97% (96%-97%)	96% (95%-97%)	92% (90%-93%)	82% (79%-85%)
	66-75	1299	125	130	156	203	238	90% (89%-92%)	90% (88%-91%)	88% (86%-90%)	81% (78%-83%)	69% (64%-73%)
	76-95	885	147	148	185	243	273	83% (81%-86%)	83% (81%-86%)	79% (76%-81%)	64% (59%-68%)	36% (28%-44%)
NSTEMI	22-55	1286	29	29	35	54	78	98% (97%-98%)	98% (97%-98%)	97% (96%-98%)	95% (93%-96%)	89% (86%-92%)
	56-65	1517	46	49	74	122	158	97% (96%-98%)	97% (96%-98%)	95% (94%-96%)	90% (88%-92%)	80% (76%-84%)
	66-75	1495	74	79	125	208	280	95% (94%-96%)	95% (93%-96%)	91% (90%-93%)	82% (80%-85%)	63% (58%-67%)
	76-95	1043	127	132	176	259	324	88% (86%-90%)	87% (85%-89%)	83% (80%-85%)	68% (64%-71%)	36% (29%-42%)

SAP	22-55	1494	11	12	23	57	86	99% (99%-100%)	99% (99%-100%)	98% (98%-99%)	96% (94%-97%)	91% (88%-93%)
	56-65	2079	44	49	78	150	209	98% (97%-98%)	98% (97%-98%)	96% (95%-97%)	92% (90%-93%)	83% (81%-86%)
	66-75	1977	85	91	126	235	339	96% (95%-97%)	95% (94%-96%)	94% (92%-95%)	86% (84%-88%)	68% (64%-71%)
	76-95	969	66	70	101	192	268	93% (91%-95%)	93% (91%-94%)	89% (87%-91%)	74% (71%-78%)	42% (35%-48%)

PCI, percutaneous coronary intervention; STEMI, ST-elevation myocardial infarction; NSTEMI, non STEMI; SAP, stable angina pectoris

In Table 3 the 1-year conditional survival probabilities are shown (1-year conditional survival in panel B and 1-year relative conditional survival in panel C) together with the number of patients at the start of each interval and the number of deaths during that interval (panel A). Due to a limited number of patients aged 76-95 years at the end of the follow-up period, reliable estimates could not be calculated for all 1-year survival probabilities conditional on 9 years survival. Supplementary table 1 presents the expected survival estimates used to calculate relative survival in panel C. Figure 2 gives a graphical representation of these results. Since the first period of follow-up showed to be the most interesting, additional plots were inserted into Figure 2A, plotting the conditional survival in the first six months in more detail. Supplementary Figure 1 depicts separate curves for observed and expected survival for STEMI patients, on which the relative conditional survival estimates in Figure 2B are based. As expected, overall, when conditioned on surviving the first month (CS(1 year | 1 month)), 1-year conditional survival probabilities were higher compared to the 1-year survival probability at the start of follow-up. (Table 3, panel B). This finding was most pronounced in the eldest STEMI patients. Specifically, for these patients the 1-year survival probability from the start of follow-up at the index event was 83% (95% CI: 81%-86%). For patients from this category that survived the first month after the PCI, the 1-year conditional survival was 92% (90%-94%). When conditioned on surviving the first year, the 1-year conditional survival probability was even higher for the two oldest age groups in the STEMI patients, with those aged 76-95 years having an estimated 1-year survival of 95% (93%-96%). For the younger patients, it was 99% (99%-99%). When conditioned on even longer survival (four and nine years), the 1-year conditional survival probabilities remained the same or were at most 3% lower. Higher 1-year survival is reflected by the initial sharp increase shown in Figure 2A. The

increase was most prominent in the first month, as illustrated by the subplots in Figure 2. Relative conditional survival was similar to conditional survival in the younger STEMI patients. Conditional probabilities were already close to 100% and relative conditional probabilities were only slightly higher, indicating that reporting relative survival is less useful when observed survival probabilities are already high. However, in the older age categories relative conditional survival was markedly higher than conditional survival (Table 3, panel C and Figure 2B). STEMI patients aged 76-95 years had a 1-year relative survival probability of 89% (87%-92%) (opposed to 83% 1-year cumulative survival) and if they survived the first month, their 1-year relative conditional survival was 99% (97%-101%) (92% for conditional survival).

Table 3: Conditional 1-year survival probabilities stratified by indication for PCI and age

		Alive at start of interval (Deaths during Interval)* Alive at					Conditional Survival (95% Confidence Interval) 1-y Survival Conditional on Surviving:					Relative Survival (95% Confidence Interval) 1-y Survival Conditional on Surviving:				
Ind	Age	0 y	1 mo	1 y	4 y	9 y	0 y	1 mo	1 y	4 y	9 y	0 y	1 mo	1 y	4 y	9 y
STEMI	22-55	2071	2041	1961	835	439	98%	99%	99%	99%	99%	98%	99%	99%	100%	99%
		(43)	(18)	(17)	(6)	(6)	(97%-98%)	(99%-99%)	(99%-99%)	(98%-100%)	(97%-99%)	(97%-99%)	(99%-100%)	(99%-100%)	(99%-100%)	(98%-100%)
	56-65	1661	1622	1548	565	278	97%	99%	99%	99%	97%	97%	100%	100%	100%	99%
		(55)	(20)	(15)	(8)	(7)	(96%-97%)	(98%-99%)	(98%-99%)	(97%-99%)	(95%-99%)	(96%-98%)	(99%-100%)	(99%-100%)	(98%-100%)	(96%-101%)
	66-75	1299	1234	1142	367	169	90%	95%	97%	97%	97%	92%	97%	100%	100%	102%
		(125)	(65)	(31)	(12)	(5)	(89%-92%)	(93%-96%)	(96%-98%)	(94%-98%)	(93%-99%)	(91%-94%)	(95%-98%)	(98%-100%)	(97%-101%)	(98%-104%)
	76-95	885	798	711	135	32	83%	92%	95%	92%	*	89%	99%	102%	100%	*
		(147)	(61)	(38)	(11)	(3)	(81%-86%)	(90%-94%)	(93%-96%)	(86%-95%)		(87%-92%)	(97%-101%)	(100%-104%)	(94%-104%)	
NSTEMI	22-55	1286	1274	1209	648	229	98%	99%	100%	99%	98%	98%	99%	100%	100%	99%
		(29)	(19)	(6)	(4)	(4)	(97%-98%)	(98%-99%)	(99%-100%)	(98%-100%)	(95%-99%)	(97%-99%)	(98%-99%)	(99%-100%)	(99%-100%)	(96%-100%)
	56-65	1517	1503	1433	690	185	97%	98%	98%	99%	96%	98%	98%	99%	100%	97%
		(46)	(35)	(28)	(8)	(8)	(96%-98%)	(97%-98%)	(97%-99%)	(98%-99%)	(92%-98%)	(97%-98%)	(98%-99%)	(98%-99%)	(99%-101%)	(93%-99%)
	66-75	1495	1471	1380	617	151	95%	96%	96%	96%	92%	97%	98%	99%	99%	97%
		(74)	(55)	(51)	(23)	(12)	(94%-96%)	(95%-97%)	(95%-97%)	(94%-98%)	(86%-95%)	(96%-98%)	(97%-99%)	(97%-99%)	(98%-101%)	(91%-101%)
	76-95	1043	996	886	268	50	88%	91%	94%	95%	*	93%	97%	101%	103%	*
		(127)	(85)	(49)	(14)	(8)	(86%-90%)	(90%-93%)	(93%-96%)	(91%-97%)		(91%-95%)	(95%-99%)	(99%-102%)	(99%-105%)	

SAP	22-55	1494	1490	1460	1019	330	99%	99%	99%	99%	98%	100%	100%	100%	99%	98%
		(11)	(9)	(12)	(10)	(8)	(99%-100%)	(99%-100%)	(99%-100%)	(98%-99%)	(95%-99%)	(99%-100%)	(99%-100%)	(99%-100%)	(99%-100%)	(96%-99%)
	56-65	2079	2066	2007	1230	315	98%	98%	98%	98%	98%	99%	99%	99%	99%	100%
		(44)	(37)	(34)	(22)	(5)	(97%-98%)	(98%-99%)	(98%-99%)	(97%-99%)	(96%-99%)	(98%-99%)	(98%-100%)	(99%-100%)	(98%-100%)	(98%-101%)
	66-75	1977	1959	1848	1022	208	96%	96%	98%	97%	93%	98%	98%	100%	100%	98%
		(85)	(75)	(41)	(30)	(14)	(95%-97%)	(95%-97%)	(97%-98%)	(96%-98%)	(89%-96%)	(97%-99%)	(97%-99%)	(99%-101%)	(99%-101%)	(93%-101%)
	76-95	969	955	878	400	51	93%	94%	96%	92%	88%	99%	100%	102%	100%	*
		(66)	(56)	(35)	(32)	(6)	(91%-95%)	(92%-95%)	(94%-97%)	(89%-94%)	(76%-95%)	(97%-100%)	(98%-101%)	(101%-104%)	(97%-103%)	

1-year survival probabilities conditional on having survived certain periods of time. The first column indicates the probability of surviving the first year from the start of the study (conditional on 0 years). The probabilities can also be interpreted as – for the last column for example – surviving 10 years, given survival of the first 9 years. Relative survival is the survival compared to the general population with the same age and gender. PCI indicates percutaneous coronary intervention; Ind, Indication for intervention; STEMI, ST-elevation myocardial infarction; NSTEMI, non STEMI; SAP, stable angina pectoris

*Standard error $> 5\%$

† Deaths during interval indicate the number of deaths for each 1-year interval per group. For example, for STEMI patients aged 22-55, 43 patients died in the first year, 18 patients died in the period between one month and one year and a month, etcetera.

Similar results, albeit less outspoken, were present in the NSTEMI and SAP patients. When conditioned on the first month of survival, 1-year conditional survival probabilities were slightly higher than the 1-year survival from the start of follow-up (Table 3, panel B and Figure 2A). When conditioned on longer periods of survival (four or nine years), 1-year conditional survival probabilities decreased slightly. For example, SAP patients aged 76-95 years showed a 1-year survival probability of 96% (94% - 97%) conditional on one year of survival. This was 92% (89%-94%) conditioned on four years of survival and 88% (76%-95%) conditioned on nine years. The 1-year relative conditional survival probabilities (Table 3, panel C and Figure 2B) again shifted upwards compared to the 1-year observed conditional survival. The decrease found in the 1-year conditional survival probabilities, conditioned on surviving four and nine years, was not present in the 1-year relative conditional survival. Conditional on having survived the first months, the relative conditional survival probabilities remained 97% or higher throughout follow-up.

When comparing the three indications for PCI, overall, SAP patients had better prognosis than NSTEMI patients, and NSTEMI patients performed better than STEMI patients. For cumulative survival, the difference was most prominent in the patients aged 76-95 years (Figure 1, Table 2). In this age category, at five years of follow up, cumulative survival of SAP patients was 6% higher compared to NSTEMI patients ($p = 0.008$) and 10% higher compared to STEMI patients ($p < 0.001$). After conditioning on survival of the first month, the differences in prognosis between the patient groups were smaller, and the 1-year probabilities across the groups varied no more than 5% for the same age categories.

The 5-year survival probabilities can be found in Supplementary Tables 2 and 3. Results remained essentially the same, however, the differences between

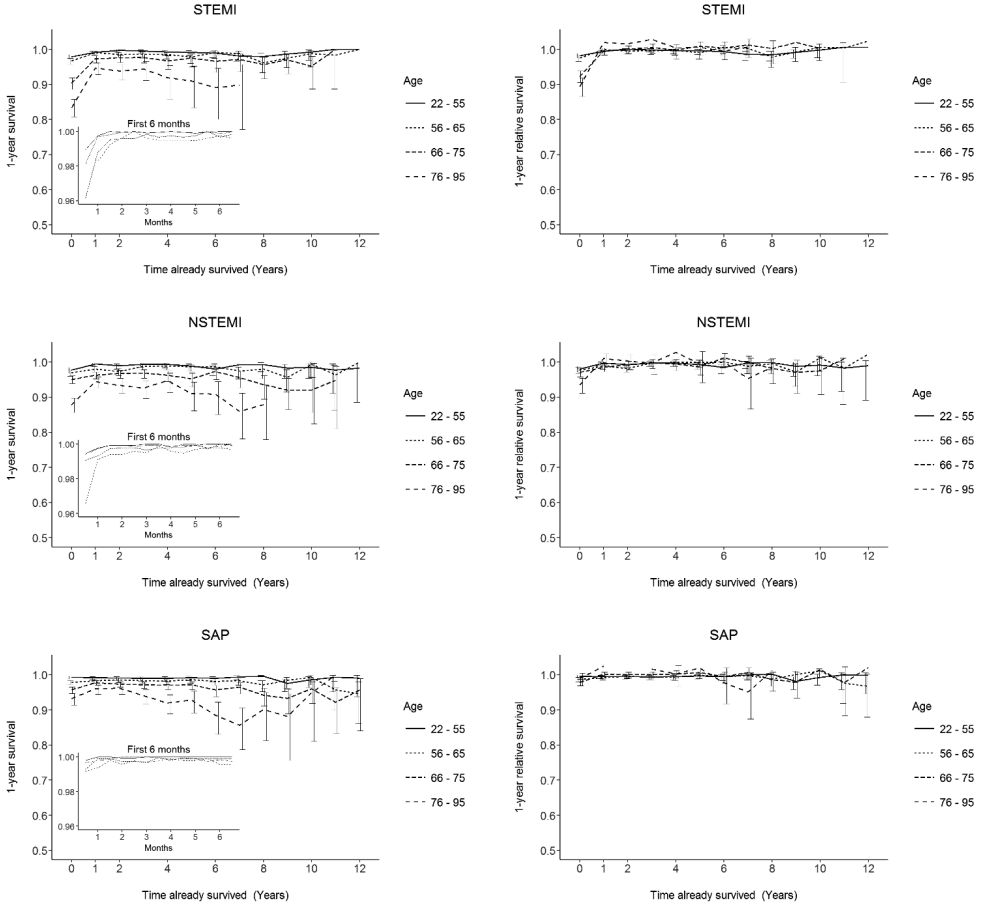


Figure 2: One-year conditional survival estimates stratified by indication for percutaneous coronary intervention (PCI) and age. **A**, One-year conditional survival estimates. **B**, One-year relative conditional survival estimates. The lines in the graph indicate the 1-year survival probabilities conditional on having survived the period of time indicated on the x axis. The graph can also be interpreted–at having survived 4 years for example—as the probability of surviving 5 years, given the survival of the first 4 years. The subplots are zoomed in on the first 6 months after PCI. Error bars, 95% confidence intervals. NSTEMI indicates non–ST-segment–elevation myocardial infarction; SAP, stable angina pectoris; and STEMI, ST-segment–elevation myocardial infarction.

conditional and relative conditional survival were more pronounced than in the 1-year survival estimates. The latter illustrates that in this PCI population, longer term (i.e., 5-year) survival is also excellent when compared to the general population.

We stratified our main analyses on indication for PCI and age. However, gender may also have important effects on survival probability. One- and 5-year survival tables stratified on gender and age can be found in Supplementary tables 4-6. Differences in cumulative survival between men and women were most prominent in patients aged 76-95 years; in the remaining age groups, all differences were no larger than 5%. Specifically, women aged 76-95 years showed higher cumulative survival ($p < 0.001$ at five years). For the 1-year conditional and 1-year relative conditional survival probabilities the differences were no larger than 3%, except for the eldest patients conditioned on nine years follow-up. Here the survival for men was 7% and 11% higher, for conditional and relative conditional survival respectively, although not statistically significant. In the table with the 5-years survival, older women displayed higher 5-year conditional survival probabilities than men until five years into follow-up. Conditioned on surviving eight years, men showed higher 5-years conditional survival probabilities, although all differences were not significant. Men seemed to perform better than women relative to the general population of their age and gender. These gender-specific survival estimates may in part have been confounded by gender differences in indication for intervention. In men, 35%, 29%, and 36% experienced STEMI, NSTEMI and SAP, respectively. In women, these proportions were 30%, 33%, and 37%, respectively.

Since reliable estimates could not be obtained after simultaneous stratification on age, gender and indication of intervention due to sample size limitations, a relative survival regression model was fitted, where the number of deaths in a

Table 4: Results From the Poisson Survival Model

Parameter	Relative Excess Risk	95% Confidence Interval	P-value
Before vs after 1y of follow-up, per indication			
STEMI	6.98	4.95-9.85	<0.001
NSTEMI	2.76	2.06-3.69	<0.001
SAP	2.33	1.53-3.56	<0.001
Between indication, after 1y of follow-up			
STEMI vs SAP	1.22	0.79-1.87	0.374
NSTEMI vs SAP	1.91	1.32-2.76	<0.001
Covariates			
Female vs Male	1.38	1.14-1.66	<0.001
Age (per year)	1.03	1.03-1.04	<0.001

Parameter estimates are interpreted as relative excess risks estimates or excess hazard ratios. For example, the estimate of 6.98 for the interaction between STEMI and follow-up time indicates that there is a 7-fold increased excess risk of mortality (compared with the general population) for patients with STEMI in the first year after PCI compared with the remainder of the follow-up, adjusted for age and sex. The nonsignificant estimate of 1.22 for the patients with STEMI versus SAP after 1 year of follow-up indicates that after the first year of follow-up, there is no excess risk for the patients with STEMI compared with the patients with SAP. NSTEMI indicates non–ST-segment–elevation myocardial infarction; SAP, stable angina pectoris; and STEMI, ST-segment–elevation myocardial infarction.

time interval was assumed to have a Poisson distribution. [20] The variables included in the model were indication for catheterization, gender, age (per one year increase), and an interaction between indication and moment of assessment (before versus after 1 year of follow-up). This model resulted in relative excess risk (RER) estimates (i.e., the difference between expected and observed hazard) for each of the three indications for the first year of follow-up compared to the remainder of the follow-up (Table 4). The RER of 6.98 (95% CI: 4.95-9.85) for the STEMI patients, indicates that the excess risk of mortality (compared to the general population) is almost seven times higher in the first year after

PCI than in the remainder of the follow-up for STEMI patients ($p < 0.001$), adjusted for age and sex. For NSTEMI and SAP patients, the RER in the first year after PCI compared to the rest of the follow-up was smaller (2.76 (2.06-3.69) and 2.33 (1.53-3.56), respectively) but also highly significant. This coincides with the findings from the stratified analyses, and illustrates that the first period after PCI (in the case of the model, the first year) is the most crucial, and that this is most pronounced in STEMI patients. In a post-hoc analysis the relative excess risk in the first month of follow-up was compared to the remainder of the follow-up, resulting in even larger and highly significant RER estimates ($p < 0.001$). Furthermore, the model provided RER estimates for STEMI and NSTEMI compared to SAP after one year of follow-up. The results demonstrated that although no significant excess hazard for STEMI compared to the SAP patients is present within the time period 1 year after PCI and onwards, an excess hazard was present for the NSTEMI patients compared to the SAP patients. The latter, with an RER estimate of 1.91 (1.32-2.76) indicates that even after the patients have stabilized, the excess risk of mortality for NSTEMI patients remains twice as high as for the SAP patients. The difference in excess risk between males and females (RER = 1.38 (1.14-1.66), i.e. 38% higher for females) coincides with the results found in the stratified analyses.

Discussion

In this study, we reported multiple measures of prognosis in 17903 patients who underwent a PCI procedure for STEMI, NSTEMI or SAP and were followed for up to 13 years post-PCI. We found that the 1-year cumulative survival probability for STEMI patients aged 76-95 years was 83% (95% CI: 81%-86%), but when conditioned on surviving the first month, the 1-year conditional survival for this group was 92% (90%-94%). Relative to the general population

of the same age, their 1-year relative conditional survival probability was estimated at 99% (97%-101%). In other words, if STEMI patients 76 years or older survive one month, their prognosis becomes the same as that of the general population. Furthermore, in the younger age categories, as well as in NSTEMI and SAP patients, similar patterns were found, albeit less pronounced. Altogether, these results demonstrate that relative conditional survival estimates may provide an important and meaningful addition when discussing prognosis with a patient. It should be noted that the incremental information conveyed by relative conditional survival estimates, as compared to cumulative survival probabilities, is most pronounced in groups where the survival probability is low, and less so in those where survival is high (as witnessed in the youngest group of patients in our study for example). Our findings were supported by a regression model, where age, gender, follow-up time and indication were estimated together, resulting in high excess hazards in the first year of follow-up, especially for STEMI patients.

Our results suggest that accounting for the time already survived is useful for providing a comprehensive picture of patient prognosis, in particular for patients who are older and who have experienced a STEMI. In such patients, 1-year survival probability conditional on surviving the first month was 9% higher than cumulative survival probability as estimated from the start of follow-up. This finding complies with existing literature, showing that most events occur in the first month after PCI, making the first month the most crucial period. The 1-year survival probability was even higher for patients who survived the first year post-PCI. Since conditional survival can be calculated at every time point during follow-up, a patient's survival status can be used to repeatedly update prognosis. This enables dynamic modelling of prognosis, which provides a more accurate and comprehensive picture at any specific time point during

follow-up. Further into the follow-up the 1-year conditional survival probabilities appeared to decline, especially in the older age groups. A possible explanation for this decrease may be that these patients were ageing; and were therefore becoming more likely to die. Calculating relative survival enables incorporation of information on ageing of patients, because it relates their survival to the expected survival of individuals in the general population with the same age and gender. Accordingly, for every age group and indication studied here, the 1-year relative conditional survival conditioned on the first month or any longer period during follow- up (up to 10 years), was 97% or higher. This indicates that for patients undergoing PCI, and in particular those experiencing STEMI, survival probabilities become close to the survival of the general population after they survive the first month after the procedure.

In oncology research, relative and conditional survival are used in order to investigate at which moment after undergoing treatment a patient's prognosis becomes equal again to the prognosis of the general population. [7–9] In cardiovascular research however, these methods are used less often, and in particular the combination of these two methods has not been examined in such a large consecutive cohort. Alabas et al. [13] and Gale et al. [12] studied the relative survival of a large cohort at 6 months post MI. They found 6-months relative survival rates of around 96% for patients aged <65 years (STEMI and NSTEMI). In patients aged 65-80 years, 6-months relative survival was between 85-92% for STEMI patients and between 83-89% for NSTEMI patients, depending on the year of admission. Although the present study reported relative survival only at one year and not six months, the survival rates we found were all higher than those described in the study by Alabas et al. The difference was largest for NSTEMI patients. These differences may possibly be explained by the difference in period of admission. Alabas et al. studied relative survival per admission

period between 2003 and 2010, and found increasing relative survival estimates. This increase might have carried on until 2014, which could potentially have resulted in a higher average relative survival, closer to our results. Longer term relative survival was presented by Velders et al. [22] (up to three years), De Carvalho et al. [11] (up to five years), and Nelson et al. [10] (up to six years). De Carvalho calculated a relative survival ratio (RSR) at three and five years post AMI for three different ethnic Asian populations. They found RSRs between 0.73 and 0.82 at three years, depending on the ethnic group. At five years the RSR had dropped another 3 or 4 points, "indicating that the residual risk of mortality persisted long after the index AMI". [11] These findings are not consistent with results from the present study, since we did not observe any residual risk in any of the subgroups after the first year of follow-up. The residual risk found by De Carvalho et al. may also have resulted from the admission period (2000-2005), which was earlier in time than the current study. Furthermore, differences in study populations could have contributed. Not only did De Carvalho study different ethnic populations, the patients also showed higher rates of diabetes and current smoking status. Velders et al. calculated interval-specific 1-year relative survival for a STEMI population aged over 80 years until three years follow-up. Their findings coincide with our results from the eldest STEMI patients, where no residual risk was found after patients survived the first month of follow-up.

Our study showed that combining conditional and relative survival results in different estimates of prognosis compared to those that are obtained by merely considering cumulative survival or conditional or relative survival separately. Patients may benefit from the information conveyed by these methods, as they provide a more up-to-date picture of the patient's situation and also take into account the survival of the general population.

Strengths of our study include large cohort size, long-term follow-up and inclusion of several indications for catheterization. However, limitations should also be mentioned. Firstly, the α used in the adjusted expected survival to account for the large prevalence of the disease of interest in the general population, is not known and had to be approximated. In the current study, however the difference between the unadjusted and adjusted expected survival was on average extremely small (the median difference was smaller than 0.0001 percentage point [IQR=0-0.0003 percentage points]), indicating that a possible over- or under-estimation for α does not have a large impact on the analysis. Furthermore, we were not able to account for age, indication for catheterization and gender in the survival tables simultaneously. Even larger study size is required for such an analysis. The survival model, however, is capable of accounting for all variables together. The results from the regression model concurred with the stratified survival estimates. Finally, since this was a single centre study, external validation is warranted.

In conclusion, long-term survival prognosis for patients undergoing PCI for STEMI, NSTEMI or SAP can be supplemented by estimates of conditional and relative conditional survival. Conditional survival probabilities incorporate information on the patient's survival up to a certain time point. Relative survival relates the patient's survival to the general population and may be particularly useful for older patients. In the current study the most prominent findings pertained to STEMI patients aged 76-95 years. Their 1-year survival at the start of follow up was 83%. Conditioned on surviving the first month, the 1-year conditional survival was 92%. Relative to the general population, it was 99%, meaning that once these patients survived the first month after PCI, their 1-year survival probability was essentially the same as that of the general population. In sum, the information obtained from these two survival methods

provides additional insights into prognosis and could therefore be helpful when communicating prognosis to patients.

References

1. Ahalt, C. *et al.* "Knowing is better": preferences of diverse older adults for discussing prognosis. *J Gen Intern Med* **27,** 568–75 (2012).
2. Conroy, R. M. *et al.* Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* **24,** 987–1003 (2003).
3. Hofma, S. H. *et al.* One year clinical follow up of paclitaxel eluting stents for acute myocardial infarction compared with sirolimus eluting stents. *Heart* **91,** 1176–80 (2005).
4. Dafni, U. Landmark analysis at the 25-year landmark point. *Circ Cardiovasc Qual Outcomes* **4,** 363–71 (2011).
5. Cho, H., Howlader, N., Mariotto, A. B. & Cronin, K. A. *Estimating relative survival for cancer patients from the SEER Program using expected rates based on Ederer I versus Ederer II method* Report (NCI, 2011).
6. Hinchliffe, S. R., Dickman, P. W. & Lambert, P. C. Adjusting for the proportion of cancer deaths in the general population when using relative survival: a sensitivity analysis. *Cancer epidemiol.* **36,** 148–52 (2012).
7. Baade, P. D., Youlden, D. R. & Chambers, S. K. When do I know I am cured? Using conditional estimates to provide better information about cancer survival prospects. *Med J Aust* **194,** 73–7 (2011).
8. Janssen-Heijnen, M. L. *et al.* Prognosis for long-term survivors of cancer. *Ann Oncol* **18,** 1408–13 (2007).
9. Shack, L., Bryant, H., Lockwood, G. & Ellison, L. F. Conditional relative survival: a different perspective to measuring cancer outcomes. *Cancer Epidemiol* **37,** 446–8 (2013).
10. Nelson, C. P., Lambert, P. C., Squire, I. B. & Jones, D. R. Relative survival: what can cardiovascular disease learn from cancer? *Eur Heart J* **29,** 941–7 (2008).
11. De Carvalho, L. P. *et al.* Differences in late cardiovascular mortality following acute myocardial infarction in three major Asian ethnic groups. *Eur Heart J Acute Cardiovasc Care* **3,** 354–62 (2014).
12. Gale, C. P. *et al.* Trends in hospital treatments, including revascularisation, following acute myocardial infarction, 2003-2010: a multilevel and relative survival analysis for the National Institute for Cardiovascular Outcomes Research (NICOR). *Heart* **100,** 582–9 (2014).
13. Alabas, O. A., Allan, V., McLenachan, J. M., Feltbower, R. & Gale, C. P. Age-dependent improvements in survival after hospitalisation with acute myocardial infarction: an analysis of the Myocardial Ischemia National Audit Project (MINAP). *Age Ageing* **43,** 779–85 (2014).

14. Simsek, C. *et al.* The unrestricted use of sirolimus- and paclitaxel-eluting stents results in better clinical outcomes during 6-year follow-up than bare-metal stents: an analysis of the RESEARCH (Rapamycin-Eluting Stent Evaluated At Rotterdam Cardiology Hospital) and T-SEARCH (Taxus-Stent Evaluated At Rotterdam Cardiology Hospital) registries. *JACC Cardiovasc Interv* **3,** 1051–8 (2010).

15. Lemos, P. A. *et al.* Early outcome after sirolimus-eluting stent implantation in patients with acute coronary syndromes: insights from the Rapamycin-Eluting Stent Evaluated At Rotterdam Cardiology Hospital (RESEARCH) registry. *J Am Coll Cardiol* **41,** 2093–9 (2003).

16. ESC. *Clinical Practice Guidelines* `http://www.escardio.org/Guidelines/Clinical-Practice-Guidelines` (2017).

17. Hieke, S., Kleber, M., Konig, C., Engelhardt, M. & Schumacher, M. Conditional Survival: A Useful Concept to Provide Information on How Prognosis Evolves over Time. *Clin Cancer Res* **21,** 1530–6 (2015).

18. CBS. *CBS population figures* `http://www.cbs.nl/en-GB/menu/themas/bevolking/cijfers/default.htm` (2016).

19. Benjamin, E. J. *et al.* Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association. *Circulation* **135,** e146–e603 (2017).

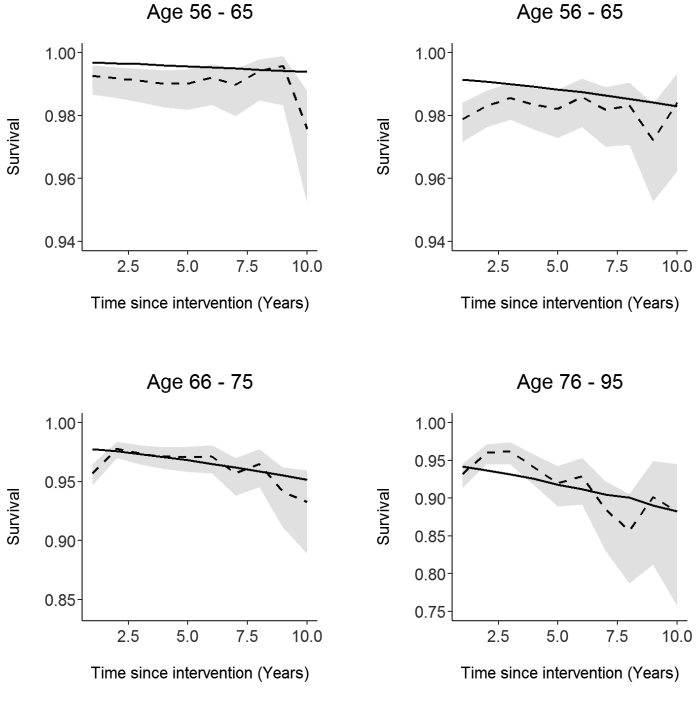
20. Dickman, P. W., Sloggett, A., Hills, M. & Hakulinen, T. Regression models for relative survival. *Stat Med* **23,** 51–64 (2004).

21. Dickman, P. Estimating and modelling relative survival using SAS (2004).

22. Velders, M. A. *et al.* Prognosis of elderly patients with ST-elevation myocardial infarction treated with primary percutaneous coronary intervention in 2001 to 2011: A report from the Swedish Coronary Angiography and Angioplasty Registry (SCAAR) registry. *Am Heart J* **167,** 666–73 (2014).

Supplemental Material

The solid line depicts the age adjusted expected survival and the dashed line depicts the observed survival with the grey band the 95% confidence interval.



Supplementary Figure 1. Observed and expected survival for STEMI patients.

Supplemental Table 1: Adjusted expected 1-year survival probabilities

Indication for PCI	Age	1 year	1 year, 1 month	2 years	5 years	10 years
STEMI	22-55	100%	100%	100%	100%	99%
	56-65	99%	99%	99%	99%	98%
	66-75	98%	98%	98%	97%	95%
	76-95	93%	93%	93%	91%	87%
NSTEMI	22-55	100%	100%	100%	100%	99%
	56-65	99%	99%	99%	99%	98%
	66-75	98%	98%	98%	97%	95%
	76-95	94%	94%	94%	92%	88%
SAP	22-55	100%	100%	100%	100%	99%
	56-65	99%	99%	99%	99%	98%
	66-75	98%	98%	98%	97%	95%
	76-95	94%	94%	94%	92%	88%

STEMI, ST-elevation myocardial infarction; NSTEMI, non STEMI; SAP, stable angina pectoris

Supplemental Table 2: Cumulative survival probabilities stratified by indication for PCI and age

			Cumulative Deaths at					Survival at (95% Confidence Interval)				
Indication for PCI	Age, y	Alive at start	5 y	5 y, 1 mo	6 y	10 y	13 y	5 y	5 y, 1 mo	6 y	10 y	13 y
STEMI	22-55	2071	76	77	82	114	116	96% (94%-97%)	95% (94%-96%)	95% (93%-96%)	89% (86%-91%)	88% (85%-90%)
	56-65	1661	107	107	118	144	147	92% (90%-93%)	92% (90%-93%)	90% (87%-91%)	82% (79%-85%)	80% (75%-84%)
	66-75	1299	203	204	211	238	243	81% (78%-83%)	81% (78%-83%)	79% (76%-82%)	69% (64%-73%)	65% (60%-70%)
	76-95	885	243	245	252	273	274	64% (59%-68%)	63% (58%-67%)	58% (53%-64%)	36% (28%-44%)	*
NSTEMI	22-55	1286	54	54	60	78	84	95% (93%-96%)	95% (93%-96%)	94% (92%-95%)	89% (86%-92%)	84% (79%-89%)
	56-65	1517	122	122	129	158	161	90% (88%-92%)	90% (88%-92%)	89% (87%-91%)	80% (76%-84%)	77% (71%-82%)
	66-75	1495	208	208	230	280	294	82% (80%-85%)	82% (80%-85%)	79% (76%-81%)	63% (58%-67%)	50% (42%-58%)
	76-95	1043	259	259	277	324	329	68% (64%-71%)	68% (64%-71%)	62% (57%-66%)	36% (29%-42%)	20% (10%-33%)

SAP	22-55	1494	57	57	64	86	92	96% (94%-97%)	96% (94%-97%)	95% (93%-96%)	91% (88%-93%)	88% (85%-91%)
	56-65	2079	150	150	164	209	220	92% (90%-93%)	92% (90%-93%)	90% (89%-92%)	83% (81%-86%)	75% (69%-80%)
	66-75	1977	235	237	258	339	354	86% (84%-88%)	86% (84%-87%)	83% (81%-85%)	68% (64%-71%)	57% (51%-63%)
	76-95	969	192	195	212	268	271	74% (71%-78%)	74% (70%-77%)	69% (65%-73%)	42% (35%-48%)	32% (18%-47%)

PCI, percutaneous coronary intervention; STEMI, ST-elevation myocardial infarction;

NSTEMI, non STEMI; SAP, stable angina pectoris

Supplemental Table 3: Conditional 5-year survival probabilities stratified by indication for PCI and age

		Alive at start of interval (Deaths during Interval)* Alive at					Conditional Survival (95% Confidence Interval) 5-y Survival Conditional on Surviving:					Relative Survival (95% Confidence Interval) 5-y Survival Conditional on Surviving:				
Ind	Age	0 y	1 mo	1 y	4 y	9 y	0 y	1 mo	1 y	4 y	9 y	0 y	1 mo	1 y	4 y	9 y
STEMI	22-55	2071	2041	1961	735	485	96%	97%	97%	93%	96%	97%	98%	99%	95%	99%
		(76)	(50)	(39)	(38)	(18)	(94%-97%)	(96%-98%)	(96%-98%)	(90%-95%)	(93%-97%)	(96%-98%)	(97%-99%)	(97%-99%)	(93%-97%)	(96%-100%)
	56-65	1661	1622	1548	491	315	92%	94%	93%	90%	91%	96%	99%	98%	97%	100%
		(107)	(70)	(63)	(37)	(22)	(90%-93%)	(92%-95%)	(91%-94%)	(86%-93%)	(86%-94%)	(94%-98%)	(97%-100%)	(96%-100%)	(93%-100%)	(94%-104%)
	66-75	1299	1234	1142	316	186	81%	85%	87%	85%	88%	92%	97%	101%	105%	119%
		(203)	(139)	(86)	(35)	(18)	(78%-83%)	(82%-87%)	(84%-90%)	(80%-89%)	(82%-93%)	(89%-95%)	(94%-100%)	(97%-104%)	(98%-110%)	(110%-125%)
	76-95	885	798	711	100	40	64%	70%	70%	*	*	95%	104%	107%	*	*
		(243)	(158)	(105)	(30)	(10)	(59%-68%)	(64%-74%)	(63%-76%)			(88%-102%)	(96%-111%)	(97%-116%)		
NSTEMI	22-55	1286	1274	1209	531	290	95%	96%	96%	94%	92%	97%	97%	98%	96%	95%
		(54)	(44)	(31)	(24)	(12)	(93%-96%)	(94%-97%)	(94%-97%)	(91%-96%)	(86%-96%)	(95%-98%)	(96%-98%)	(96%-99%)	(93%-98%)	(88%-99%)
	56-65	1517	1503	1433	545	251	90%	91%	92%	89%	90%	95%	96%	97%	96%	98%
		(122)	(108)	(83)	(36)	(16)	(88%-92%)	(89%-93%)	(90%-93%)	(85%-92%)	(82%-94%)	(93%-96%)	(94%-97%)	(95%-98%)	(91%-99%)	(90%-103%)
	66-75	1495	1471	1380	463	219	82%	84%	83%	76%	*	94%	96%	96%	95%	*
		(208)	(184)	(156)	(72)	(40)	(80%-85%)	(81%-86%)	(80%-85%)	(71%-81%)		(91%-97%)	(93%-98%)	(92%-98%)	(88%-101%)	
	76-95	1043	996	886	201	74	68%	71%	70%	52%	*	97%	102%	103%	*	*
		(259)	(212)	(150)	(65)	(22)	(64%-71%)	(67%-74%)	(65%-74%)	(43%-61%)		(91%-102%)	(96%-107%)	(96%-109%)		

SAP	22-55	1494	1490	1460	883	472	96%	96%	96%	95%	94%	97%	98%	98%	97%	98%
		(57)	(54)	(53)	(29)	(16)	(94%-97%)	(95%-97%)	(94%-97%)	(92%-97%)	(90%-97%)	(96%-98%)	(96%-99%)	(96%-99%)	(95%-99%)	(94%-100%)
	56-65	2079	2066	2007	993	468	92%	92%	92%	91%	86%	96%	97%	97%	98%	95%
		(150)	(138)	(120)	(59)	(29)	(90%-93%)	(91%-93%)	(91%-93%)	(88%-93%)	(78%-91%)	(95%-98%)	(95%-98%)	(96%-99%)	(95%-100%)	(86%-100%)
	66-75	1977	1959	1848	802	343	86%	86%	87%	79%	74%	98%	99%	101%	97%	*
		(235)	(221)	(173)	(104)	(49)	(84%-88%)	(85%-88%)	(85%-89%)	(74%-82%)	(66%-81%)	(96%-100%)	(97%-101%)	(99%-103%)	(92%-102%)	
	76-95	969	955	878	280	81	74%	75%	74%	56%	*	107%	107%	110%	*	*
		(192)	(181)	(146)	(76)	(17)	(71%-78%)	(71%-78%)	(70%-78%)	(47%-64%)		(102% -111%)	(102%-112%)	(104%-115%)		

5-year survival probabilities conditional on having survived certain periods of time. The first column indicates the probability of surviving the first five years from the start of the study (conditional on 0 years). The probabilities can also be interpreted as – for the last column for example – surviving 13 years, given survival of the first 8 years. Relative survival is the survival compared to the general population with the same age and gender. PCI indicates percutaneous coronary intervention; STEMI, ST-elevation myocardial infarction; NSTEMI, non STEMI; SAP, stable angina pectoris

*Standard error $> 5\%$

† Deaths during interval indicate the number of deaths for each 5-year interval per group. For example, for STEMI patients aged 22-55, 76 patients died in the first five years, 50 patients died in the period between one month and five years and a month, etcetera.

Supplemental Table 4: Cumulative survival probabilities stratified by gender and age

			Cumulative Deaths at						Survival at (95% Confidence Interval)					
Gender	Age, y	Alive at start	1 y	1 y, 1 mo	2 y	5 y	10 y	13 y	1 y	1 y, 1 mo	2 y	5 y	10 y	13y
Male	22-55	3928	59	62	89	146	221	231	98% (98%-99%)	98% (98%-99%)	98% (97%-98%)	95% (95%-96%)	90% (88%-91%)	87% (85%-89%)
	56-65	4036	94	102	157	277	387	396	98% (97%-98%)	97% (97%-98%)	96% (95%-97%)	91% (90%-92%)	82% (79%-84%)	77% (74%-81%)
	66-75	3198	202	213	293	462	609	632	94% (93%-94%)	93% (92%-94%)	91% (90%-92%)	83% (81%-84%)	65% (62%-68%)	56% (50%-61%)
	76-95	1650	211	216	290	433	525	531	87% (86%-89%)	87% (85%-88%)	82% (80%-84%)	66% (63%-69%)	35% (30%-41%)	*
Female	22-55	923	24	24	29	41	57	61	97% (96%-98%)	97% (96%-98%)	97% (95%-98%)	95% (93%-96%)	90% (87%-92%)	84% (76%-90%)
	56-65	1221	51	53	65	102	124	132	96% (95%-97%)	96% (94%-97%)	95% (93%-96%)	90% (88%-92%)	84% (80%-87%)	72% (63%-80%)
	66-75	1573	82	87	114	184	248	259	95% (94%-96%)	94% (93%-95%)	93% (91%-94%)	85% (83%-87%)	69% (65%-73%)	57% (49%-65%)
	76-95	1247	129	134	172	261	340	343	90% (88%-91%)	89% (87%-91%)	86% (84%-88%)	73% (70%-76%)	41% (34%-47%)	37% (29%-44%)

Supplemental Table 5: Conditional 1-year survival probabilities stratified by gender and age

		Alive at start of interval (Deaths during Interval)* Alive at					Conditional Survival (95% Confidence Interval) 1-y Survival Conditional on Surviving:					Relative Survival (95% Confidence Interval) 1-y Survival Conditional on Surviving:				
Gender	Age	0 y	1 mo	1 y	4 y	9 y	0 y	1 mo	1 y	4 y	9 y	0 y	1 mo	1 y	4 y	9 y
Male	22-55	3928	3895	3764	2036	829	98%	99%	99%	99%	98%	99%	99%	100%	99%	99%
		(59)	(35)	(30)	(19)	(17)	(98%-99%)	(99%-99%)	(99%-99%)	(99%-99%)	(97%-99%)	(98%-99%)	(99%-100%)	(99%-100%)	(99%-100%)	(97%-99%)
	56-65	4036	3987	3843	1910	591	98%	99%	98%	98%	97%	99%	99%	99%	100%	99%
		(94)	(56)	(63)	(33)	(16)	(97%-98%)	(98%-99%)	(98%-99%)	(98%-99%)	(96%-98%)	(98%-99%)	(99%-100%)	(99%-100%)	(99%-100%)	(97%-100%)
	66-75	3198	3121	2924	1313	358	94%	96%	97%	97%	95%	96%	98%	100%	101%	100%
		(202)	(138)	(91)	(41)	(19)	(93%-94%)	(95%-96%)	(96%-97%)	(96%-98%)	(92%-97%)	(95%-97%)	(97%-99%)	(99%-100%)	(99%-101%)	(97%-102%)
	76-95	1650	1564	1393	423	64	87%	92%	94%	92%	91%	94%	99%	102%	101%	105%
		(211)	(130)	(79)	(35)	(6)	(86%-89%)	(90%-93%)	(93%-95%)	(89%-94%)	(80%- 96%)	(92%-96%)	(97%-100%)	(101%-103%)	(98%-104%)	(93%-111%)
Female	22-55	923	910	866	466	169	97%	99%	99%	100%	99%	98%	99%	100%	100%	100%
		(24)	(11)	(5)	(1)	(1)	(96%-98%)	(98%-99%)	(99%-100%)	(98%-100%)	(96%-100%)	(96%-98%)	(98%-100%)	(99%-100%)	(99%-100%)	(96%-100%)
	56-65	1221	1204	1145	575	187	96%	97%	99%	99%	98%	96%	98%	99%	100%	99%
		(51)	(36)	(14)	(5)	(4)	(95%-97%)	(96%-98%)	(98%-99%)	(98%-100%)	(94%-99%)	(95%-97%)	(96%-98%)	(99%-100%)	(99%-100%)	(96%-100%)
	66-75	1573	1543	1446	693	170	95%	96%	98%	97%	93%	96%	98%	99%	99%	97%
		(82)	(57)	(32)	(24)	(12)	(94%-96%)	(95%-97%)	(97%-98%)	(95%-98%)	(88%-96%)	(95%-97%)	(97%-99%)	(98%-100%)	(97%-100%)	(91%-100%)
	76-95	1247	1185	1082	380	69	90%	94%	96%	94%	84%	94%	99%	101%	101%	94%
		(129)	(72)	(43)	(22)	(11)	(88%-91%)	(92%-95%)	(95%-97%)	(91%-96%)	(73%-91%)	(92%-96%)	(97%-100%)	(100%-103%)	(98%-103%)	(82%-102%)

1-year survival probabilities conditional on having survived certain periods of time. The first column indicates the probability of surviving the first year from the start of the study (conditional on 0 years). The probabilities can also be interpreted as – for the last column for example – surviving 10 years, given survival of the first 9 years. Relative survival is the survival compared to the general population with the same age and gender.

*Standard error $> 5\%$

† Deaths during interval indicate the number of deaths for each 1-year interval per group. For example, for males aged 22-55, 59 patients died in the first year, 35 patients died in the period between one month and one year and a month, etcetera.

Supplemental Table 6: Conditional 5-year survival probabilities stratified by gender and age

		Alive at start of interval (Deaths during Interval)* Alive at					Conditional Survival (95% Confidence Interval) 5-y Survival Conditional on Surviving:					Relative Survival (95% Confidence Interval) 5-y Survival Conditional on Surviving:				
Gender	Age	0 y	1 mo	1 y	5 y	8 y	0 y	1 mo	1 y	5 y	8 y	0 y	1 mo	1 y	5 y	8 y
Male	22-55	3928	3895	3764	1734	1020	95%	96%	96%	94%	94%	97%	98%	98%	96%	98%
		(146)	(120)	(104)	(75)	(38)	(95%-96%)	(95%-97%)	(95%-97%)	(92%-95%)	(92%-96%)	(96%-98%)	(97%-98%)	(97%-99%)	(95%-98%)	(95%-100%)
	56-65	4036	3987	3843	1548	780	91%	93%	92%	89%	89%	96%	98%	98%	97%	99%
		(277)	(231)	(207)	(110)	(50)	(90%-92%)	(92%-93%)	(91%-93%)	(87%-91%)	(85%-93%)	(95%-97%)	(97%-99%)	(96%-99%)	(94%-99%)	(94%-103%)
	66-75	3198	3121	2924	1037	489	83%	84%	85%	79%	76%	96%	98%	100%	101%	104%
		(462)	(390)	(302)	(147)	(73)	(81%-84%)	(83%-86%)	(83%-86%)	(76%-82%)	(68%-82%)	(94%-98%)	(97%-100%)	(98%-102%)	(96%-105%)	(93%-112%)
	76-95	1650	1564	1393	303	87	66%	69%	70%	54%	*	101%	106%	112%	*	*
		(433)	(349)	(242)	(92)	(20)	(63%-69%)	(66%-72%)	(67%-74%)	(46%-61%)		(96%-105%)	(101%-111%)	(106%-117%)		
Female	22-55	923	910	866	415	227	95%	96%	97%	95%	92%	96%	98%	98%	96%	94%
		(41)	(28)	(19)	(16)	(8)	(93%-96%)	(95%-97%)	(95%-98%)	(91%-97%)	(82%-97%)	(94%-98%)	(96%-99%)	(97%-100%)	(93%-99%)	(83%-99%)
	56-65	1221	1204	1145	481	254	90%	92%	93%	93%	83%	94%	95%	96%	98%	*
		(102)	(85)	(59)	(22)	(17)	(88%-92%)	(90%-93%)	(90%-94%)	(89%-95%)	(70%-90%)	(91%-95%)	(93%-97%)	(94%-98%)	(94%-101%)	
	66-75	1573	1543	1446	544	259	85%	87%	88%	81%	*	94%	96%	98%	94%	*
		(184)	(154)	(113)	(64)	(34)	(83%-87%)	(85%-89%)	(86%-90%)	(76%-85%)		(91%-96%)	(93%-98%)	(95%-100%)	(88%-99%)	
	76-95	1247	1185	1082	278	108	73%	76%	74%	55%	*	98%	103%	102%	*	*
		(261)	(202)	(159)	(79)	(29)	(70%-76%)	(73%-79%)	(69%-77%)	(47%-63%)		(94%-102%)	(98%-107%)	(96%-107%)		

5-year survival probabilities conditional on having survived certain periods of time. The first column indicates the probability of surviving the first five years from the start of the study (conditional on 0 years). The probabilities can also be interpreted as – for the last column for example – surviving 13 years, given survival of the first 8 years. Relative survival is the survival compared to the general population with the same age and gender.

*Standard error $> 5\%$

† Deaths during interval indicate the number of deaths for each 1-year interval per group. For example, for males aged 22-55, 146 patients died in the first five years, 120 patients died in the period between one month and five years and a month, etcetera.

Chapter 3

Relative conditional survival analysis provides additional insights into the prognosis of heart failure patients

Baart SJ, van den Berge JC, Akkerhuis KM, Deckers JW, van Domburg RT, Boersma H, Kardys I

Submitted

Abstract

Background: Heart failure is a diagnosis that entails high mortality. Cumulative survival is the standard method of reporting survival information and communicating prognosis to the patient. Additional survival measures could provide further insights into prognosis. Such measures have proven to be useful in oncology.

Methods and results: In total 1810 consecutive patients admitted with acute heart failure were included between 1985 and 2008. Three measures of survival were calculated; standard cumulative survival using the Kaplan-Meier method until 10 years after diagnosis, 1-year conditional survival measures, and 1-year relative conditional survival measures. One-year survival rates ranged from 56% to 65%, depending on age category. Cumulative survival at ten years follow-up ranged from 9% to 35%. If patients survived the first year, their 1-year conditional survival rates ranged from 79% to 89% and remained stable throughout follow-up. However, the relative 1-year survival rates remained below 1.

Conclusion: Conditional and relative conditional survival estimates could be useful for communicating prognosis to heart failure patients. The conditional survival estimates showed that the first period after diagnosis is the most critical, and survival is higher and stable after having survived this period. Relative conditional survival analyses demonstrated however, that the survival of heart failure patients always remains lower than that of the general population. This holds particular relevance for younger patients, who will carry the burden of their disease for the rest of their lives.

Introduction

Heart failure (HF) is a diagnosis associated with high short- and long-term mortality rates. [1] Mortality rates presented in literature are usually based on cumulative probabilities, providing statistical estimates of mortality for the entire duration of the follow-up period. However, by presenting survival this way, certain aspects of prognosis in HF are missed that may contain important information for the patient or treating physician. Additional insights into prognosis may be gained by using relative and conditional survival analysis, approaches commonly used and proven highly useful in oncology. [2] Studies applying these approaches in HF patients are scarce. We aimed to explore these measures of survival in a large cohort of patients with acute HF.

Methods

We used a single-center cohort of acute HF patients with long-term follow-up on survival. [3] In short, patients admitted with acute heart failure to the Intensive Coronary Care Unit (ICCU) of the Erasmus MC, Rotterdam, the Netherlands, were included in a prospective registry between 1st January 1985 and 31st December 2008. Consecutive patients were included to reduce the chance of selection bias. Patients were excluded if their HF was caused by acute coronary syndrome and if there was no evidence of sustained systolic of diastolic dysfunction. In January 2017, vital status of the patients was obtained from the Municipal Civil Registries. Since this cohort was a registry, and patients were not subjected to acts, nor was any mode of behavior imposed, approval from the local research ethics committee was not required. The study was conducted in accordance with the Declaration of Helsinki. [4] To assess prognosis of the patients, we calculated and compared three types of survival measures. First

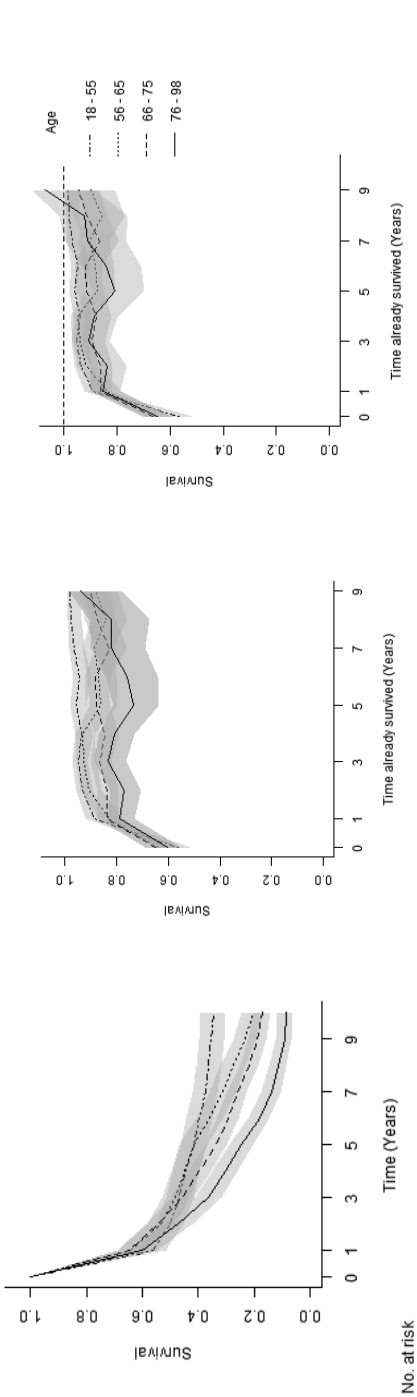
we estimated the standard cumulative survival using the Kaplan-Meier method. Standard errors were calculated using Greenwood's formula. [5] Second, we calculated conditional survival, which incorporates the time already survived since diagnosis, and herewith shows how a patient's estimated survival changes over time. For example, the 1-year survival probability, conditional on having already survived 5 years, is calculated by dividing the survival at 5+1 years by the survival at 5 years. The third measure of survival was relative conditional survival. When estimating relative survival, a patient's survival is directly compared to the survival of a person in the general population of the same age and gender, retrieved from Statistics Netherlands (Centraal Bureau voor de Statistiek). [6] Relative conditional survival is obtained similar to conditional survival, however relative survival estimates are used instead of the observed survival estimates. As such, relative conditional survival demonstrates at which point in time the survival probability of a patient becomes similar to that of someone in the general population. [7] Survival estimates were considered reliable and were reported if the standard error was $\leq$5%. Analyses were performed in SAS version 9.4, where a publicly available macro was used for relative survival. [8] The graphs were made with R version 3.5.0.

Results

The cohort consisted of 1810 patients. Baseline characteristics have been described previously. [3] Mean (standard deviation) age was 64 (15) years, 64% were male and 41% had ischemic cardiomyopathy. No information on the race or ethnicity of the patients was collected. For the analyses, patients were stratified in four age categories (18-55, 56-65, 66-65, 76-98 years). Information on age was available for all patients in the cohort. Due to the limited number of patients still available further along in the follow-up, patients were not additionally stratified

on gender. For 41 patients (2.3%) no follow-up information was available after diagnosis and they were removed for this analysis. Twenty-five patients (1.4%) were lost to follow-up and the last recorded date on which they were still alive was used. The median follow-up time was 28 months (IQR 3 – 95 months). Over the course of the follow-up period 1572 patients (87%) experienced the combined end point of implantation of a left ventricular assist device (LVAD), heart transplantation or death. For this analysis we considered LVAD and heart transplantation to be equivalent to death. Cumulative survival rates until 10 years after diagnosis of acute heart failure, are shown in the Figure, panel A and the Supplemental Table. Survival rates were low, especially in the first year after inclusion. One-year survival rates ranged from 56% to 65%, and survival ranged from 9% to 35% at ten years of follow-up. Overall, younger patients had better survival than older patients. The steep decline in survival observed in the first year attenuated during subsequent follow-up. This became even more evident when one year conditional survival rates were calculated (Panel B in the Figure and the Supplemental Table). For example, a patient aged 55 or younger had a cumulative survival probability of 56% at one year of follow-up and of 50% at two years of follow-up. However, if that patient survived the first year, the survival probability of also surviving the second year changed to 89%. In fact, all 1-year estimates conditional on surviving the first year ranged from 79% to 89%. These much higher one-year survival rates persisted throughout the remainder of the follow-up (panel B). Younger patients had higher one-year conditional survival rates than the older age groups. How much of the all-cause mortality in the current cohort can be attributed to the diagnosis of heart failure is shown by relative (conditional) survival. Panel C in the Figure and the Supplemental Table show the one-year relative conditional survival estimates. For example, a patient aged 75 or older had a one-year conditional

survival probability of 79%, but relative to the general population this was 85%. The relative conditional survival probability lines are shifted upwards towards one compared to the conditional one-year survival in Panel B. It is clear, however, that in general the lines never actually reach the value of one, indicating that during the entire 15 year follow-up period patients continue to have an increased mortality rate compared to the general population. Moreover, relative survival rates show more overlap for the different patient groups than they do in the conditional survival analysis. This means that although younger patients have better prognosis than older patients, compared to their peers survival remains suboptimal and increased heart failure mortality persists even years after diagnosis.



(a) (b) (c)

Figure. Cumulative, conditional and relative conditional survival curves
Panel A shows cumulative survival in years (with 95% confidence intervals) for four different age groups. It also shows the number of patients still at risk at 0, 5, and 10 years. Panel B shows the 1-year survival estimates, conditional on surviving the number of years indicated on the x-axis (with 95% confidence intervals). Panel C shows the 1-year conditional survival curves relative to the general population. Additionally in panel C, the line is plotted that indicates a survival equal to that of the general population (relative survival = 1).

Discussion

In this investigation we reported several measures of prognosis in patients with acute HF. The standard cumulative survival curves showed that prognosis was poor and that ten year survival rates were low (<35%). However, by calculating conditional survival, we showed that mortality mostly occurred in the first year after diagnosis. Once patients survive the first crucial year, their estimated survival improved greatly. Moreover, the 1-year conditional survival rates remained more or less stable during follow-up. On the other hand, when compared to the general population, the survival rates of HF patients never reached the population level. Our report has several limitations. First, since this was a single-center cohort study, generalizability of the results is limited. Second, when calculating conditional survival, multiple age categories were used; however the number of patients included in this study did not permit additional stratification on gender or etiology of HF. Moreover, the sample size also prohibited calculation of reliable survival estimates in the later years of follow up. Larger registries are necessary for these purposes.

Conclusion

In conclusion, acute HF is a diagnosis with high mortality. Additional to standard mortality estimates, (relative) conditional survival can be used to gain further insights into the disease course and communicate this more clearly to patients. Using these measures, we found that once patients are diagnosed with acute heart failure, their survival never becomes 'normal' again, even after surviving the first, crucial year. This notion is particularly important for patients diagnosed at a young age.

References

1. Bui, A. L., Horwich, T. B. & Fonarow, G. C. Epidemiology and risk profile of heart failure. *Nat Rev Cardiol* **8,** 30–41 (2011).
2. Nelson, C. P., Lambert, P. C., Squire, I. B. & Jones, D. R. Relative survival: what can cardiovascular disease learn from cancer? *Eur Heart J* **29,** 941–7 (2008).
3. Van den Berge, J. C. *et al.* Temporal trends in long-term mortality of patients with acute heart failure: Data from 1985-2008. *Int J Cardiol* **224,** 456–460 (2016).
4. Goodyear, M. D., Krleza-Jeric, K. & Lemmens, T. The Declaration of Helsinki. *BMJ* **335,** 624–5 (2007).
5. Cho, H., Howlader, N., Mariotto, A. B. & Cronin, K. A. *Estimating relative survival for cancer patients from the SEER Program using expected rates based on Ederer I versus Ederer II method* Report (NCI, 2011).
6. CBS. *CBS population figures* `http://www.cbs.nl/en-GB/menu/themas/bevolking/cijfers/default.htm` (2016).
7. Baart, S. J. *et al.* Impact of Relative Conditional Survival Estimates on Patient Prognosis After Percutaneous Coronary Intervention. *Circ Cardiovasc Qual Outcomes* **10(6)** (2017).
8. Dickman, P. Estimating and modelling relative survival using SAS (2004).

Supplemental Material

Supplemental Table. Cumulative, conditional and relative conditional survival estimates

	Cumulative survival at:				Conditional Survival (95% CI) 1y survival, conditional on surviving:				Relative Conditional Survival (95% CI) 1y survival, conditional on surviving:			
Age	1y	2y	5y	10y	0y	1y	4y	9y	0y	1y	4y	9y
18 - 55	56%	50%	41%	35%	56%	89%	94%	98%	57%	89%	94%	99%
	(51%-1%)	(45%-5%)	(37%-6%)	(30%-9%)	(51%-61%)	(84%-92%)	(89%-96%)	(94%-99%)	(52%-61%)	(84%-92%)	(60%-97%)	(95%-100%)
56 - 65	64%	53%	42%	21%	64%	83%	93%	88%	65%	84%	95%	90%
	(59%-69%)	(48%-58%)	(37%-47%)	(17%-25%)	(59%-69%)	(78%-87%)	(89%-96%)	(79%-94%)	(60%-70%)	(79%-88%)	(90%-97%)	(81%-95%)
66 - 75	65%	55%	34%	17%	65%	84%	85%	90%	67%	86%	88%	95%
	(61%-69%)	(50%-59%)	(30%-38%)	(14%-20%)	(61%-69%)	(80%-87%)	(79%-89%)	(83%-95%)	(62%-71%)	(82%-90%)	(82%-92%)	(87%-99%)
76 - 98	60%	47%	25%	9%	60%	79%	81%	94%	65%	85%	89%	107%
	(55%-65%)	(42%-52%)	(21%-29%)	(6%-12%)	(55%-65%)	(73%-83%)	(73%-87%)	(78%-98%)	(59%-70%)	(79%-90%)	(80%-95%)	(89%-112%)

The three types of survival measures calculated at selected point during follow-up. Cumulative survival is the standard survival method calculated using the Kaplan-Meier method. One year conditional survival estimated are conditional on having survived the previous period. Relative conditional survival estimates related the survival to a comparable person from the general population. Survival estimates include the 95% confidence interval.

Part III: Predicting outcomes for cardiovascular disease in women

Chapter 4

Cardiovascular risk prediction models for women in the general population: a systematic review

Baart SJ*, Dam V*, Scheres LJJ*, Damen JAAG, Spijker R, Schuit E, Debray TPA, Fauser BCJM, Boersma H, Moons KGM, van der Schouw YT, on behalf of the CREW consortium

*These authors contributed equally

Plos ONE 2019; 14(1)

Abstract

Aim: To provide a comprehensive overview of cardiovascular disease (CVD) risk prediction models for women and models that include female-specific predictors.

Methods and Results: We performed a systematic review of CVD risk prediction models for women in the general population by updating a previous review. We searched Medline and Embase up to July 2017 and included studies in which; (a) a new model was developed, (b) an existing model was validated, or (c) a predictor was added to an existing model. A total of 285 prediction models for women have been developed, of these 160 (56%) were female-specific models, in which a separate model was developed solely in women and 125 (44%) were sex-predictor models. Out of the 160 female-specific models, 2 (1.3%) included one or more female-specific predictors (mostly reproductive risk factors). A total of 591 validations of sex-predictor or female-specific models were identified in 206 papers. Of these, 333 (56%) validations concerned nine models (five versions of Framingham, SCORE, Pooled Cohort Equations and QRISK). The median and pooled C statistics were comparable for sex-predictor and female-specific models. In 260 articles the added value of new predictors to an existing model was described, however in only 3 of these female-specific predictors (reproductive risk factors) were added.

Conclusions: There is an abundance of models for women in the general population. Female-specific and sex-predictor models have similar predictors and performance. Female-specific predictors are rarely included. Further research is needed to assess the added value of female-specific predictors to CVD models for women and provide physicians with a well-performing prediction model for women.

Introduction

Differences between women and men in cardiovascular disease (CVD) have been recognized decades ago, [1] pertaining to clinical presentation, pathophysiological mechanisms, course of disease and prognosis. [2–6] As symptoms of CVD are more subtle in women, there is often delayed diagnosis, and thus treatment and consequently poorer prognosis and outcomes compared with men. [7] It is crucial to identify sex differences to optimize diagnostic and management strategies for both women and men. [8] Although women and men share many CVD risk factors, which are often used in prediction models for the general population, there are also female-specific risk factors. Well known examples are early menarche and menopause, primary ovarian insufficiency, pregnancy complications, polycystic ovary syndrome, and use of hormones. [9–11] Preventive measures are available to reduce the cardiovascular disease burden. Numerous strategies to reduce the CVD burden have been implemented to identify persons at high risk. As seen in a systematic review published in 2016, over 350 prediction models have been developed in recent years aiming to identify individuals at high CVD risk in the general population. [12] Guidelines in Europe and the Unites States currently recommend the use of Systematic COronary Risk Evaluation (SCORE) or the Pooled Cohort Equations in the general population, both for women and men. [13, 14] Although several female-specific CVD risk factors have been identified, predictors in most implemented CVD prediction models seem generally similar for women and men. As clinical presentation, pathophysiological mechanisms, course of disease and prognosis differ between women and men; risk prediction likely differs between the sexes as well. Therefore, we aimed to provide an overview of available CVD risk prediction models for women and of models that include female-specific predictors.

Methods

Systematic literature search

For this review we used the results of the review by Damen et al. on all future CVD prediction models for the general population, both men and women. [12] As shown by this review, the number of newly developed CVD prediction models grew excessively in recent years. For this reason, we complemented the results of Damen et al., by performing an update of their search. Details of the review by Damen et al. were published previously. [12] In the original search, Medline and Embase were searched until June 1st 2013 in order to identify articles on prediction models for the occurrence of CVD in the general population, published after 2004. Articles which dated before 2004 were subtracted from the review by Beswick et al. [15] Articles were included when they reported one or more multivariable (i.e. including at least 2 predictors) prediction models, tools or scores to predict future CVD in the general population (development papers), articles that investigated the added value of certain predictors (incremental value papers) and articles that validated existing models (validation papers). Table 1 provides an overview of the key terminology. For the present systematic review, we updated the search of Damen et al. until 26th of July 2017. Title and abstract screening were conducted using the same in- and exclusion criteria as Damen et al. However, in the full text screening we included only models specifically developed to predict CVD in women. We defined 'model developed for women' as 1) female-specific models, in which a separate model was developed in women only and 2) sex-predictor models, in which sex was included as a predictor (e.g. covariate) in the model (Table 1). Models that were developed on men only or models that did not include sex as a predictor were excluded. For the validation papers, only studies that validated a prediction model developed for women

were included. Studies in which a predictor was added to an existing model (incremental value papers) were also included. Incremental value or validation studies in men only were excluded.

Screening and data extraction

The titles and abstracts retrieved by the search were divided randomly among the researchers (SJB, VD or LJJS) and screened independently. Studies were not screened in duplicate, but to guarantee uniformity in screening, 30 abstracts were screened by all three researchers and discussed afterwards. In the screening stage, all papers that were labeled as 'any doubt' were included for full text screening. For full text screening the papers were divided in three different subsets for independent screening by one of the three researchers (SJB, VD or LJJS). Again, full text screening was not performed in duplicate, a subset of 20 papers from each researcher was screened by all three researchers to achieve uniformity. Articles labeled as "any doubt" were resolved by discussion among the three reviewers to reach consensus. Hand searching based on included articles and 'snowballing' were used to search for additional studies.

Finally, data extraction was performed in a pre-specified data-extraction format based on the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS). [16] All three reviewers read the papers and subsequently filled in the data-extraction format together to guarantee agreement on the extracted information. In this stage, disagreements were settled by an additional reviewer (JAAGD or YTvdS). For papers in which a model was developed we extracted the same information as Damen et al. and additionally determined whether the model was a female-specific or sex-predictor model. All developed models were then assessed for quality based on reliability defined as 1) model externally validated 2) model

Table 1: Key definitions

Model developed for women	A model developed for women, either separately for women (female-specific model) or where sex is incorporated as a predictor (sex-predictor model)
Female-specific model	A model developed in a dataset of women only, with a separate regression model or risk chart for women
Sex-predictor model	A model developed in a dataset of women and men, which uses sex as a predictor in the model
Development	When a new model is derived from a dataset
Incremental value paper	When one or more predictors are added to an existing model to study whether the performance of the model improves after adding the predictor(s)
Validation paper	When the performance of an existing model is verified in a different population
Female-specific predictor	A risk factor that is very clearly female specific such as: early menarche and menopause, primary ovarian insufficiency, pregnancy complications, and polycystic ovary syndrome
Discrimination	Indicates how well the model distinguishes between persons with an outcome event and persons without an outcome event, often depicted as the C statistic
C statistic	Measure of discrimination of the model and quantifies the area under the receiver operator curve (ROC). Ranges from 0.5 to 1.0, where 0.5 resembles a coin-toss and 1.0 is a perfect discrimination.

externally validated in a separate investigation/paper and 3) C statistic >0.7. If the development model did not report a C statistic, we used the mean C statistic of the external validations. Reliable models, which met these criteria were assessed for clinical usability for 1) 10 predictors or fewer, 2) full regression model or chart reported and 3) availability of an online calculator. For every

included incremental value paper we extracted author, year, journal, the model that was used to calculate incremental value and whether this model was female-specific or sex-predictor and which predictors actually had incremental value. In addition, predictors considered for incremental value were also extracted. Finally, for the validation papers we extracted author, year, journal and which model was validated. For the models that were validated >5 times and at least once in an external study, we subsequently extracted additional information: characteristics of the validation cohort (country, number of participants, age range, number of events), and performance measures (Table 1). We also extracted whether the validation cohort existed of men and women or women only (studies with men only were previously excluded). When studies used a cohort consisting of both men and women, the model could be validated on men and women together or separately. When validated in men and women separately we only included the validation on women.

Descriptive analyses

Results are presented as counts or percentages where indicated. Combined summary measures of studies and models (e.g. C statistics and number of participants in a cohort) are presented as medians and/or ranges. Proportions were compared with the Chi-square test. C statistics of the most frequently validated models were pooled with the R package metamisc. [17] We estimated random-effect models using restricted maximum likelihood estimation, and derived approximate 95% prediction intervals using the methods described in metamisc. [17] Analyses were performed using SPSS 24 (IBM, Armonk, New York) or R (R Core Team, Vienna, Austria).

Results

Figure 1 depicts the study flow diagram. From the study by Damen and colleagues, 249 articles were included that described models developed for women. The updated search, after removing duplicates, resulted in 9348 new references. After title and abstract screening, 2290 articles were eligible for full text assessment. Full text screening resulted in 244 included articles from the updated search and two additional references identified through snowballing. These 246 papers were added to the 249 papers from Damen et al. and in total, this review includes 495 papers on models for women (Figure 1). In 133 papers prediction models for women were developed. In 206 papers a model was validated and 260 papers concerned incremental value studies. Since papers can develop a model, validate a model and calculate the incremental value of a predictor on an existing model in the same paper, these numbers do not add up to the total of 495 papers.

Development of new prediction models

In 133 distinct papers, 285 cardiovascular risk prediction models were developed. Of these, 160 (56%) were developed solely on women and are henceforth denoted as female-specific models. The remaining 125 (44%) were sex-predictor models (Table 2). Table 2 shows the year in which the models were published. Clearly, new models are still being developed in large numbers, with the majority of the models developed in the last decade (on average 16 new models developed each year). Before 1990, 62% of the developed models were sex-predictor models. Between 1991 and 2010 female-specific models were developed more often than sex-predictor models, since 2010 these proportions are equally divided.

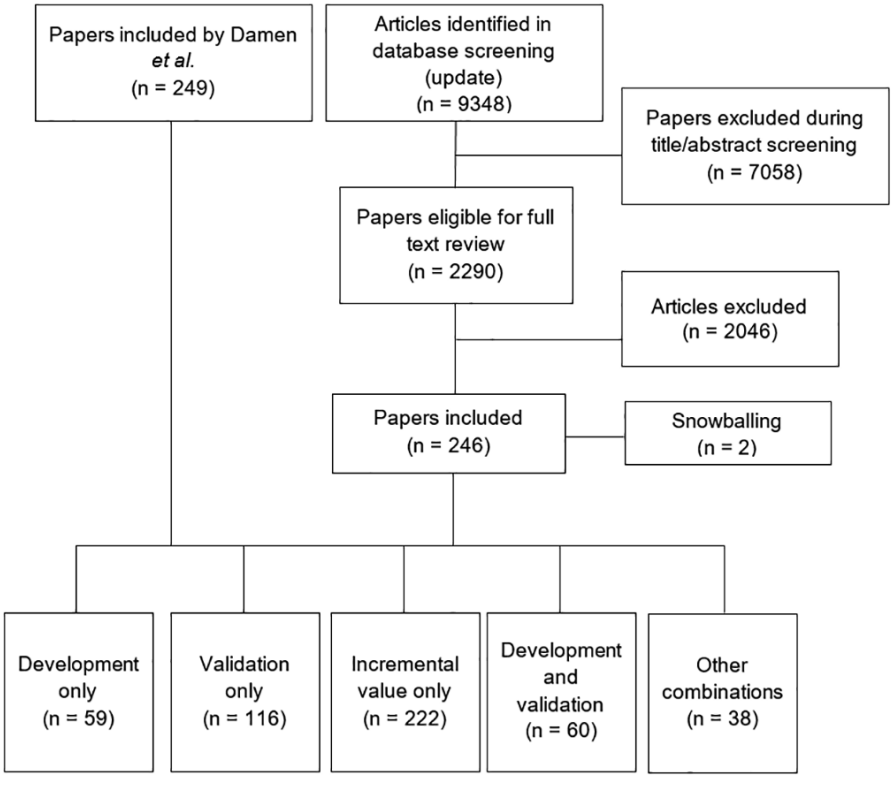


Figure 1: Study flow diagram. The papers that were identified by the updated search were added to the papers from the study by Damen and colleagues, resulting in a total of 495 papers.

Table 2: Number of developed models over time.

Year	1967 – 1990	1991 - 2000	2001 – 2010	2011 - 2017	Total
Sex predictor	21 (62%)	21 (35%)	28 (35%)	55 (50%)	125 (44%)
Female specific	13 (38%)	39 (65%)	52 (65%)	56 (50%)	160 (56%)
Total	34 (100%)	60 (100%)	80 (100%)	111 (100%)	285 (100%)

Predictors in the development papers

For the models that were specifically developed for women, it was of particular interest whether female-specific predictors were included in the model. Only

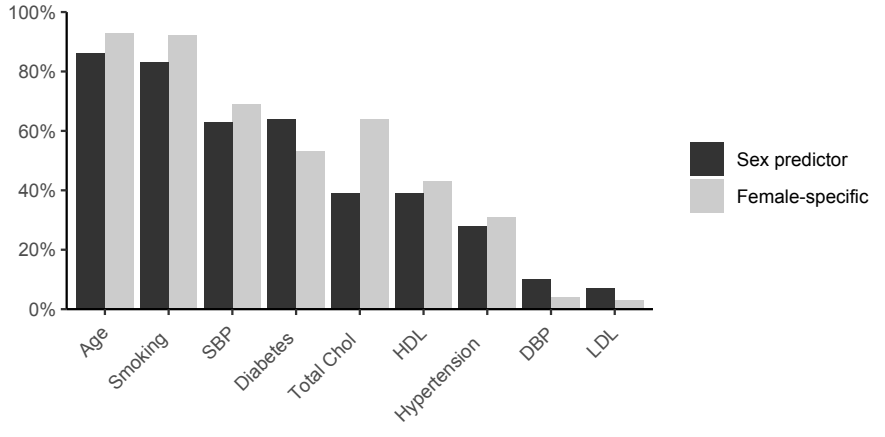


Figure 2: Most frequently used predictors for the sex predictor and female-specific models. HDL; High-density lipoprotein. Total Chol; total cholesterol. LDL; Low-density lipoprotein. SBP; systolic blood pressure. DBP; Diastolic blood pressure.

2 out of the 160 developed female-specific models (1.3%) included a female-specific risk factor. In the first, D'Agostino and colleagues developed a model including menopause (yes/no) and an interaction with menopause and age as predictors. [18] In the second, Parikh and colleagues considered the predictors pregnancy status, number of live births, age at menarche, menstrual irregularity, age at first birth, stillbirths, miscarriages, infertility $\geq$1 year, infertility cause and breastfeeding for inclusion in a model with established risk factors. The final model presented included in addition to age the female-specific risk factors: menstrual irregularity, age at first birth, still births, miscarriages and breast-feeding and had a C statistic of 0.675 in the derivation cohort. [19] The median number of predictors for the female-specific models was 6 [IQR: 5 - 8] and for the sex-predictor models was 8 [IQR: 7 - 10], including the predictor for sex. Figure 2 shows the percentage of sex-predictor and female-specific models that included the nine most often-used predictors. By definition sex was not a predictor in

any of the female-specific models. Total cholesterol was used more frequently in female-specific models (58% vs. 36%, difference 22% 95%CI 10%-33%). For the remaining eight predictors most frequently identified in the models (age, smoking, diabetes mellitus, systolic blood pressure, HDL, hypertension, diastolic blood pressure, and LDL), the frequency of predictors used was similar for the both model types. The apparent C statistic (i.e. the C statistic in the development models) was reported in 66 (53%) of the sex-predictor models and in 59 (37%) of the female-specific models. The median of the C statistics were similar (0.797 for the sex-predictor models [range: 0.610 - 1.000] and 0.787 for the female-specific models [range: 0.660 - 0.918]). The full list of identified development papers in the updated search is available as S1 Table.

Validation of prediction models

A total of 206 articles described 591 validations of sex-predictor or female-specific models. The models that were validated more than five times and at least once in a separate paper, were; SCORE Conroy 2003 (n=63), Framingham Wilson 1998 (n=61 validations), Pooled Cohort Equations Goff 2013 (n=52), Framingham D'Agostino 2008 (n=48), Framingham Anderson 1991a (n=40), Framingham ATP III 2002 (n=29), Framingham Wolf 1991 (n=20), Framingham Anderson 1991b (n=14), and QRISK Hippisley-Cox 2007 (n=6) (Table 3). The 333 validations of these nine models will be discussed further. The only model that is a sex-predictor model is Framingham Anderson 1991a, which was validated 15 (37%) times in men and women and 25 (63%) times in women only. The eight female-specific models were validated 119 (41%) times in men and women together. The other 174 validations (59%) were performed in women only. A C statistic was reported in 70% of these validation studies and ranged from 0.449 to 0.993. Pooled C statistics showed similar performances in validations

performed on women only and validations on men and women together (Table 4). The full list of validated models identified in the updated search is available as S2 Table.

Table 3: Characteristics of the validations of the nine most frequently validated prediction models.

	1 SCORE Conroy 2003 n = 63	2 Framingham Wilson 1998 n = 61	3 PCE Goff 2013 n = 52	4 Framingham D'Agostino 2008 n = 48	5 Framingham Anderson 1991 n = 40	6 Framingham ATP III 2002 n = 29	7 Framingham Wolf 1991 n = 20	8 Framingham Anderson 1991 n = 14	9 QRISK Hippisley-Cox 2007 n = 6
Composition of validation cohorts									
Men and Women	26	27	16	28	15	15	6	1	0
Women Separately	37	34	36	20	25	14	14	13	6
Location of the validation cohorts									
Asia	8	7	8	10	1	1	1	1	0
Australia	4	0	1	1	10	1	0	1	0
Europe	43	20	7	22	28	3	9	8	6
North America	8	32	34	13	1	24	10	4	0
Age of the validation cohorts									
Min, median	40	40	40	40	35	45	55	35	35
Max, median	65	74	79	79	74	82	99	64	74
Size of the validation cohorts									
Sample size, median	7573	3554	4218	2613	2105	3716	3507	3014	542987
range	[203-44649]	[246-163627]	[392-307591]	[136-542987]	[302-797373]	[613-36517]	[401-23983]	[331-542783]	[306111-797373]
Events, median	157	213	150	146	86	384	160	158	29057
range	[10-4842]	[8-24659]	[9-4658]	[15-18173]	[1-29057]	[35-2343]	[24-939]	[5-18173]	[18027-29057]

Incremental value

In 260 articles the added value of a predictor to an existing female-specific or sex-predictor model was described. In 3 (1.1%) papers female-specific risk factors were added to an existing model, all of which were recently published (2016 n=2 and 2017 n=1). [19–21] In the previously discussed paper by Parikh and colleagues, female-specific predictors were added to established risk factors, resulting in a final model including age at first birth, still births, miscarriages and breastfeeding. This slightly improved the model, C statistic of 0.730, where the model with only established risk factors had a C statistic of 0.726. [19] In a study by van der Meer and colleagues, the female-specific predictors age at menarche, menopausal status/age, hormone use, gestational hypertension and diabetes, number of children, miscarriages/stillbirths were added to established risk factors. The addition of these predictors did non apparently improve the discrimination or calibration of the model beyond the established risk factors. [20] In the third paper, Zhou and colleagues added amongst other predictors (African American ethnicity, physical exercise level, BMI, waist circumference, height, HDL cholesterol), use of hormone replacement therapy in postmenopausal women to the Framingham Stroke Risk Score (Wolf 1991). The addition of this predictor set improved discrimination and calibration of the model in women; however, the separate performance of hormone use was not reported. [21] The full list of incremental value papers identified by the updated search is available as S3 Table.

Table 4: Pooled C statistics of the validations of the nine most frequently validated prediction models

	1 SCORE Conroy 2003	2 Framingham Wilson 1998	3 PCE Goff 2013	4 Framingham D'Agostino 2008	5 Framingham Anderson 1991	6 Framingham ATP III 2002	7 Framingham Wolf 1991	8 Framingham Anderson 1991	9 QRISK Hippisley-Cox 2007
Validations on men and women									
	n = 15	n = 17	n = 9	n = 21	n = 8	n = 11	n = 3	n = 0	n = 0
Pooled C statistic	0.768	0.717	0.739	0.734	0.673	0.72	0.653	–	–
95% Prediction Interval	(0.709–0.826)	(0.542–0.893)	(0.679-0.799)	(0.600-0.868)	–*	(0.593-0.846)	–*	–	–
Validations on women separately									
	n = 13	n = 18	n = 28	n = 10	n = 13	n = 8	n = 8	n = 2	n = 3
Pooled C statistic	0.772	0.682	0.757	0.730	0.776	0.687	0.678	0.767	0.796
95% Prediction Interval	(0.591–0.954)	(0.491–0.874)	(0.696-0.819)	(0.544-0.916)	(0.755-0.796)	(0.568-0.806)	(0.447-0.908)	–†	(0.750-0.843)

*Due to limited information the resulting prediction interval lies outside the possible interval (values >1 and/or <0)
†Not enough validations were available to calculate the prediction interval

Reliability and clinical usability of available models

All 285 models developed for women were first assessed for reliability and were regarded so if they met the following criteria: 1) model externally validated 2) externally validated in a separate investigation/paper and 3) a C statistic >0.7. Of the 285 models, 40 (14%) met these criteria and were considered reliable (Table 5). Of these 40, 25 (63%) were female-specific and 15 (37%) were sex-predictor models. Following, these models were assessed for clinical usability based on the presence of 1) 10 predictors or fewer, 2) full regression model or chart reported and 3) online calculator available (Table 5). The SCORE and Framingham 2008 model had the highest usability score as they met all criteria. Other models with high usability are the Pooled Cohort Equations (African American), Framingham 30 year and the Framingham stroke models as they have 10 or fewer predictors and an online calculator available. The remaining models either had more than 10 predictors or no calculator available, rendering them less appealing for clinical practice.

Discussion

In this study we provided an overview of the available CVD risk assessment models for women in the general population. We identified a wide range of models that have been developed over the past decades, including 160 female-specific models (i.e. models that are developed for use in women only) and 125 sex-predictor models (i.e. models that include sex as a predictor). Despite this large quantity, only two of the 160 (1.3%) female-specific models included female-specific predictors. [18, 19] Of the 260 studies in which the added value of a predictor was assessed, only three (1.1%) investigated the added value of a female-specific predictor. [19–21]

Table 5: Clinical usability of models that met the reliability criteria.

Model – study name	Author - Year	Number of separate models	< 10 predictors	Full regression formula	Risk Chart	Online Calculator
Framingham	Anderson 1991a	12	✓	✓	x	x
Framingham	Anderson 1991b	2	✓	✓	✓	x
–	Assmann 2007	2	✓	x	✓	x
ARIC	Chambless 2003	2	x	✓	x	✓
SCORE	Conroy 2003	6	✓	✓	✓	✓
Framingham	D'Agostino 2008	2	✓	✓	✓	✓
Framingham	ATP II	1	✓	✓	✓	x
–	Gaziano 2008	2	✓	x	✓	x
Pooled Cohort Equations (African American)	Goff 2013	1	✓	✓	x	✓
Pooled Cohort Equations (White)	Goff 2013	1	x	✓	x	✓
QRISK	Hippisley-Cox 2007	1	✓	x	x	x
QRISK2	Hippisley-Cox 2008	2	x	x	x	x
QRISK lifetime	Hippisley-Cox 2010	1	x	x	x	✓
–	Lumley 2002	1	✓	x	✓	x
Framingham (30 yrs)	Pencina 2009	1	✓	x	x	✓
–	Schnabel 2009	1	✓	x	✓	x
Framingham	Wilson 1998	1	✓	✓	✓	x
Framingham (Stroke)	Wolf 1991	1	✓	x	✓	✓

Clinical usability was scored for the models which met all criteria for reliability: 1) model externally validated 2) externally validated in a separate investigation/paper and 3) a C statistic >0.7.

Our study has several major strengths. We performed an extensive search up to July 2017 and systematically selected studies for inclusion. Detailed and thorough data extraction of essential information such as type of models, predictors, population and model discrimination, was performed by means of standardized forms and was done by three investigators together for the development models to ensure uniformity. Limitations of our study should be mentioned. First, we did not include models specifically made for men and thus could not compare differences in performance and predictors between men and women. Second, in some validation studies it was not clear which models were validated when the original development article reported on more than one model. We assumed that all models in the article were validated, but this may have led to an overestimation of the actual number of times prediction models were validated. Third, we did not include articles written in a language different than English and articles of which the full text could not be retrieved. Furthermore, validation papers were excluded from the pooled C statistic analyses when insufficient information necessary for pooling was reported. In addition, since we did not conduct a formal risk of bias assessment, we were only able to include all validation studies in which reporting was complete, instead of including for example studies with the smallest risk of bias. Therefore, results on the pooled C statistics, should be interpreted with caution. Finally, as calibration was reported in a heterogeneous manner, conclusions for this performance measure could not be drawn. Furthermore, in papers the measure for calibration was often not reported. In order to guarantee uniformity, new studies reporting on prediction models should adhere to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement. [22, 23]

The models described in this review often comprise several variations of

established, sex-independent predictors such as age, blood pressure, lipid levels and smoking indicating that these predictors attribute most to the current performance of the models. Interestingly, the results showed that both the female-specific models as well as the sex-predictor models often comprise these same established predictors and do not differ substantially in estimated C statistic. This might imply that using sex as predictor in a model is just as effective as developing a female-specific model. Of the nine most frequently validated models in women the C statistic as a measure of performance was reported in 59% of the validation studies. Pooled C statistics indicated good performance in general (pooled C statistic >0.70 for most models), although the range of reported C statistics varied from 0.45 to 0.99. This indicates that although these models generally perform well, they can definitely be improved. Of all 285 developed models, only 40 (14%) met the quality criteria for reliability. When these models were further assessed for clinical usability only 2/40 (5%), the SCORE and Framingham 2008 model, met all criteria. Other models which met most criteria and had a risk calculator available were the Pooled cohort equations, Framingham 30 years and Framingham stroke model. Based on both these reliability and clinical usability criteria, these models seem best suitable for implementation in clinical practice. Models without an online calculator are likely less attractive for use in clinical practice.

Our findings are in line with a previous study by Goh and colleagues, in which the utility of CVD prediction models for women was appraised. [24] They also concluded that there is room for improvement in CVD prediction models for women and this could be achieved by adding predictors which may perform well in women. Remarkable is that none of the predictors suggested by Goh, such as obesity, physical activity and coronary artery calcium, are female-specific. It must be noted that in the study by Goh and colleagues the search was

limited to five years before publication (2008-2013). The study was restricted to six models, where we in our study considered any model identified by the search strategy. The 2011 guidelines for the prevention of CVD in women [25] categorize women as 'at risk' when having one or more major risk factors. Aside from the established risk factors found in most prediction models, they explicitly include the female-specific risk factors of a history of preeclampsia, gestational diabetes, or pregnancy-induced hypertension. However, none of these disorders are used in any of the prediction models for women in this review.

Although many models have been developed in women only, it seems that differences between men and women in CVD risk assessment are still not fully recognized. Many female specific risk factors for CVD have been identified in recent years, but their predictive potential has not been tested or even considered in risk prediction models within the scope of our review. Our search only identified two development studies that included a female-specific predictor in the model. [18, 19] Improvement of the existing models might be achieved in adding female-specific predictors. However, in most of the incremental value studies we found, female-specific predictors were not even considered as potential predictors for added value. Of the 260 incremental value studies, three added a female-specific predictor. Of these, one reported no improvement in performance and one observed a slight improvement in discrimination. The third did not report on improvement of individual predictors. A reason for not finding any substantial improvement could be that studies missed information on several important female-specific risk factors like preeclampsia, polycystic ovary syndrome and infant birth weight. Therefore, it is important to further investigate the potential added value of female-specific predictors. Most female-specific predictors become apparent at an early stage in life whereas CVD events mostly occur after the age of 50. An additional benefit is that these

predictors can be easily obtained from the medical history. This underlines the potential of these predictors, as risk assessment is ideally performed decades before the anticipated event, in order to implement and optimize effect of preventive strategies. Although we identified a total of 495 papers on CVD prediction models for women, it is still uncertain whether these can be improved by female-specific predictors. However, it should be mentioned that finding new predictors that improve model performance on top of the well-known predictors seems challenging. [26] It is possible that current models, which often aim to estimate the 10-year risk based on a single assessment, have reached their maximal predictive potential and cannot be further improved. A new type of model, for example the dynamic model, in which an individual's risk is continuously updated over time, could further advance preventive strategies.

Conclusions

In conclusion, there is an abundance of models for women in the general population, but female-specific predictors are rarely included. The few studies that add female-specific risk factors to existing CVD risk models do not show substantial improved performance, but lacked important potential predictors. Further research in order to provide physicians with a well-performing and properly validated prediction model for women is therefore warranted, considering all female-specific predictors. Ideally their added value to models which already perform well is assessed instead of developing completely new models. [12]

References

1. Healy, B. The Yentl syndrome. *N Engl J Med* **325,** 274–6 (1991).
2. Dean, J., Cruz, S. D., Mehta, P. K. & Merz, C. N. Coronary microvascular dysfunction: sex-specific risk, diagnosis, and therapy. *Nat Rev Cardiol* **12,** 406–14. ISSN: 1759-5010 (Electronic) 1759-5002 (Linking) (2015).

3. Gillis, A. M. Atrial Fibrillation and Ventricular Arrhythmias: Sex Differences in Electrophysiology, Epidemiology, Clinical Presentation, and Clinical Outcomes. *Circulation* **135,** 593–608 (2017).
4. Grodecki, K. *et al.* Gender-related differences in post-discharge bleeding among patients with acute coronary syndrome on dual antiplatelet therapy: A BleeMACS sub-study. *Thromb Res* **168,** 156–163 (2018).
5. Regitz-Zagrosek, V. & Kararigas, G. Mechanistic Pathways of Sex Differences in Cardiovascular Disease. *Physiol Rev* **97,** 1–37 (2017).
6. Ventura-Clapier, R. *et al.* Sex in basic research: concepts in the cardiovascular field. *Cardiovasc Res* **113,** 711–724 (2017).
7. Mosca, L. *et al.* National study of physician awareness and adherence to cardiovascular disease prevention guidelines. *Circulation* **111,** 499–510 (2005).
8. Baggio, G., Corsini, A., Floreani, A., Giannini, S. & Zagonel, V. Gender medicine: a task for the third millennium. *Clin Chem Lab Med* **51,** 713–27 (2013).
9. Appelman, Y., van Rijn, B. B., Ten Haaf, M. E., Boersma, E. & Peters, S. A. Sex differences in cardiovascular risk factors and disease prevention. *Atherosclerosis* **241,** 211–8 (2015).
10. Tan, Y. Y., Gast, G. C. & van der Schouw, Y. T. Gender differences in risk factors for coronary heart disease. *Maturitas* **65,** 149–60 (2010).
11. Zoet, G. A. *et al.* Prevalence of Subclinical Coronary Artery Disease Assessed by Coronary Computed Tomography Angiography in 45- to 55-Year-Old Women With a History of Preeclampsia. *Circulation* **137,** 877–879 (2018).
12. Damen, J. A. *et al.* Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* **353,** i2416 (2016).
13. Goff D. C., J. *et al.* 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol* **63,** 2935–59 (2014).
14. Piepoli, M. F. *et al.* 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts)Developed with the special contribution of the European Association for Cardiovascular Prevention and Rehabilitation (EACPR). *Eur Heart J* **37,** 2315–2381 (2016).
15. Beswick, A. D., Brindle, P., Fahey, T. & Ebrahim, S. A Systematic Review of Risk Scoring Methods and Clinical Decision Aids Used in the Primary Prevention of Coronary Heart Disease, 1–361 (2008).
16. Moons, K. G. *et al.* Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med* **11,** e1001744 (2014).
17. Debray, T. P. *et al.* A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. *Stat Methods Med Res,* 962280218785504 (2018).

18. D'Agostino, R. B. *et al.* Primary and subsequent coronary risk appraisal: new results from the Framingham study. *Am Heart J* **139,** 272–81 (2000).

19. Parikh, N. I. *et al.* Reproductive Risk Factors and Coronary Heart Disease in the Women's Health Initiative Observational Study. *Circulation* **133,** 2149–58 (2016).

20. Van der Meer, M. G. *et al.* Added Value of Female-Specific Factors Beyond Traditional Predictors for Future Cardiovascular Disease. *J Am Coll Cardiol* **67,** 2084–6 (2016).

21. Zhou, X. H., Wang, X., Duncan, A., Hu, G. & Zheng, J. Statistical evaluation of adding multiple risk factors improves Framingham stroke risk score. *BMC Med Res Methodol* **17,** 58 (2017).

22. Collins, G. S., Reitsma, J. B., Altman, D. G., Moons, K. G. M. & members of the, T. g. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *Eur Urol* **67,** 1142–1151 (2015).

23. Moons, K. G. *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* **162,** W1–73 (2015).

24. Goh, L. G., Dhaliwal, S. S., Lee, A. H., Bertolatti, D. & Della, P. R. Utility of established cardiovascular disease risk score models for the 10-year prediction of disease outcomes in women. *Expert Rev Cardiovasc Ther* **11,** 425–35 (2013).

25. Mosca, L. *et al.* Effectiveness-based guidelines for the prevention of cardiovascular disease in women–2011 update: a guideline from the American Heart Association. *J Am Coll Cardiol* **57,** 1404–23 (2011).

26. Wang, T. J. *et al.* Multiple biomarkers for the prediction of first major cardiovascular events and death. *N Engl J Med* **355,** 2631–9 (2006).

27. Mosca, L. *et al.* Effectiveness-based guidelines for the prevention of cardiovascular disease in women–2011 update: a guideline from the american heart association. *Circulation* **123,** 1243–62 (2011).

Supplemental Material

S1. The CREW Consortium

The CREW consortium consists of (in alphabetical order):

Yolande E Appelman[1], Sara J Baart[2,3], Laura Benschop[2,3], Eric Boersma[2], Laura Brouwers[3,4], Ricardo P J Budde[2], Suzanne C Cannegieter[5], Veerle Dam[3,6], Rene M J C Eijkemans[6], Bart C J M Fauser[4], Michel D Ferrari[5], Arie Franx[4], Christianne J M de Groot[1], Marlise N Gunning[3,4], Annemiek Hoek[7], Hendrik Koffijberg[6,8], Maria P H Koster[2], Mark C Kruit[5], Ghizelda R Lagerwij[3,6], Cornelis B Lambalk[1], Joop S E Laven[2], Katie M Linstra[2,3,5], Aad van der Lugt[2], Angela H E M Maas[9], Antoinette Maassen van den Brink[2], Cindy Meun[2,3], Saskia Middeldorp[10], Karel G M Moons[6], Bas B van Rijn[4], Jeanine E Roeters van Lennep[2], Jolien W Roos-Hesselink[2], Luuk J J Scheres[3,5,10], Yvonne T van der Schouw[6], Eric A P Steegers[2], Regine P M Steegers-Theunissen[2], Gisela M Terwindt[5], Birgitta K Velthuis[4], Marieke J H Wermer[5], Gerbrand A Zoet[3,4]

[1]VU University Medical Center, Amsterdam, the Netherlands

[2]Erasmus University Medical Center, Rotterdam, the Netherlands

[3]Netherlands Heart Institute, Utrecht, the Netherlands

[4]University Medical Center Utrecht, Utrecht, the Netherlands

[5]Leiden University Medical Center, Leiden, the Netherlands

[6]Julius Center, University Medical Center, Utrecht, the Netherlands

[7]University Medical Center Groningen, Groningen, the Netherlands

[8]University of Twente, Enschede, the Netherlands

[9]Radboud University Medical Center, Nijmegen, the Netherlands

[10]Academic Medical Center, Amsterdam, the Netherlands

S2. Supplemental Tables

Supplemental Table 1. Articles that developed a new model in the updated search and their external validation

First author, publication year	Number of models developed	Female-specific or sex-predictor	Number of articles in which model is validated
Artigao-Rodenas, 2015	1	Sex-predictor	1 (Artigao-Rodenas, 2015)
Backholer, 2017	3	Sex-predictor	1 (Backholer, 2017)
Bali, 2016	1	Sex-predictor	1 (Bali, 2016)
Borglykke, 2010	5	Female-specific	1 (Borglykke, 2010)
Chahal, 2015	7	Sex-predictor	-
Chiuve, 2014	1	Female-specific	-
Cooney, 2012	1	Female-specific	1 (Cooney, 2012)
Cross, 2013	1	Sex-predictor	-
Deo, 2016	1	Sex-predictor	1 (Deo, 2016)
Dhoble, 2014	4	Sex-predictor	-
Fox, 2016	6	Sex-predictor	1 (Fox, 2016)
Goff, 2013	2 (PCE)	Female-specific	20 (Muntner, 2014; Lee, 2015; Khalili, 2015; Kavousi, 2014; Jung, 2015; DeFilippis, 2015; Chia, 2014; Andersson, 2015; Yang, 2016; Cook, 2014; De Las Heras Gala, 2016; DeFilippis, 2017; Emdin, 2017; Foraker, 2016; Goff, 2013; Karmali, 2015; Mortensen, 2017; Qureshi, 2016; Rana, 2016; Zhang, 2017)
Hajifathalian, 2015	2	Female-specific	1 (Hajifathalian, 2015)
Hensley, 1998	1	Female-specific	-
Hippisley-Cox, 2013	1 (QSTROKE)	Female-specific	2 (Hippisley-Cox, 2014; Parmar, 2015)
Hippisley-Cox, 2017	3	Female-specific	1 (Hippisley-Cox, 2017)
Ho, 2016	2	Sex-predictor	1 (Ho, 2016)
Howard, 2017	1	Sex-predictor	1 (Howard, 2017)

Hu, 2014	1	Sex-predictor	-
Jairam, 2015	1	Sex-predictor	
Jee, 2014	4	Female-specific	-
Johansson	1	Sex-predictor	
Jung, 2015	1	Female-specific	-
Kovalchik, 2013	1	Female-specific	1 (Kovalchik, 2013)
Kusmana, 2002	1	Sex-predictor	-
Liu, 2016	1	Female-specific	
Manuel, 2015	1	Female-specific	
Marino, 2014	2	Female-specific	-
Marrugat, 2014	6	Female-specific	1 (Marrugat, 2014)
McClelland, 2015	2	Sex-predictor	1 (McClelland, 2015)
McNeil, 2001	2	Female-specific	-
Nishimura, 2014	4	Sex-predictor	-
Nobel, 2014	1	Sex-predictor	-
Onat, 2017	1	Female-specific	
Parikh, 2016	1	Female-specific	-
Parmar, 2014	1	Sex-predictor	1 (Parmar, 2014)
Paynter, 2014	3	Female-specific	-
Piotrowski, 2016	2	Female-specific	
Selmer, 2017	1	Female-specific	
Stam-Slob, 2017	1	Sex-predictor	
Vartiainen, 2016	3	Female-specific	-
Wang, 2016	1	Female-specific	
Wickramasinghe, 2014	1	Female-specific	-
Woodward, 2007	1	Female-specific	
Woodward, 2006	2	Female-specific	-
Würtz, 2015	2	Sex-predictor	1 (Würtz, 2015)
Yang, 2016	1	Female-specific	-
Yatsuya, 2016	2	Sex-predictor	-
Yudkin, 1999	1	Female-specific	-

Supplemental Table 2. Models validated in the update

Model Validated	Author, year model developed	Number of articles in which model is validated
Framingham	Anderson - 1991	2 (Goh, 2014; Tilin, 2014)
Framingham	ATP III - 2002	5 (DeFilippis, 2015; Dhoble, 2014; Hu, 2014; Qureshi, 2016; Kavousi, 2014)
YDR	Colditz - 2000	1 (De Vito, 2015)
SCORE	Conroy - 2003	15 (Goh, 2014; Jdanov, 2014; Jorstad 2014; Kavousi, 2014; Mortsensen, 2015; Selvarajah, 2014; Vikhireva, 2014-a; Vikhireva, 2014-b, Baena-Diez, 2017; Mortensen, 2017; De Las Heras Gala, 2016; Qureshi, 2016; Berard, 2016; Sawano, 2016; Piotrowski, 2016)
SSVMod	Counsell - 2002	1 (Sim, 2016)
Framingham	D'Agostino - 2008	9 (Artigao-Rodenas, 2013; Chia, 2015; DeFilipis, 2015; Marino, 2014; Selvarajah, 2014, Fatema, 2016, Qureshi, 2016; Chamnan, 2016, Sepanlou, 2015)
CHADS2	Gage - 2001	1 (Yuan, 2017)
QRISK2	Hippisley-Cox - 2008	2 (Hippisley-Cox, 2014; Tilin, 2014)
CBC Score	Horne - 2009	1 (Horne, 2015)
SCORE – Germany	Keil – 2005	1 (Rucker, 2016)
CHA2DS2 - VASC	Lip - 2010	1 (Yuan, 2017)
Framingham - Regicor	Marrugat - 2003	1 (Marrugat, 2014)
HellenicSCORE	Panagiotakos - 2007	1 (Panagiotakos, 2015)
Framingham	Pencina - 2009	1 (van Kempen, 2014)
Reynolds Risk	Ridker - 2007	1 (DeFilippis, 2015)
Dubbo	Simons - 2003	1 (Weatherley, 2011)
SCORE-NL	Van Dis - 2010	1 (Van Dis, 2014)
NIHSSMod	Weimar – 2004	1 (Sim, 2016)
WHO/ISH	WHO - 2007	2 (Raghu, 2015; Selvarajah, 2014)
SCORE – Sweden	Wilhelmsen - 2004	1 (Karjalainen, 2017)
Framingham	Wilson - 1998	4 (DeFilippis, 2015; Hu, 2914; Nishimura, 2014; Fowkes, 2014)
Framingham	Wolf - 1991	6 (Hippisley-Cox, 2013; McClure, 2014; Parmar, 2014; Sabayan, 2013; Dufouil, 2017; Howard, 2017)

Supplemental Table 3. Models used for incremental value in the update.

Model used for incremental value	Author, year model developed	Number of articles used for IV
ARIC HF	Agarwal - 2012	1 (Nambi, 2013)
Framingham	Anderson - 1991	1 (Wassertheil-Smel, 2014)
Framingham	ATP III - 2002	2 (Hadamitzky, 2013; Valentini, 2015)
SCORE	Conroy - 2003	7 (Faeh, 2013; Ferrario, 2014; Groot, 2015; Schnohr, 2015; Sehestedt, 2011; Vikhireva, 2014; Woznicka-Leskiew, 2015)
Framingham	Cupples - 1988	1 (Lluis-Ganella, 2012)
Framingham	D'Agostino - 1994	2 (Gibson, 2014; Ziegelbauer)
Framingham	D'Agostino - 2000	1 (Aljaroudi, 2013)
Framingham	D'Agostino - 2001	1 (Yeboah, 2014)
Framingham	D'Agostino - 2008	5 (Armstrong, 2014; Criqui, 2013; Goh, 2014; Kunutsor, 2015; Lopez-Suarez, 2014)
AGLA	Eckardstein - 2012	1 (Romanens, 2014)
–	Ferrario -2005	1 (Veronesi, 2014)
Framingham	Unspecified	8 (Badheka, 2013-a; Badheka, 2013-b; Brouwers, 2014; Gaibazzi, 2014; Lindberg, 2014; Okwuosa, 2014; Willeit, 2014; Woznicka-Leskiew, 2015)
Pooled Cohort Equations	Goff - 2013	3 (Everett, 2015; Kim, 2014; Okwuosa, 2014)
QRISK2	Hippisley-Cox - 2008	1 (Weng, 2015)
REGICOR	Marrugat - 2003	2 (Velescu, 2015; Llius-Ganella, 2012)
Laboratory Report Model	Nambi - 2013	Nambi - 2013
HellenicSCORE	Panagiotakos - 2007	1 (Georgousopoulou, 2015)
Reynolds Risk	Ridker - 20017	4 (Everett, 2015; Everett, 2014; Kim, 2014; Shah, 2014)
NL-SCORE	Smulders - 2008	1 (Van Dis, 2012)
Traditional Risk Factors	-	7 (Baber, 2015; Candell-Riera, 2013; Funke-Kaiser, 2014; Gardin, 2014; Kunutsor, 2014; Nielson, 2014; Nimomiya, 2013)
Framingham	Wilson - 1998	11 (Bérard, 2013; Britton, 2013; Fowkes, 2014; Gronewold, 2014; Kalsch, 2014; Lyngbaek, 2012; Mahabadi, 2015; Polak, 2015; Valentini, 2015; Weng, 2015; Zalawadiya, 2015)

S3. Full list of included papers from the update

1. Abbasi, A. *et al.* Plasma N-terminal Prosomatostatin and Risk of Incident Cardiovascular Disease and All-Cause Mortality in a Prospective Observational Cohort: the PREVEND Study. *Clin Chem* **63,** 278–287 (2017).
2. Abraham, G. *et al.* Genomic prediction of coronary heart disease. *Eur Heart J* **37,** 3267–3278 (2016).
3. Aijala, M. *et al.* The fat mass and obesity-associated (FTO) gene variant rs9939609 predicts long-term incidence of cardiovascular disease and related death independent of the traditional risk factors. *Ann Med* **47,** 655–63 (2015).
4. AlJaroudi, W. A. *et al.* Incremental prognostic value of diastolic dysfunction in low risk patients undergoing echocardiography: beyond Framingham score. *International Journal of Cardiovascular Imaging* **29,** 1441–1450 (2013).
5. Amato, M. *et al.* Carotid plaque-thickness and common carotid IMT show additive value in cardiovascular risk prediction and reclassification. *Atherosclerosis* **263,** 412–419 (2017).
6. Ambale-Venkatesh, B. *et al.* Left ventricular shape predicts different types of cardiovascular events in the general population. *Heart* **103,** 507–515 (2017).
7. Andersson, C., Enserro, D., Larson, M. G., Xanthakis, V. & Vasan, R. S. Implications of the US cholesterol guidelines on eligibility for statin therapy in the community: comparison of observed and predicted risks in the Framingham Heart Study Offspring Cohort. *J Am Heart Assoc* **4(4)** (2015).
8. Armstrong, A. C. *et al.* Left atrial dimension and traditional cardiovascular risk factors predict 20-year clinical cardiovascular events in young healthy adults: the CARDIA study. *European heart journal cardiovascular Imaging* **15,** 893–9 (2014).
9. Artigao-Rodenas, L. M. *et al.* Framingham risk score for prediction of cardiovascular diseases: a population-based study from southern Europe. *PLoS ONE* **8,** e73529 (2013).
10. Artigao-Rodenas, L. M. *et al.* Construction and Validation of a 14-Year Cardiovascular Risk Score for Use in the General Population: The Puras-GEVA Chart. *Medicine (Baltimore)* **94,** e1980 (2015).
11. Astor, B. C. *et al.* Novel Markers of Kidney Function as Predictors of ESRD, Cardiovascular Disease, and Mortality in the General Population. *American Journal of Kidney Diseases* **59,** 653–662 (2012).
12. Baber, U. *et al.* Prevalence, impact, and predictive value of detecting subclinical coronary and carotid atherosclerosis in asymptomatic adults: the BioImage study. *J Am Coll Cardiol* **65,** 1065–74 (2015).
13. Backholer, K. *et al.* Development of an Australian cardiovascular disease mortality risk score using multiple imputation and recalibration from national statistics. *BMC Cardiovasc Disord* **17,** 17 (2017).
14. Badheka, A. O. *et al.* ST-T wave abnormality in lead aVR and reclassification of cardiovascular risk (from the National Health and Nutrition Examination Survey-III). *Am J Cardiol* **112,** 805–10 (2013).

15. Badheka, A. O. *et al.* QRS duration on electrocardiography and cardiovascular mortality (from the National Health and Nutrition Examination Survey-III). *Am J Cardiol* **112,** 671–7 (2013).

16. Baena-Diez, J. M. *et al.* Validity Assessment of Low-risk SCORE Function and SCORE Function Calibrated to the Spanish Population in the FRESCO Cohorts. *Rev Esp Cardiol (Engl Ed)* (2017).

17. Bali, V., Yermilov, I., Coutts, K. & Legorreta, A. P. Novel screening metric for the identification of at-risk peripheral artery disease patients using administrative claims data. *Vasc Med* **21,** 33–40 (2016).

18. Barr, E. L. *et al.* Cystatin C estimated glomerular filtration rate and all-cause and cardiovascular disease mortality risk in the general population: AusDiab study. *Nephrology (Carlton)* **22,** 243–250 (2017).

19. Bellinazzi, V. R. *et al.* Carotid flow velocity/diameter ratio is a predictor of cardiovascular events in hypertensive patients. *J Hypertens* **33,** 2054–60 (2015).

20. Berard, E., Bongard, V., Ruidavets, J. B., Amar, J. & Ferrieres, J. Pulse wave velocity, pulse pressure and number of carotid or femoral plaques improve prediction of cardiovascular death in a population at low risk. *Journal of Human Hypertension* **27,** 529–534 (2013).

21. Berard, E. *et al.* Predictive Accuracy of the European Society of Cardiology SCORE Among French People. *Journal of Cardiopulmonary Rehabilitation and Prevention* **36,** 38–48 (2016).

22. Biering-Sorensen, T. *et al.* Global Longitudinal Strain by Echocardiography Predicts Long-Term Risk of Cardiovascular Morbidity and Mortality in a Low-Risk General Population: The Copenhagen City Heart Study. *Circ Cardiovasc Imaging* **10** (2017).

23. Biering-Sorensen, T., Mogelvang, R., Schnohr, P. & Jensen, J. S. Cardiac Time Intervals Measured by Tissue Doppler Imaging M-mode: Association With Hypertension, Left Ventricular Geometry, and Future Ischemic Cardiovascular Diseases. *J Am Heart Assoc* **5** (2016).

24. Blaha, M. J. *et al.* Improving the CAC Score by Addition of Regional Measures of Calcium Distribution: Multi-Ethnic Study of Atherosclerosis. *JACC Cardiovasc Imaging* **9,** 1407–1416 (2016).

25. Blankenberg, S. *et al.* Troponin I and cardiovascular risk prediction in the general population: the BiomarCaRE consortium. *European Heart Journal* **37,** 2428–+ (2016).

26. Borglykke, A. *et al.* Stroke risk estimation across nine European countries in the MORGAM project. *Heart* **96,** 1997–2004 (2010).

27. Bos, D., Ikram, M. A., Leening, M. J. G. & Ikram, M. K. The Revised Framingham Stroke Risk Profile in a Primary Prevention Population: The Rotterdam Study. *Circulation* **135,** 2207–2209 (2017).

28. Bozorgmanesh, M., Sardarinia, M., Hajsheikholeslami, F., Azizi, F. & Hadaegh, F. CVD-predictive performances of "a body shape index" versus simple anthropometric measures: Tehran lipid and glucose study. *European Journal of Nutrition* **55,** 147–157 (2016).

29. Britton, K. A. *et al.* Body fat distribution, incident cardiovascular disease, cancer, and all-cause mortality. *J Am Coll Cardiol* **62,** 921–5 (2013).

30. Brouwers, F. P. *et al.* Elevated urinary albumin excretion complements the Framingham Risk Score for the prediction of cardiovascular risk - response to treatment in the PREVEND IT trial. *IJC Heart and Vessels* **4,** 193–197 (2014).

31. Brownrigg, J. R. *et al.* Microvascular disease and risk of cardiovascular events among individuals with type 2 diabetes: a population-level cohort study. *Lancet Diabetes Endocrinol* **4,** 588–97 (2016).

32. Brutsaert, E. F. *et al.* Relations of Postload and Fasting Glucose With Incident Cardiovascular Disease and Mortality Late in Life: The Cardiovascular Health Study. *J Gerontol A Biol Sci Med Sci* **71,** 370–7 (2016).

33. Bye, A. *et al.* Circulating microRNAs predict future fatal myocardial infarction in healthy individuals - The HUNT study. *J Mol Cell Cardiol* **97,** 162–8 (2016).

34. Candell-Riera, J. *et al.* Usefulness of exercise test and myocardial perfusion-gated single photon emission computed tomography to improve the prediction of major events. *Circ Cardiovasc Imaging* **6,** 531–41 (2013).

35. Chahal, H. *et al.* Heart failure risk prediction in the Multi-Ethnic Study of Atherosclerosis. *Heart* **101,** 58–64 (2015).

36. Chambless, L. E. *et al.* Coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC) study. *J Clin Epidemiol* **56,** 880–90 (2003).

37. Chamnan, P. *et al.* Repeat Cardiovascular Risk Assessment after Four Years: Is There Improvement in Risk Prediction? *PLoS One* **11,** e0147417 (2016).

38. Chang, X. *et al.* Utility of genetic and non-genetic risk factors in predicting coronary heart disease in Singaporean Chinese. *Eur J Prev Cardiol* **24,** 153–160 (2017).

39. Chia, Y. C., Gray, S. Y. W., Ching, S. M., Lim, H. M. & Chinna, K. Validation of the Framingham general cardiovascular risk score in a multiethnic Asian population: a retrospective cohort study. *Bmj Open* **5** (2015).

40. Chia, Y. C., Lim, H. M. & Ching, S. M. Validation of the pooled cohort risk score in an Asian population - a retrospective cohort study. *Bmc Cardiovascular Disorders* **14** (2014).

41. Chiuve, S. E. *et al.* Lifestyle-based prediction model for the prevention of CVD: the Healthy Heart Score. *J Am Heart Assoc* **3,** e000954 (2014).

42. Cho, I. *et al.* Incremental prognostic utility of coronary CT angiography for asymptomatic patients based upon extent and severity of coronary artery calcium: results from the COronary CT Angiography EvaluatioN For Clinical Outcomes InteRnational Multicenter (CONFIRM) study. *Eur Heart J* **36,** 501–8 (2015).

43. Christophersen, I. E. *et al.* A comparison of the CHARGE-AF and the CHA2DS2-VASc risk scores for prediction of atrial fibrillation in the Framingham Heart Study. *American Heart Journal* **178,** 45–54 (2016).

44. Colantonio, L. D. *et al.* Performance of the Atherosclerotic Cardiovascular Disease Pooled Cohort Risk Equations by Social Deprivation Status. *J Am Heart Assoc* **6** (2017).

45. Cook, N. R. & Ridker, P. M. Further Insight Into the Cardiovascular Risk Calculator The Roles of Statins, Revascularizations, and Underascertainment in the Women's Health Study. *Jama Internal Medicine* **174,** 1964–1971 (2014).

46. Cooney, M. T. *et al.* Cardiovascular risk age: concepts and practicalities. *Heart* **98,** 941–6 (2012).

47. Criqui, M. H. *et al.* Calcium density of coronary artery plaque and risk of incident cardiovascular events. *JAMA* **311,** 271–8 (2014).

48. Cross, D. S., McCarty, C. A., Steinhubl, S. R., Carey, D. J. & Erlich, P. M. Development of a multi-institutional cohort to facilitate cardiovascular disease biomarker validation using existing biorepository samples linked to electronic health records. *Clin Cardiol* **36,** 486–91 (2013).

49. Daniels, L. B. *et al.* Serial measurement of N-terminal pro-B-type natriuretic peptide and cardiac troponin T for cardiovascular disease risk assessment in the Multi-Ethnic Study of Atherosclerosis (MESA). *American Heart Journal* **170,** 1170–1183 (2015).

50. De Las Heras Gala, T. *et al.* Recalibration of the ACC/AHA risk score in two population-based German cohorts. *PLoS ONE* **11,** e0164688 (2016).

51. De Lemos, J. A. *et al.* Multimodality Strategy for Cardiovascular Risk Assessment: Performance in 2 Population-Based Cohorts. *Circulation* **135,** 2119–2132 (2017).

52. De Vito, K. M. *et al.* Validation of a risk prediction tool for coronary heart disease in middle-aged women. *BMC Womens Health* **15,** 101 (2015).

53. DeFilippis, A. P., Young, R. & Blaha, M. J. Calibration and Discrimination Among Multiple Cardiovascular Risk Scores in a Modern Multiethnic Cohort RESPONSE. *Annals of Internal Medicine* **163,** 68–69 (2015).

54. DeFilippis, A. P. *et al.* Risk score overestimation: the impact of individual cardiovascular risk factors and preventive therapies on the performance of the American Heart Association-American College of Cardiology-Atherosclerotic Cardiovascular Disease risk score in a modern multi-ethnic cohort. *European Heart Journal* **38,** 598–608 (2017).

55. Deo, R. *et al.* Development and Validation of a Sudden Cardiac Death Prediction Model for the General Population. *Circulation* **134,** 806–16 (2016).

56. Desai, C. S., Bartz, T. M., Gottdiener, J. S., Lloyd-Jones, D. M. & Gardin, J. M. Usefulness of Left Ventricular Mass and Geometry for Determining 10-Year Prediction of Cardiovascular Disease in Adults Aged ¿65 Years (from the Cardiovascular Health Study). *Am J Cardiol* **118,** 684–90 (2016).

57. Dhoble, A. *et al.* Predicting long-term cardiovascular risk using the mayo clinic cardiovascular risk score in a referral population. *Am J Cardiol* **114,** 704–10 (2014).

58. Di Angelantonio, E. *et al.* Glycated hemoglobin measurement and prediction of cardiovascular disease. *JAMA - Journal of the American Medical Association* **311,** 1225–1233 (2014).

59. Dufouil, C. *et al.* Revised Framingham Stroke Risk Profile to Reflect Temporal Trends. *Circulation* **135,** 1145–1159 (2017).

60. Eggers, K. M. *et al.* Evaluation of Temporal Changes in Cardiovascular Biomarker Concentrations Improves Risk Prediction in an Elderly Population from the Community. *Clin Chem* **62,** 485–93 (2016).

61. Emdin, C. A. *et al.* Evaluation of the Pooled Cohort Equations for Prediction of Cardiovascular Risk in a Contemporary Prospective Cohort. *American Journal of Cardiology* **119,** 881–885 (2017).

62. Evans, T. E. *et al.* White Matter Microstructure Improves Stroke Risk Prediction in the General Population. *Stroke* **47,** 2756–2762 (2016).

63. Everett, B. M., Berger, J. S., Manson, J. E., Ridker, P. M. & Cook, N. R. B-type natriuretic peptides improve cardiovascular disease risk prediction in a cohort of women. *J Am Coll Cardiol* **64,** 1789–97 (2014).

64. Everett, B. M., Ridker, P. M., Cook, N. R. & Pradhan, A. D. Usefulness of B-type Natriuretic Peptides to Predict Cardiovascular Events in Women (from the Women's Health Study). *Am J Cardiol* **116,** 532–7 (2015).

65. Faeh, D., Rohrmann, S. & Braun, J. Better risk assessment with glycated hemoglobin instead of cholesterol in CVD risk prediction charts. *Eur J Epidemiol* **28,** 551–5 (2013).

66. Fatema, K., Rahman, B., Zwar, N. A., Milton, A. H. & Ali, L. Short-term predictive ability of selected cardiovascular risk prediction models in a rural Bangladeshi population: a case-cohort study. *BMC Cardiovasc Disord* **16,** 105 (2016).

67. Ferrario, M. M. *et al.* The contribution of educational class in improving accuracy of cardiovascular risk prediction across European regions: The MORGAM Project Cohort Component. *Heart* **100,** 1179–87 (2014).

68. Fogacci, F. *et al.* Serum lipoprotein(a) level as long-term predictor of cardiovascular mortality in a large sample of subjects in primary cardiovascular prevention: data from the Brisighella Heart Study. *Eur J Intern Med* **37,** 49–55 (2017).

69. Foraker, R. E. *et al.* Comparison of risk scores for the prediction of stroke in African Americans: Findings from the Jackson Heart Study. *Am Heart J* **177,** 25–32 (2016).

70. Fowkes, F. G. *et al.* Development and validation of an ankle brachial index risk model for the prediction of cardiovascular events. *Eur J Prev Cardiolog* **21,** 310–20 (2014).

71. Fox, E. R. *et al.* Development and Validation of Risk Prediction Models for Cardiovascular Events in Black Adults: The Jackson Heart Study Cohort. *JAMA Cardiol* **1,** 15–25 (2016).

72. Funke-Kaiser, A. *et al.* Predictive value of midregional pro-adrenomedullin compared to natriuretic peptides for incident cardiovascular disease and heart failure in the population-based FINRISK 1997 cohort. *Annals of Medicine* **46,** 155–162 (2014).

73. Gaibazzi, N. *et al.* Ultrasound carotid intima-media thickness, carotid plaque and cardiac calcium incrementally add to the Framingham Risk Score for the prediction of angiographic coronary artery disease: A multicenter prospective study. *International Journal of Cardiology* **177,** 708–710 (2014).

74. Garcia-Gil, M. *et al.* Role of renal function in cardiovascular risk assessment: A retrospective cohort study in a population with low incidence of coronary heart disease. *Prev Med* **89,** 200–6 (2016).

75. Gardin, J. M., Bartz, T. M., Polak, J. F., O'Leary, D. H. & Wong, N. D. What do carotid intima-media thickness and plaque add to the prediction of stroke and cardiovascular disease risk in older adults? The cardiovascular health study. *J Am Soc Echocardiogr* **27,** 998–1005 e2 (2014).

76. Georgousopoulou, E. N. *et al.* Physical Activity Level Improves the Predictive Accuracy of Cardiovascular Disease Risk Score: The ATTICA Study (2002-2012). *Int J Prev Med* **7,** 52 (2016).

77. Georgousopoulou, E. N., Panagiotakos, D. B., Pitsavos, C., Stefanadis, C. & group, A. s. Assessment of diet quality improves the classification ability of cardiovascular risk score in predicting future events: The 10-year follow-up of the ATTICA study (2002-2012). *Eur J Prev Cardiol* **22,** 1488–98 (2015).

78. Gibson, A. O. *et al.* Coronary artery calcium and incident cerebrovascular events in an asymptomatic cohort. The MESA Study. *JACC Cardiovasc Imaging* **7,** 1108–15 (2014).

79. Goff D. C., J. *et al.* 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol* **63,** 2935–59 (2014).

80. Goh, L. G., Dhaliwal, S. S., Welborn, T. A., Lee, A. H. & Della, P. R. Anthropometric measurements of general and central obesity and the prediction of cardiovascular disease risk in women: a cross-sectional study. *BMJ Open* **4,** e004138 (2014).

81. Goh, L. G., Welborn, T. A. & Dhaliwal, S. S. Independent external validation of cardiovascular disease mortality in women utilising Framingham and SCORE risk models: a mortality follow-up study. *BMC Womens Health* **14,** 118 (2014).

82. Grau, M. *et al.* Validation of a population coronary disease predictive system: the CASSANDRA model. *J Epidemiol Community Health* **68,** 1012–9 (2014).

83. Graversen, P., Abildstrom, S. Z., Jespersen, L., Borglykke, A. & Prescott, E. Cardiovascular risk prediction: Can Systematic Coronary Risk Evaluation (SCORE) be improved by adding simple risk markers? Results from the Copenhagen City Heart Study. *Eur J Prev Cardiol* **23,** 1546–56 (2016).

84. Greve, S. V. *et al.* Estimated carotid-femoral pulse wave velocity has similar predictive value as measured carotid-femoral pulse wave velocity. *Journal of Hypertension* **34,** 1279–1289 (2016).

85. Gronewold, J. *et al.* Ankle-brachial index predicts stroke in the general population in addition to classical risk factors. *Atherosclerosis* **233,** 545–50 (2014).

86. Groot, A. *et al.* Measurement of ECG abnormalities and cardiovascular risk classification: a cohort study of primary care patients in the Netherlands. *Br J Gen Pract* **65,** e1–8 (2015).

87. Guarrera, S. *et al.* Gene-specific DNA methylation profiles and LINE-1 hypomethylation are associated with myocardial infarction risk. *Clin Epigenetics* **7,** 133 (2015).

88. Gullu, O., Tekindal, M. A., Ates, C., Ekici, B. & Yavuz, Y. Assessing the classification performance of the mean platelet volume (MPV) in a cardiovascular risk prediction model. *Biomedical Research (India)* **28,** 705–710 (2017).

89. Hadamitzky, M. *et al.* Optimized Prognostic Score for Coronary Computed Tomographic Angiography Results From the CONFIRM Registry (COronary CT Angiography EvaluatioN For Clinical Outcomes: An InteRnational Multicenter Registry). *Journal of the American College of Cardiology* **62,** 468–476 (2013).

90. Hajifathalian, K. *et al.* A novel risk score to predict cardiovascular disease risk in national populations (Globorisk): a pooled analysis of prospective cohorts and health examination surveys. *Lancet Diabetes Endocrinol* **3,** 339–55 (2015).

91. Hartaigh, B. O. *et al.* 15-Year prognostic utility of coronary artery calcium scoring for all-cause mortality in the elderly. *Atherosclerosis* **246,** 361–6 (2016).

92. Hensley, W. J. & Mansfield, C. H. Lipoproteins, atherogenicity, age and risk of myocardial infarction. *Australian and New Zealand Journal of Public Health* **23,** 174–178 (1999).

93. Hippisley-Cox, J., Coupland, C. & Brindle, P. Derivation and validation of QStroke score for predicting risk of ischaemic stroke in primary care and comparison with other risk scores: a prospective open cohort study. *BMJ* **346,** f2573 (2013).

94. Hippisley-Cox, J., Coupland, C. & Brindle, P. The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study. *BMJ Open* **4,** e005809 (2014).

95. Hippisley-Cox, J., Coupland, C. & Brindle, P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* **357,** j2099 (2017).

96. Ho, J. E. *et al.* Predicting Heart Failure With Preserved and Reduced Ejection Fraction: The International Collaboration on Heart Failure Subtypes. *Circ. Heart fail.* **9(6)** (2016).

97. Hoffmann, U. *et al.* Cardiovascular Event Prediction and Risk Reclassification by Coronary, Aortic, and Valvular Calcification in the Framingham Heart Study. *J Am Heart Assoc* **5** (2016).

98. Horne, B. D., Anderson, J. L., Muhlestein, J. B., Ridker, P. M. & Paynter, N. P. Complete blood count risk score and its components, including RDW, are associated with mortality in the JUPITER trial. *European Journal of Preventive Cardiology* **22,** 519–526 (2015).

99. Hoshide, S. *et al.* Morning and Evening Home Blood Pressure and Risks of Incident Stroke and Coronary Artery Disease in the Japanese General Practice Population: The Japan Morning Surge-Home Blood Pressure Study. *Hypertension* **68,** 54–61 (2016).

100. Howard, G. *et al.* Self-Reported Stroke Risk Stratification: Reasons for Geographic and Racial Differences in Stroke Study. *Stroke* **48,** 1737–1743 (2017).

101. Htun, N. M. *et al.* Prediction of acute coronary syndromes by urinary proteome analysis. *PLoS One* **12,** e0172036 (2017).

102. Hu, G., Root, M. & Duncan, A. W. Adding multiple risk factors improves Framingham coronary heart disease risk scores. *Vasc Health Risk Manag* **10,** 557–62 (2014).

103. Hwang, Y. C., Park, C. Y., Ahn, H. Y. & Cho, N. H. Prediction of future development of cardiovascular disease with an equation to estimate apolipoprotein B: A community-based cohort study. *Medicine (Baltimore)* **95,** e3644 (2016).

104. Inzhakova, G. *et al.* Potential of Risk-Based Population Guidelines to Reduce Cardiovascular Risk in a Large Integrated Health System. *American Journal of Managed Care* **22,** E161–+ (2016).

105. Iribarren, C. *et al.* High-sensitivity cardiac troponin I and incident coronary heart disease among asymptomatic older adults. *Heart* **102,** 1177–82 (2016).

106. Iribarren, C. *et al.* Clinical Utility of Multimarker Genetic Risk Scores for Prediction of Incident Coronary Heart Disease: A Cohort Study Among Over 51 Thousand Individuals of European Ancestry. *Circ Cardiovasc Genet* **9,** 531–540 (2016).

107. Israel, A. *et al.* Use of exercise capacity to improve SCORE risk prediction model in asymptomatic adults. *Eur Heart J* **37,** 2300–6 (2016).

108. Jairam, P. M. *et al.* Cardiovascular disease prediction: do pulmonary disease-related chest CT features have added value? *Eur Radiol* **25,** 1646–54 (2015).

109. Jdanov, D. A. *et al.* Recalibration of the SCORE risk chart for the Russian population. *European Journal of Epidemiology* **29,** 621–628 (2014).

110. Jee, S. H. *et al.* A coronary heart disease prediction model: the Korean Heart Study. *BMJ Open* **4,** e005025 (2014).

111. Johansson, J. K. *et al.* Health 2000 score - development and validation of a novel cardiovascular risk score. *Ann Med* **48,** 403–409 (2016).

112. Jorstad, H. T., Boekholdt, S. M., Wareham, N. J., Khaw, K. T. & Peters, R. J. The Dutch SCORE-based risk charts seriously underestimate the risk of cardiovascular disease. *Netherlands Heart Journal* **25,** 173–180 (2017).

113. Jorstad, H. T. *et al.* The Systematic COronary Risk Evaluation (SCORE) in a large UK population: 10-year follow-up in the EPIC-Norfolk prospective population study. *European Journal of Preventive Cardiology* **22,** 119–26 (2015).

114. Jung, C. H. *et al.* 2013 ACC/AHA versus 2004 NECP ATP III Guidelines in the Assignment of Statin Treatment in a Korean Population with Subclinical Coronary Atherosclerosis. *PLoS ONE [Electronic Resource]* **10,** e0137478 (2015).

115. Jung, K. J. *et al.* The ACC/AHA 2013 pooled cohort equations compared to a Korean Risk Prediction Model for atherosclerotic cardiovascular disease. *Atherosclerosis* **242,** 367–75 (2015).

116. Kaess, B. M. *et al.* Ultra-sensitive troponin I is an independent predictor of incident coronary heart disease in the general population. *Eur J Epidemiol* **32,** 583–591 (2017).

117. Kalsch, H. *et al.* Beyond Framingham risk factors and coronary calcification: does aortic valve calcification improve risk prediction? The Heinz Nixdorf Recall Study. *Heart* **100,** 930–7 (2014).

118. Kanjilal, S. *et al.* Application of cardiovascular disease risk prediction models and the relevance of novel biomarkers to risk stratification in Asian Indians. *Vasc Health Risk Manag* **4,** 199–211 (2008).

119. Karas, M. G. *et al.* Incremental value of biochemical and echocardiographic measures in prediction of ischemic stroke: the Strong Heart Study. *Stroke* **43,** 720–6 (2012).

120. Karjalainen, T. *et al.* An evaluation of the performance of SCORE Sweden 2015 in estimating cardiovascular risk: The Northern Sweden MONICA Study 1999-2014. *Eur J Prev Cardiol* **24,** 103–110 (2017).

121. Karmali, K. N., Ning, H., Goff, D. C. & Lloyd-Jones, D. M. Identifying Individuals at Risk for Cardiovascular Events Across the Spectrum of Blood Pressure Levels. *J Am Heart Assoc* **4,** e002126 (2015).

122. Kavousi, M. *et al.* Prevalence and Prognostic Implications of Coronary Artery Calcification in Low-Risk Women: A Meta-analysis. *JAMA* **316,** 2126–2134 (2016).

123. Kavousi, M. *et al.* Comparison of application of the ACC/AHA guidelines, Adult Treatment Panel III guidelines, and European Society of Cardiology guidelines for cardiovascular disease prevention in a European cohort. *JAMA* **311,** 1416–23 (2014).

124. Khalili, D. *et al.* A new approach to test validity and clinical usefulness of the 2013 ACC/AHA guideline on statin therapy: A population-based study. *Int J Cardiol* **184,** 587–94 (2015).

125. Kim, T. H. *et al.* Subclinical hypothyroidism in addition to common risk scores for prediction of cardiovascular disease: a 10-year community-based cohort study. *Eur J Endocrinol* **171,** 649–57 (2014).

126. Konety, S. H. *et al.* Echocardiographic Predictors of Sudden Cardiac Death: The Atherosclerosis Risk in Communities Study and Cardiovascular Health Study. *Circ Cardiovasc Imaging* **9** (2016).

127. Konfino, J. *et al.* Comparing Strategies for Lipid Lowering in Argentina: An Analysis from the CVD Policy Model-Argentina. *J Gen Intern Med* **32,** 524–533 (2017).

128. Kovalchik, S. A. & Pfeiffer, R. M. Population-based absolute risk estimation with survey data. *Lifetime Data Anal* **20,** 252–75 (2014).

129. Kreger, B. E., Cupples, L. A. & Kannel, W. B. The electrocardiogram in prediction of sudden death: Framingham Study experience. *Am Heart J* **113,** 377–82 (1987).

130. Kumarathurai, P. *et al.* Atrial ectopy and N-terminal pro-B-type natriuretic peptide as predictors of atrial fibrillation: a population-based cohort study. *Europace* **19,** 364–370 (2017).

131. Kunutsor, S. K., Bakker, S. J., Kootstra-Ros, J. E., Gansevoort, R. T. & Dullaart, R. P. Circulating gamma glutamyltransferase and prediction of cardiovascular disease. *Atherosclerosis* **238,** 356–64 (2015).

132. Kunutsor, S. K. *et al.* Serum Alkaline Phosphatase and Risk of Incident Cardiovascular Disease: Interrelationship with High Sensitivity C-Reactive Protein. *PLoS One* **10,** e0132822 (2015).

133. Kusmana, D. The influence of smoking cessation, regular physical exercise and/or physical activity on survival: A 13 years cohort study of the Indonesian population in Jakarta. *Medical Journal of Indonesia* **11,** 230–242 (2002).

134. Langley, S. R. *et al.* Extracellular matrix proteomics identifies molecular signature of symptomatic carotid plaques. *J Clin Invest* **127,** 1546–1560 (2017).

135. Lee, C. H. *et al.* Validation of the Pooled Cohort equations in a long-term cohort study of Hong Kong Chinese. *J Clin Lipidol* **9,** 640–6 e2 (2015).

136. Lee, J. W. *et al.* Incremental prognostic value of computed tomography in stroke: rationale and design of the IMPACTS study. *Int J Cardiovasc Imaging* **32 Suppl 1,** 83–9 (2016).

137. Lin, Y. H., Glei, D., Weinstein, M., Wu, S. I. & Chien, K. L. Additive value of interleukin-6 and C-reactive protein in risk prediction for all-cause and cardiovascular mortality among a representative adult cohort in Taiwan. *J Formos Med Assoc* **116,** 982–992 (2017).

138. Lindberg, S. *et al.* Plasma Neutrophil Gelatinase-Associated Lipocalinin in the General Population Association With Inflammation and Prognosis. *Arteriosclerosis Thrombosis and Vascular Biology* **34,** 2135–2142 (2014).

139. Liu, L. *et al.* A novel risk score to the prediction of 10-year risk for coronary artery disease among the elderly in Beijing based on competing risk model. *Medicine (United States)* **95** (2016).

140. Lluis-Ganella, C. *et al.* Assessment of the value of a genetic risk score in improving the estimation of coronary risk. *Atherosclerosis* **222,** 456–63 (2012).

141. Lopez-Suarez, A. *et al.* Metabolic syndrome does not improve the prediction of 5-year cardiovascular disease and total mortality over standard risk markers. Prospective population based study. *Medicine (Baltimore)* **93,** e212 (2014).

142. Lyngbaek, S. *et al.* Cardiovascular risk prediction in the general population with use of suPAR, CRP, and Framingham Risk Score. *Int J Cardiol* **167,** 2904–11 (2013).

143. Mahabadi, A. A. *et al.* Association of epicardial fat with cardiovascular risk factors and incident myocardial infarction in the general population: the Heinz Nixdorf Recall Study. *J Am Coll Cardiol* **61,** 1388–95 (2013).

144. Mahabadi, A. A. *et al.* Noncoronary Measures Enhance the Predictive Value of Cardiac CT Above Traditional Risk Factors and CAC Score in the General Population. *JACC Cardiovasc Imaging* **9,** 1177–1185 (2016).

145. Manuel, D. G. *et al.* Predicting Stroke Risk Based on Health Behaviours: Development of the Stroke Population Risk Tool (SPoRT). *PLoS One* **10,** e0143342 (2015).

146. Marino, M. *et al.* Quantifying cardiometabolic risk using modifiable non-self-reported risk factors. *Am J Prev Med* **47,** 131–40 (2014).

147. Marrugat, J. *et al.* Derivation and validation of a set of 10-year cardiovascular risk predictive functions in Spain: the FRESCO Study. *Prev Med* **61,** 66–74 (2014).

148. Matsushita, K. *et al.* Cardiac and kidney markers for cardiovascular prediction in individuals with chronic kidney disease: the Atherosclerosis Risk in Communities study. *Arterioscler Thromb Vasc Biol* **34,** 1770–7 (2014).

149. McClelland, R. L. *et al.* 10-Year Coronary Heart Disease Risk Prediction Using Coronary Artery Calcium and Traditional Risk Factors: Derivation in the MESA (Multi-Ethnic Study of Atherosclerosis) With Validation in the HNR (Heinz Nixdorf Recall) Study and the DHS (Dallas Heart Study). *J Am Coll Cardiol* **66,** 1643–53 (2015).

150. McClure, L. A. *et al.* Assessing the performance of the Framingham Stroke Risk Score in the reasons for geographic and racial differences in stroke cohort. *Stroke* **45,** 1716–20 (2014).

151. McNeil, J. J., Peeters, A., Liew, D., Lim, S. & Vos, T. A model for predicting the future incidence of coronary heart disease within percentiles of coronary heart disease risk. *J Cardiovasc Risk* **8,** 31–7 (2001).

152. Mody, P., Joshi, P. H., Khera, A., Ayers, C. R. & Rohatgi, A. Beyond Coronary Calcification, Family History, and C-Reactive Protein: Cholesterol Efflux Capacity and Cardiovascular Risk Prediction. *J Am Coll Cardiol* **67,** 2480–7 (2016).

153. Morris, R. W. *et al.* Marginal role for 53 common genetic variants in cardiovascular disease prediction. *Heart* **102,** 1640–7 (2016).

154. Mortensen, M. B., Afzal, S., Nordestgaard, B. G. & Falk, E. The high-density lipoprotein-adjusted SCORE model worsens SCORE-based risk classification in a contemporary population of 30,824 Europeans: the Copenhagen General Population Study. *Eur Heart J* **36,** 2446–53 (2015).

155. Mortensen, M. B., Nordestgaard, B. G., Afzal, S. & Falk, E. ACC/AHA guidelines superior to ESC/EAS guidelines for primary prevention with statins in non-diabetic Europeans: the Copenhagen General Population Study. *European Heart Journal* **38,** 586–594 (2017).

156. Muntner, P. *et al.* Validation of the atherosclerotic cardiovascular disease Pooled Cohort risk equations. *JAMA* **311,** 1406–15 (2014).

157. Nambi, V. *et al.* Troponin T and N-terminal pro-B-type natriuretic peptide: a biomarker approach to predict heart failure risk–the atherosclerosis risk in communities study. *Clin Chem* **59,** 1802–10 (2013).

158. Nargesi, A. A. *et al.* Contribution of vitamin D deficiency to the risk of coronary heart disease in subjects with essential hypertension. *Atherosclerosis* **244,** 165–71 (2016).

159. Nauman, J. *et al.* Prediction of Cardiovascular Mortality by Estimated Cardiorespiratory Fitness Independent of Traditional Risk Factors: The HUNT Study. *Mayo Clin Proc* **92,** 218–227 (2017).

160. Ndumele, C. E. *et al.* N-Terminal Pro-Brain Natriuretic Peptide and Heart Failure Risk Among Individuals With and Without Obesity: The Atherosclerosis Risk in Communities (ARIC) Study. *Circulation* **133,** 631–8 (2016).

161. Nielsen, J. B. *et al.* Risk prediction of cardiovascular death based on the QTc interval: evaluating age and gender differences in a large primary care population. *Eur Heart J* **35,** 1335–44 (2014).

162. Ninomiya, T. *et al.* Brachial-ankle pulse wave velocity predicts the development of cardiovascular disease in a general Japanese population: the Hisayama Study. *J Hypertens* **31,** 477–83, discussion 483 (2013).

163. Nishimura, K. *et al.* Predicting coronary heart disease using risk factor categories for a Japanese urban population, and comparison with the framingham risk score: the suita study. *J Atheroscler Thromb* **21,** 784–98 (2014).

164. Nobel, L., Mayo, N. E., Hanley, J., Nadeau, L. & Daskalopoulou, S. S. MyRisk Stroke Calculator: A Personalized Stroke Risk Assessment Tool for the General Population. *J Clin Neurol* **10,** 1–9 (2014).

165. Ogunwale, A. N. *et al.* The impact of multiple single day blood pressure readings on cardiovascular risk estimation: The Atherosclerosis Risk in Communities study. *Eur J Prev Cardiol* **23,** 1529–36 (2016).

166. Okwuosa, T. M. *et al.* Left ventricular hypertrophy and cardiovascular disease risk prediction and reclassification in blacks and whites: the Atherosclerosis Risk in Communities Study. *Am Heart J* **169,** 155–61 e5 (2015).

167. Onat, A. *et al.* Algorithm for predicting CHD death risk in Turkish adults: conventional factors contribute only moderately in women. *Anatol J Cardiol* **17,** 436–444 (2017).

168. Paixao, A. R. *et al.* Coronary Artery Calcium Improves Risk Classification in Younger Populations. *JACC Cardiovasc Imaging* **8,** 1285–93 (2015).

169. Panagiotakos, D. B., Georgousopoulou, E. N., Fitzgerald, A. P., Pitsavos, C. & Stefanadis, C. Validation of the HellenicSCORE (a Calibration of the ESC SCORE Project) Regarding 10-Year Risk of Fatal Cardiovascular Disease in Greece. *Hellenic Journal of Cardiology* **56,** 302–308 (2015).

170. Parikh, N. I. *et al.* Reproductive Risk Factors and Coronary Heart Disease in the Women's Health Initiative Observational Study. *Circulation* **133,** 2149–58 (2016).

171. Park, G. M. *et al.* Model for assessing cardiovascular risk in a Korean population. *Circ Cardiovasc Qual Outcomes* **7,** 944–51 (2014).

172. Parmar, P. *et al.* The Stroke Riskometer(TM) App: validation of a data collection tool and stroke risk predictor. *Int J Stroke* **10,** 231–44 (2015).

173. Paynter, N. P., Crainiceanu, C. M., Sharrett, A. R., Chambless, L. E. & Coresh, J. Effect of correcting for long-term variation in major coronary heart disease risk factors: relative hazard estimation and risk prediction in the Atherosclerosis Risk in Communities Study. *Ann Epidemiol* **22,** 191–7 (2012).

174. Paynter, N. P. *et al.* Comparison of lifestyle-based and traditional cardiovascular disease prediction in a multiethnic cohort of nonsmoking women. *Circulation* **130,** 1466–73 (2014).

175. Pencina, M. J. *et al.* Apolipoprotein B improves risk assessment of future coronary heart disease in the Framingham Heart Study beyond LDL-C and non-HDL-C. *Eur J Prev Cardiol* **22,** 1321–7 (2015).

176. Perez, H. A., Garcia, N. H., Spence, J. D. & Armando, L. J. Adding carotid total plaque area to the Framingham risk score improves cardiovascular risk classification. *Arch Med Sci* **12,** 513–20 (2016).

177. Piotrowski, W., Waskiewicz, A. & Cicha-Mikolajczyk, A. Global cardiovascular mortality risk in the adult Polish population: prospective assessment of the cohorts studied in multicentre national WOBASZ and WOBASZ Senior studies. *Kardiol Pol* **74,** 262–73 (2016).

178. Polak, J. F., Szklo, M. & O'Leary, D. H. Associations of Coronary Heart Disease with Common Carotid Artery Near and Far Wall Intima-Media Thickness: The Multi-Ethnic Study of Atherosclerosis. *J Am Soc Echocardiogr* **28,** 1114–21 (2015).

179. Polenz, G. F. *et al.* CHA2 DS2 VASc Score as a Predictor of Cardiovascular Events in Ambulatory Patients without Atrial Fibrillation. *Pacing Clin Electrophysiol* **38,** 1412–7 (2015).

180. Qureshi, W. T. *et al.* Impact of Replacing the Pooled Cohort Equation With Other Cardiovascular Disease Risk Scores on Atherosclerotic Cardiovascular Disease Risk Assessment (from the Multi-Ethnic Study of Atherosclerosis [MESA]). *Am J Cardiol* **118,** 691–6 (2016).

181. Raggi, P., Cooil, B. & Callister, T. Q. Use of electron beam tomography data to develop models for prediction of hard coronary events. *Am Heart J* **141,** 375–82 (2001).

182. Raghu, A., Praveen, D., Peiris, D., Tarassenko, L. & Clifford, G. Implications of Cardiovascular Disease Risk Assessment Using the WHO/ISH Risk Prediction Charts in Rural India. *PLoS One* **10,** e0133618 (2015).

183. Rana, J. S. *et al.* Accuracy of the Atherosclerotic Cardiovascular Risk Equation in a Large Contemporary, Multiethnic Population. *J Am Coll Cardiol* **67,** 2118–2130 (2016).

184. Romanens, M., Ackermann, F., Sudano, I., Szucs, T. & Spence, J. D. Arterial age as a substitute for chronological age in the AGLA risk function could improve coronary risk prediction. *Swiss Med Wkly* **144,** w13967 (2014).

185. Rucker, V. *et al.* Predicting 10-Year Risk of Fatal Cardiovascular Disease in Germany: An Update Based on the SCORE-Deutschland Risk Charts. *PLoS One* **11,** e0162188 (2016).

186. Sabayan, B., Gussekloo, J., de Ruijter, W., Westendorp, R. G. & de Craen, A. J. Framingham stroke risk score and cognitive impairment for predicting first-time stroke in the oldest old. *Stroke* **44,** 1866–71 (2013).

187. Salim, A. *et al.* C-reactive protein and serum creatinine, but not haemoglobin A1c, are independent predictors of coronary heart disease risk in non-diabetic Chinese. *Eur J Prev Cardiolog* **23,** 1339–49 (2016).

188. Sanchez-Inigo, L., Navarro-Gonzalez, D., Fernandez-Montero, A., Pastrana-Delgado, J. & Martinez, J. A. The TyG index may predict the development of cardiovascular events. *Eur J Clin Invest* **46,** 189–97 (2016).

189. Sardarinia, M., Ansari, R., Azizi, F., Hadaegh, F. & Bozorgmanesh, M. Mortality prediction of a body shape index versus traditional anthropometric measures in an Iranian population: Tehran Lipid and Glucose Study. *Nutrition* **33,** 105–112 (2017).

190. Sawano, M. *et al.* Validation of the european SCORE risk chart in the healthy middle-aged Japanese. *Atherosclerosis* **252,** 116–21 (2016).

191. Schiopu, A. *et al.* Associations Between Macrophage Colony-Stimulating Factor and Monocyte Chemotactic Protein 1 in Plasma and First-Time Coronary Events: A Nested Case-Control Study. *J Am Heart Assoc* **5** (2016).

192. Schnohr, P. *et al.* Ranking of psychosocial and traditional risk factors by importance for coronary heart disease: the Copenhagen City Heart Study. *Eur Heart J* **36,** 1385–93 (2015).

193. Sehestedt, T. *et al.* Can ambulatory blood pressure measurements substitute assessment of subclinical cardiovascular damage? *Journal of Hypertension* **30,** 513–21 (2012).

194. Seidelmann, S. B. *et al.* Retinal Vessel Calibers in Predicting Long-Term Cardiovascular Outcomes: The Atherosclerosis Risk in Communities Study. *Circulation* **134,** 1328–1338 (2016).

195. Selmer, R. *et al.* NORRISK 2: A Norwegian risk model for acute cerebral stroke and myocardial infarction. *European Journal of Preventive Cardiology* **24,** 773–782 (2017).

196. Selvarajah, S. *et al.* Comparison of the Framingham Risk Score, SCORE and WHO/ISH cardiovascular risk prediction models in an Asian population. *Int J Cardiol* **176,** 211–8 (2014).

197. Sepanlou, S. G. *et al.* The clinical performance of an office-based risk scoring system for fatal cardiovascular diseases in North-East of Iran. *PLoS One* **10,** e0126779 (2015).

198. Shah, N. *et al.* Red cell distribution width and risk of cardiovascular mortality: Insights from National Health and Nutrition Examination Survey (NHANES)-III. *Int J Cardiol* **232,** 105–110 (2017).

199. Shah, N. *et al.* Neutrophil lymphocyte ratio significantly improves the Framingham risk score in prediction of coronary heart disease mortality: insights from the National Health and Nutrition Examination Survey-III. *Int J Cardiol* **171,** 390–7 (2014).

200. Sharma, A., Ghatge, M., Mundkur, L. & Vangala, R. K. Translational informatics approach for identifying the functional molecular communicators linking coronary artery disease, infection and inflammation. *Mol Med Rep* **13,** 3904–12 (2016).

201. Shoamanesh, A. *et al.* Circulating biomarkers and incident ischemic stroke in the Framingham Offspring Study. *Neurology* **87,** 1206–11 (2016).

202. Sim, J., Teece, L., Dennis, M. S., Roffe, C. & Team, S. S. Validation and Recalibration of Two Multivariable Prognostic Models for Survival and Independence in Acute Stroke. *Plos One* **11** (2016).

203. Solbu, M. D. *et al.* N-Acetyl-beta-D-Glucosaminidase Does Not Enhance Prediction of Cardiovascular or All-Cause Mortality by Albuminuria in a Low-Risk Population. *J Am Soc Nephrol* **27,** 533–42 (2016).

204. Sonneveld, M. A. *et al.* Low ADAMTS13 activity is associated with an increased risk of ischemic stroke. *Blood* **126,** 2739–46 (2015).

205. Stam-Slob, M. C. *et al.* Personalized absolute benefit of statin treatment for primary or secondary prevention of vascular disease in individual elderly patients. *Clinical Research in Cardiology* **106,** 58–68 (2017).

206. Tada, H. *et al.* Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. *Eur Heart J* **37,** 561–7 (2016).

207. Tang, W. H. *et al.* Prognostic value of estimated functional capacity incremental to cardiac biomarkers in stable cardiac patients. *J Am Heart Assoc* **3,** e000960 (2014).

208. Thompson, S. G., Pyke, S. D. & Wood, D. A. Using a coronary risk score for screening and intervention in general practice. British Family Heart Study. *J Cardiovasc Risk* **3,** 301–6 (1996).

209. Tillin, T. *et al.* Ethnicity and prediction of cardiovascular disease: performance of QRISK2 and Framingham scores in a U.K. tri-ethnic prospective cohort study (SABRE–Southall And Brent REvisited). *Heart* **100,** 60–7 (2014).

210. Valenti, V. *et al.* A 15-Year Warranty Period for Asymptomatic Individuals Without Coronary Artery Calcium: A Prospective Follow-Up of 9,715 Individuals. *JACC Cardiovasc Imaging* **8,** 900–9 (2015).

211. Van der Meer, M. G. *et al.* Added Value of Female-Specific Factors Beyond Traditional Predictors for Future Cardiovascular Disease. *J Am Coll Cardiol* **67,** 2084–6 (2016).

212. Van Dis, I. *et al.* Effect of including nonfatal events in cardiovascular risk estimation, illustrated with data from The Netherlands. *Eur J Prev Cardiol* **21,** 377–83 (2014).

213. Van Dis, I. *et al.* Do obesity and parental history of myocardial infarction improve cardiovascular risk prediction? *European Journal of Preventive Cardiology* **20,** 793–9 (2013).

214. Van Kempen, B. J. *et al.* Performance of Framingham cardiovascular disease (CVD) predictions in the Rotterdam Study taking into account competing risks and disentangling CVD into coronary heart disease (CHD) and stroke. *Int J Cardiol* **171,** 413–8 (2014).

215. Vartiainen, E., Laatikainen, T., Peltonen, M. & Puska, P. Predicting Coronary Heart Disease and Stroke The FINRISK Calculator. *Global Heart* **11,** 213–216 (2016).

216. Velescu, A. *et al.* Adding low ankle brachial index to classical risk factors improves the prediction of major cardiovascular events. The REGICOR study. *Atherosclerosis* **241,** 357–63 (2015).

217. Venkatesh, S., O'Neal, W. T., Broughton, S. T., Shah, A. J. & Soliman, E. Z. Utility of Normal Findings on Electrocardiogram and Echocardiogram in Subjects >= 65 Years. *American Journal of Cardiology* **119,** 856–861 (2017).

218. Verbeek, R. *et al.* Lipoprotein(a) Improves Cardiovascular Risk Prediction Based on Established Risk Algorithms. *J Am Coll Cardiol* **69,** 1513–1515 (2017).

219. Veronesi, G. *et al.* Combined use of short-term and long-term cardiovascular risk scores in primary prevention: An assessment of clinical utility. *Journal of Cardiovascular Medicine* **18,** 318–324 (2017).

220. Veronesi, G. *et al.* Validity of a long-term cardiovascular disease risk prediction equation for low-incidence populations: the CAMUNI-MATISS Cohorts Collaboration study. *Eur J Prev Cardiol* **22,** 1618–25 (2015).

221. Veronesi, G. *et al.* Improving long-term prediction of first cardiovascular event: The contribution of family history of coronary heart disease and social status. *Preventive Medicine* **64,** 75–80 (2014).

222. Vikhireva, O. *et al.* Does Inclusion of Education and Marital Status Improve SCORE Performance in Central and Eastern Europe and Former Soviet Union? Findings from MONICA and HAPIEE Cohorts. *Plos One* **9** (2014).

223. Vikhireva, O. *et al.* SCORE performance in Central and Eastern Europe and former Soviet Union: MONICA and HAPIEE results. *Eur Heart J* **35,** 571–7 (2014).

224. Waks, J. W. *et al.* Global Electric Heterogeneity Risk Score for Prediction of Sudden Cardiac Death in the General Population: The Atherosclerosis Risk in Communities (ARIC) and Cardiovascular Health (CHS) Studies. *Circulation* **133,** 2222–34 (2016).

225. Wang, Y. *et al.* Lifetime risk of stroke in young-aged andmiddle-aged Chinese population: the Chinese Multi-Provincial Cohort Study. *Journal of Hypertension* **34,** 2434–2440 (2016).

226. Wassertheil-Smoller, S. *et al.* Improvement in stroke risk prediction: role of C-reactive protein and lipoprotein-associated phospholipase A2 in the women's health initiative. *Int J Stroke* **9,** 902–9 (2014).

227. Weatherley, N. D. & Jackson, P. R. The new Sheffield risk and benefit tables for the elderly. *QJM* **104,** 3–12 (2011).

228. Welsh, P. *et al.* Prediction of Cardiovascular Disease Risk by Cardiac Biomarkers in 2 United Kingdom Cohort Studies: Does Utility Depend on Risk Thresholds For Treatment? *Hypertension* **67,** 309–15 (2016).

229. Weng, S. F., Kai, J., Guha, I. N. & Qureshi, N. The value of aspartate aminotransferase and alanine aminotransferase in cardiovascular disease risk assessment. *Open Heart* **2,** e000272 (2015).

230. Wickramasinghe, C. D. *et al.* Prediction of 30-Year Risk for Cardiovascular Mortality by Fitness and Risk Factor Levels The Cooper Center Longitudinal Study. *Circulation-Cardiovascular Quality and Outcomes* **7,** 597–602 (2014).

231. Willeit, P. *et al.* Discrimination and net reclassification of cardiovascular risk with lipoprotein(a): prospective 15-year outcomes in the Bruneck Study. *J Am Coll Cardiol* **64,** 851–60 (2014).

232. Woodward, M. Coronary risk prediction for those with and without diabetes. *European Journal of Cardiovascular Prevention and Rehabilitation* **13,** 30–36 (2006).

233. Woodward, M., Brindle, P., Tunstall-Pedoe, H. & estimation, S. g. o. r. Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). *Heart* **93,** 172–6 (2007).

234. Woznicka-Leskiewicz, L., Posadzy-Malaczynska, A. & Juszkat, R. The impact of ankle brachial index and pulse wave velocity on cardiovascular risk according to SCORE and Framingham scales and sex differences. *J Hum Hypertens* **29,** 502–10 (2015).

235. Wurtz, P. *et al.* Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* **131,** 774–85 (2015).

236. Yang, X. *et al.* Predicting the 10-Year Risks of Atherosclerotic Cardiovascular Disease in Chinese Population: The China-PAR Project (Prediction for ASCVD Risk in China). *Circulation* **134,** 1430–1440 (2016).

237. Yatsuya, H. *et al.* Development of a Risk Equation for the Incidence of Coronary Artery Disease and Ischemic Stroke for Middle-Aged Japanese- Japan Public Health Center-Based Prospective Study. *Circ J* **80,** 1386–95 (2016).

238. Yeboah, J. *et al.* Computed tomography-derived cardiovascular risk markers, incident cardiovascular events, and all-cause mortality in nondiabetics: the Multi-Ethnic Study of Atherosclerosis. *Eur J Prev Cardiol* **21,** 1233–41 (2014).

239. Yeboah, J. *et al.* Utility of Nontraditional Risk Markers in Atherosclerotic Cardiovascular Disease Risk Assessment. *J Am Coll Cardiol* **67,** 139–147 (2016).

240. Yuan, Z. *et al.* Risk Prediction for Ischemic Stroke and Transient Ischemic Attack in Patients Without Atrial Fibrillation: A Retrospective Cohort Study. *J Stroke Cerebrovasc Dis* **26,** 1721–1731 (2017).

241. Yudkin, J. Beta-blockers in heart failure. *Diabet Med* **16,** 785–7 (1999).

242. Zalawadiya, S. K. *et al.* Uric acid and cardiovascular disease risk reclassification: findings from NHANES III. *Eur J Prev Cardiol* **22,** 513–8 (2015).

243. Zhang, Y. *et al.* Smoking-Associated DNA Methylation Biomarkers and Their Predictive Value for All-Cause and Cardiovascular Mortality. *Environ Health Perspect* **124,** 67–74 (2016).

244. Zhang, Z., Gillespie, C., Bowman, B. & Yang, Q. Prediction of atherosclerotic cardiovascular disease mortality in a nationally representative cohort using a set of risk factors from pooled cohort risk equations. *PLoS ONE* **12** (2017).

245. Zhou, X. H., Wang, X., Duncan, A., Hu, G. & Zheng, J. Statistical evaluation of adding multiple risk factors improves Framingham stroke risk score. *BMC Med Res Methodol* **17,** 58 (2017).

246. Ziegelbauer, K., Schaefer, C., Steinmetz, H., Sitzer, M. & Lorenz, M. W. Clinical usefulness of carotid ultrasound to improve stroke risk assessment: ten-year results from the Carotid Atherosclerosis Progression Study (CAPS). *Eur J Prev Cardiol* **20,** 837–43 (2013).

Chapter 5

Influence of socioeconomic factors on pregnancy outcome in women with structural heart disease

van Hagen IM, **Baart SJ**, Fong Soe Khioe R, Sliwa-Hahnle K, Taha N, Lelonek M, Tavazzi L, Maggioni AP, Johnson MR, Maniadakis N, Fordham R, Hall R, Roos-Hesselink JW, on behalf of the ROPAC investigators

Heart 2018; 104:745-752

Abstract

Objective: Cardiac disease is the leading cause of indirect maternal mortality. The aim of this study was to analyze to what extent socioeconomic factors influence the outcome of pregnancy in women with heart disease.

Methods: The Registry of Pregnancy and Cardiac disease is a global prospective registry. For this analysis, countries that enrolled ≥10 patients were included. A combined cardiac endpoint included maternal cardiac death, arrhythmia requiring treatment, heart failure, thromboembolic event, aortic dissection, endocarditis, acute coronary syndrome, hospitalisation for cardiac reason or intervention. Associations between patient characteristics, country characteristics (income inequality expressed as Gini coefficient, health expenditure, schooling, gross domestic product, birth rate and hospital beds) and cardiac endpoints were checked in a three-level model (patient–centre–country).

Results: A total of 30 countries enrolled 2924 patients from 89 centres. At least one endpoint occurred in 645 women (22.1%). Maternal age, New York Heart Association classification and modified WHO risk classification were associated with the combined endpoint and explained 37% of variance in outcome. Gini coefficient and country-specific birth rate explained an additional 4%. There were large differences between the individual countries, but the need for multilevel modelling to account for these differences disappeared after adjustment for patient characteristics, Gini and country-specific birth rate.

Conclusions: While there are definite interregional differences in pregnancy outcome in women with cardiac disease, these differences seem to be mainly driven by individual patient characteristics. Adjustment for country characteristics refined the results to a limited extent, but maternal condition seems to be the main determinant of outcome.

Introduction

Cardiac disease is an important cause of maternal mortality and morbidity. Recent data from the Global Burden of Disease study have demonstrated that geographical disparities widened between 1990 and 2015, and that in 2015, 24 countries still had a maternal mortality ratio greater than 400 per 100 000. Those recent data have shown that overall maternal mortality pattern is influenced by Socio-Demographic Index (SDI) with women in the highest SDI quintile dying frequently due to indirect maternal disorders as cardiovascular and thrombotic disease. [1, 2] The Registry Of Pregnancy And Cardiac disease (ROPAC) is a global cohort including pregnant patients from both advanced and emerging countries. Several analyses from ROPAC data have been published with marked differences between advanced and emerging countries. [3–5] These differences could be partly explained by variations in underlying cardiac condition, with acquired valvular disease being more prevalent in emerging countries [6] and congenital heart disease in advanced countries. In addition, the demographic differences may also influence outcome. For instance, in some cultures, women gain status by having (many) children and thus they may be reluctant to take a doctor's advice to avoid pregnancy. Also, there is widespread difference in the availability of healthcare and access to female contraception. Although tertiary care is provided in the urban areas, many women in less developed countries are from rural areas and, consequently, might present with pregnancy complaints much later than their peers from rural areas in countries with more advanced economies. [7]

Interpretation of ROPAC results needs to be done with caution in the light of these differences. Insights in country-level socioeconomic data and the associated pregnancy outcomes will help in interpreting existing and future analyses. Such an analysis could define the influence of socioeconomic background on pregnancy

outcome exerted by the countries of residence; the alternative approach of an in-depth analysis of individual socioeconomic data is not possible.

The aim of this study was to elucidate the inter-regional differences in the countries contributing to ROPAC, by analysing to what extent socioeconomic factors on country level, such as gross domestic product (GDP), income distribution/ inequality (Gini coefficient), Human Development Index (HDI), health expenditure, birth rate, number of hospital beds and schooling, influence the outcome of pregnancy in women with heart disease. We hypothesised that country-level socioeconomic indices do influence pregnancy outcome and that cardiac status (such as severity of disease and New York Heart Association (NYHA) classification) affects the outcome of mother and baby to a greater extent.

Methods

The ROPAC is an ongoing prospective worldwide registry that includes all consecutive pregnant women with structural heart disease. Study design and methods have been described in detail previously. [3] Patient enrolment started from January 2008, and for this interim analysis, we included patients with a term date up to October 2013, [1] 6 months' follow-up in April 2014. Patient informed consent was obtained when required by the local independent review board. Patients with either congenital, valvular or ischaemic heart disease, a cardiomyopathy, pulmonary hypertension and aortic pathology were included. Women with non-structural disease such as arrhythmia were excluded. More specific details on disease have been published previously. [3, 8]

Data

The patient characteristics collected at baseline (before pregnancy) included age, ECG rhythm, NYHA functional classification, diagnosis, risk factors for cardiovascular disease (smoking, diabetes, hypertension), previous interventions, medication, parity, obstetric history and, if available, echocardiographic parameters. Every patient was stratified according to the modified WHO classification, as stated in the latest guidelines [9, 10] by two authors (IMvH, JWRH). Modified WHO class I implies no increased risk of events during pregnancy, compared with the general pregnant population. Modified WHO class II has a small increased risk, class II–III a moderate increased risk and class III has a 'significantly' increased risk. Class IV bears an unacceptable high risk of complications, and consensus suggests that pregnancy should be avoided.

For this study, prepregnancy patient characteristics that were included in statistical modelling were age, nulliparity, modified WHO class, NYHA class and signs of heart failure.

Socioeconomic data on patient level were not available. As a result, predefined socioeconomic factors were assigned to represent country characteristics and included HDI, Gini coefficient, health expenditure, schooling, gross domestic product per capita based on purchasing power parity (GDP), birth rate per 1000 and hospital beds per 1000. Definitions and sources of these characteristics are listed in online supplementary appendix 1. HDI is a combination of three factors: life expectancy from birth, mean years of schooling and the country standard of living. As these factors correlate with the other predefined country characteristics, the HDI was not included in further modelling. The HDI categories (low, medium, high, very high) were only used to categorise and understand the frequency of events within the different categories.

Endpoints

The following endpoints that occurred up to 1 week after delivery were studied: combined cardiac endpoint (including maternal cardiac death, arrhythmia requiring treatment, heart failure, thromboembolic event, aortic dissection, endocarditis, acute coronary syndrome, hospitalisation for cardiac reason or a cardiac intervention), heart failure, fetal or neonatal mortality (excluding miscarriage in the first trimester) and small for gestational age (SGA, birth weight <10th percentile). All-cause mortality data were also collected, but not used for statistical modelling due to low numbers. Heart failure was defined according to American College of Cardiology/American Heart Association guidelines [11] as a clinical syndrome that is characterised by specific symptoms (dyspnoea and fatigue) and signs (of fluid retention, such as oedema, rales) on the physical examination as judged by the treating cardiologist. The heart failure (HF) episode was only registered when signs or symptoms of HF were present, which required new treatment, change of treatment or hospital admission.

Statistical Analysis

Categorical variable differences were tested using χ^2 tests and are presented as percentages; in case of three categories, Pearson χ^2 tests were performed. Continuous variables are presented as mean and SD, or as median and first and third quartiles (Q1– Q3), as appropriate. Differences were tested using Student's t-tests; in case of three categories, one-way ANOVA tests were performed.

Generalised linear mixed models (GLMMs) were used as a result of the multilevel structure in the data. The ROPAC database consists of three levels: patients (level 1) were nested in centres (level 2), and centres were nested in countries (level 3). To account for differences in outcome between countries and between centres, random effects for country and centre were added to the

model. Patient and country characteristics were entered as fixed effects and those with a significant trend (P<0.10) in univariable analysis were assessed in multivariable analysis. Countries that included less than 10 patients were excluded from this study.

To determine the influence of fixed and random effects in our cohort, we further analysed the model for the combined cardiac endpoint. A conditional R^2 (for GLMM) was derived from the model before and after including the fixed effects (patient characteristics, followed by country characteristics). [12] This is an estimate of the percentage explained variance by the complete model (fixed and random effects). The random effect estimates of the individual countries for the combined cardiac endpoint were plotted with 95% CIs (caterpillar plot), unadjusted and adjusted for the fixed effects.

The rate of missing patient and country characteristics was relatively low, and therefore a complete case analysis approach was taken (96%). All analyses, except for multilevel modelling, were performed in SPSS V.21.0 (IBM, Armonk, New York, USA). Multilevel modelling was performed in R V.3.1, package lme4. [13]

Results

Patients

From January 2008 until April 2014, 2966 patients were included, from 99 centres in 39 countries. Nine countries enrolled less than 10 patients and were excluded. The remaining 30 countries enrolled 2924 patients from 89 centres. An overview of the countries is presented and arranged according to the HDI categories in Table 1. Socioeconomic indexes, including HDI, Gini coefficient, health expenditure, schooling, GDP, birth rate per 1000 and hospital beds per

1000, are presented for all countries (see online supplementary table S1).

Baseline characteristics are presented for patients per HDI category (Table 2). Maternal age at conception was higher in women from countries with a very high HDI, while these women were also more often nulliparous. Fewer women from countries with a medium or high HDI had a prior cardiac intervention and were in NYHA class I, compared with women from countries with a very high HDI. Indeed, signs of HF prior to pregnancy were more common; cardiac medication, mainly diuretics, were more commonly used before pregnancy by women from countries with a medium or high HDI compared with those from countries with a very high HDI. Valvular heart disease was much more common in women from countries with a medium HDI, while women from countries with a high or very high HDI more often had congenital heart disease.

Women with modified WHO class III or IV more often came from countries with a medium or high HDI, while women with a lower risk WHO class more often came from countries with a very high HDI.

Frequency of endpoints

Clinical event rates are presented for each HDI group (Figure 1) and for all countries separately (Table 3). A combined cardiac endpoint occurred in 645 women (22.1%), heart failure in 365 (12.5%), fetal/neonatal loss in 60 (2.1%) and small for gestational age in 270 (10.6%). Maternal mortality up to 1 week post partum occurred in 11 cases (0.9% medium HDI, 0.8% high HDI and 0.2% very high HDI, P=0.016) and was not included in the univariable or multivariable analysis.

Table 1: Human Development Index categories

	Low	Medium	High	Very High
Human Development Index*	<0.555	0.555-0.699	0.700-0.799	≥0.800
		(n=634)	(n=118)	(n=2130)
≥10 patients per country				
Countries in ROPAC	–	Egypt	Azerbaijan	Argentina
		South Africa	Russian Federation	Australia
				Austria
				Belgium
				Canada
				Czech Republic
				France
				Greece
				Germany
				Hungary
				Italy
				Japan
				Lithuania
				Israel
				Malta
				Netherlands
				Norway
				Poland
				Portugal
				Slovenia
				Spain
				Sweden
				Switzerland
				United Arab Emirates
				United Kingdom
				USA
<10 patients per country				
	–	–	Brazil	Ireland
			Bulgaria	
			Georgia	
			Macedonia	
			Romania	
			Serbia and Montenegro	
			Turkey	

*Human Development Index for women according to United Nations Development Report 2013. No value was available for Bosnia and Herzegovina (<10 inclusions).
ROPAC, Registry of Pregnancy and Cardiac disease.

Table 2: Baseline characteristics

	Total*		Low HDI	Medium HDI		High HDI		Very High HDI		
N (% of total inclusions)	2966	100%	0	634	22%	118	4%	2172	74%	P value
Mean age (SD)	29.3	± 5.6		27.7	± 5.9	26.4	± 5.3	29.9	± 5.4	<0.001
Nulliparous	1334	45%		160	25%	57	48%	1099	51%	<0.001
Pre-existent hypertension	188	6%		26	4%	18	16%	139	7%	<0.001
Current smoker	110	4%		11	2%	4	4%	95	5%	0.001
Pre-existent diabetes	46	2%		10	2%	1	0.8%	34	2%	1.000
Prior cardiac intervention	1585	54%		223	35%	44	37%	1304	60%	<0.001
NYHA functional class										<0.001
NYHA I	2154	74%		399	63%	48	42%	1686	79%	
NYHA II	659	23%		191	30%	62	54%	395	19%	
NYHA III	86	35%		42	7%	4	4%	39	2%	
NYHA IV	7	0.2%		2	0.3%	0	0%	5	0.2%	
Signs of HF before pregnancy	283	10%		138	22%	66	58%	74	4%	<0.001
AF before pregnancy	68	2%		47	7%	1	1%	20	1%	<0.001
Prior medication	824	28%		292	46%	17	17%	510	24%	<0.001
Beta-blocker	365	12%		75	12%	7	6%	280	13%	0.073
Antiarrythmic	90	3%		58	9%	3	3%	28	1%	<0.001
ACE inhibitor	116	4%		38	6%	9	8%	67	3%	0.001
Diuretic	170	6%		93	15%	7	6%	68	3%	<0.001
Cardiac diagnosis										<0.001
Congenital heart disease	1654	56%		88	14%	91	78%	1458	67%	
Valvular heart disease	942	32%		489	77%	15	13%	424	20%	
Ischaemic heart disease	47	2%		7	1%	0	0%	40	2%	
Cardiomyopathy	201	7%		45	7%	4	3%	151	7%	
Aortic pathology	101	3%		3	1%	6	5%	90	4%	
Pulmonary hypertension	13	0.4%		2	0.3%	2	2%	9	0.4%	
WHO classification										<0.001
WHO class I	583	20%		73	12%	27	23%	474	22%	
WHO class II	520	18%		18	3%	17	14%	481	22%	
WHO class II-III	932	32%		150	24%	34	29%	735	34%	
WHO class III	486	16%		187	30%	8	7%	286	13%	
WHO class IV	437	15%		206	33%	32	27%	196	9%	

Percentages are of total valid cases, excluding missing cases.
*Total cohort includes countries with less than 10 patients.
ACE, angiotensin-converting enzyme; AF, atrial fibrillation; HDI, Human Development Index; HF, heart failure; NYHA, New York Heart Association.

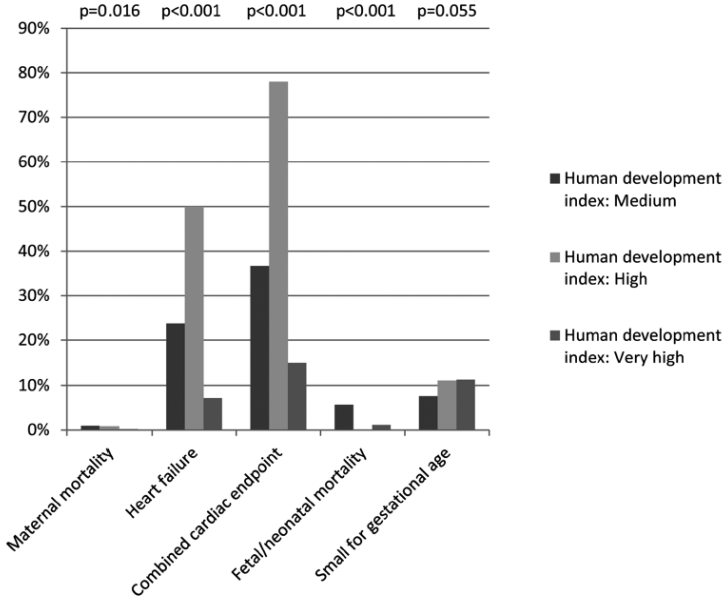


Figure 1: Event rate for Human Development Index categories.

Table 3: Events per country

		Maternal mortality (all cause)		Heart failure		Combined cardiac endpoint		Fetal/neonatal mortality (no miscarriage)		Small for gestational age	
	N	n	%	n	%	n	%	n	%	n	%
Argentina	10	0	0.0	0	0.0	0	0.0	0	0.0	1	10.0
Australia	19	0	0.0	2	10.5	4	21.1	0	0.0	2	10.5
Austria	83	0	0.0	1	1.2	4	4.8	1	1.2	1	1.2
Azerbaijan	10	0	0.0	2	20.0	2	20.0	0	0.0	0	0.0
Belgium	125	0	0.0	2	1.6	5	4.0	0	0.0	3	2.4
Canada	57	1	1.8	3	5.3	6	10.5	2	3.5	2	3.5
Czech Republic	14	0	0.0	0	0.0	0	0.0	0	0.0	1	7.1
Egypt	573	6	1.0	120	20.9	198	34.6	31	5.4	30	5.2
France	58	0	0.0	13	22.4	26	44.8	0	0.0	10	17.2
Germany	229	0	0.0	3	1.3	10	4.4	1	0.4	23	10.0
Greece	27	0	0.0	3	11.1	11	40.7	0	0.0	6	22.2
Hungary	44	0	0.0	0	0.0	1	2.3	1	2.3	4	9.1
Israel	61	0	0.0	19	31.1	25	41.0	1	1.6	7	11.5
Italy	238	1	0.4	12	5.0	33	13.9	3	1.3	28	11.8
Japan	33	0	0.0	2	6.1	2	6.1	0	0.0	6	18.2
Lithuania	60	0	0.0	5	8.3	5	8.3	1	1.7	8	13.3
Malta	19	0	0.0	0	0.0	1	5.3	0	0.0	2	10.5
The Netherlands	299	0	0.0	9	3.0	38	12.7	2	0.7	23	7.7
Norway	28	0	0.0	4	14.3	6	21.4	0	0.0	1	3.6
Poland	113	0	0.0	11	9.7	27	23.9	3	2.7	13	11.5
Portugal	13	0	0.0	0	0.0	0	0.0	1	7.7	0	0.0
Russian Federation	108	1	0.9	57	52.8	90	83.3	0	0.0	13	12.0
Slovenia	128	0	0.0	2	1.6	10	7.8	3	2.3	12	9.4
South Africa	61	0	0.0	30	49.2	34	55.7	5	8.2	8	13.1
Spain	221	1	0.5	20	9.0	32	14.5	3	1.4	29	13.1
Sweden	33	0	0.0	5	15.2	7	21.2	1	3.0	6	18.2
Switzerland	45	0	0.0	2	4.4	5	11.1	0	0.0	5	11.1
United Arab Emirates	31	0	0.0	13	41.9	16	51.6	0	0.0	5	16.1
UK	120	1	0.8	16	13.3	31	25.8	0	0.0	15	12.5
USA	64	0	0.0	9	14.1	16	25.0	1	1.6	6	9.4
Total	2924	11	0.4	365	12.5	645	22.1	60	2.1	270	9.2

Associations of patient and country characteristics with clinical endpoints

Univariable analysis of prepregnancy patient characteristics for the combined cardiac endpoint is shown in table 4. The only variable that was not significantly associated with the combined cardiac endpoint was nulliparity. Modified WHO II was not significantly different from modified WHO I. Of the country characteristics, Gini coefficient (P=0.017) and birth rate (although P=0.050) were independently associated with the combined cardiac endpoint, in addition to age, NYHA class, modified WHO class and signs of heart failure before pregnancy. The univariable and multivariable analyses of the remaining endpoints are shown in the online supplementary data. The results for HF as a separate endpoint were largely comparable to the results of the combined cardiac endpoint (see online supplementary table S2). While schooling, GDP, birth rate and number of hospital beds were associated with fetal/neonatal mortality in the univariable analysis, only GDP was independently associated with this endpoint (see online supplementary table S3). None of the country characteristics were associated with SGA, on top of NYHA II and III, and modified WHO class III and IV (see online supplementary table S4).

Influence of variability between countries and centres

The total explained variability of the model, the conditional R^2, for the combined cardiac endpoint including patient characteristics only was 37%. By adding the country characteristics, the R^2 increased by 4% to 41%. Without any of these fixed effects in the model, the conditional R^2 including random effects only was 33%. Figure 2 depicts the estimated unadjusted and adjusted ORs for a combined cardiac endpoint for each country compared with the average OR. Several countries do not include the 0 in their 95% CI in the unadjusted model.

Table 4: Univariable and multivariable analyses of patient and country characteristics with the combined cardiac endpoint.

	Univariable		Multivariable	
Variable	OR	95% CI	OR	95% CI
Age	1.026	1.008 - 1.045	1.020	1.000 - 1.039
Nulliparity	0.955	0.777 - 1.174		
NYHA I	NA		NA	
NYHA II	2.735	2.179 - 3.434	1.944	1.487 - 2.541
NYHA III	9.18	5.435 - 15.506	3.062	1.657 - 5.658
NYHA IV	26.01	2.634 - 256.826	7.456	0.792 - 70.209
WHO I	NA		NA	
WHO II	1.088	0.689 - 1.719	0.997	0.618 - 1.607
WHO II–III	2.261	1.575 - 3.246	1.992	1.371 - 2.895
WHO III	4.351	2.947 - 6.426	3.862	2.586 - 5.767
WHO IV	8.383	5.67 - 12.394	4.954	3.238 - 7.578
Signs of heart failure	4.165	3.037 - 5.711	1.708	1.167 - 2.502
Gini*	1.706	1.266 - 2.297	1.393	1.06 - 1.831
Health expenditure*	0.739	0.463 - 1.178		
Schooling*	0.965	0.468 - 1.991		
GDP*	0.737	0.453 - 1.200		
Birth rate*	2.896	1.742 - 4.815		
Hospital beds*	0.708	0.446 - 1.123	1.622	1.001 - 2.629

Data are clustered within hospitals within countries. The categorical variable NYHA classification and WHO are tested against the reference category I. WHO II is not significantly different from WHO I. The only variable that is not significant is nulliparity.
*Numerical data were standardised before analysis.
GDP, gross domestic product, NYHA, New York Heart Association.

However, when adjusted for patient and country characteristics, the 95% CIs of almost all countries do include 0. This means that for the vast majority of the countries, the need to account for random effects (patient within centre, within country) disappears when adjusting for patient and country characteristics.

Discussion

The ROPAC registry is the largest recorded cohort of pregnant women with cardiac disease. Women from many different countries were included. Results

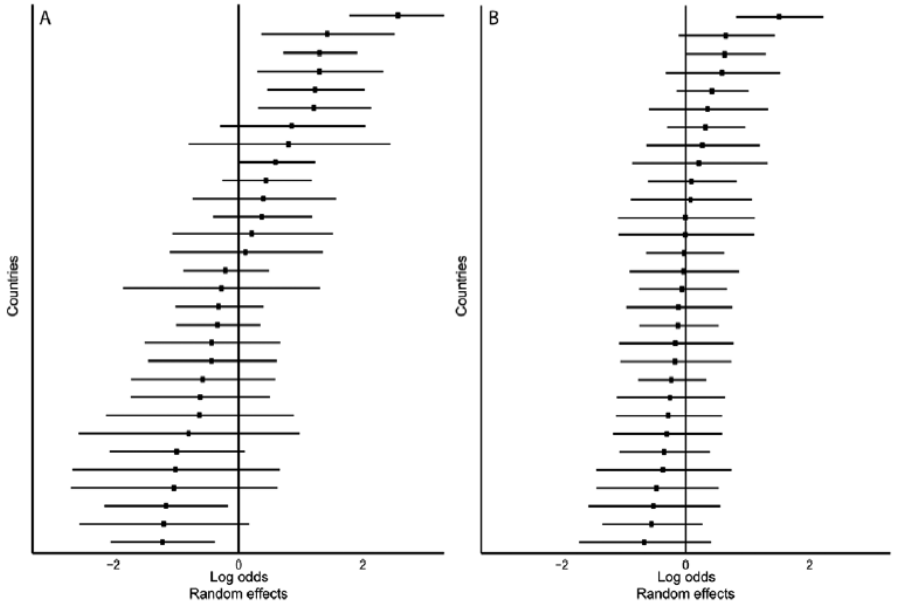


Figure 2: Between-country differences in outcome, unadjusted for fixed effects (A) and adjusted for fixed effects (B). Estimated unadjusted and adjusted ORs for a combined cardiac endpoint for each country compared with the average OR. Several countries do not include the 0 in their 95% CI in the unadjusted model. However, when adjusted for patient and country characteristics, the 95% CIs of almost all countries do include 0. This means that for the vast majority of the countries, the need to account for random effects (patient within centre, within country) disappears when adjusting for patient and country characteristics.

may be influenced by the multicentre and multinational nature of the registry. This study shows that indeed there are differences in outcome between centres and countries, but these differences are largely explained by differences in individual patient characteristics, such as NYHA classification, prior signs of heart failure and modified WHO classification. Only a few country characteristics had some impact: maternal cardiac event was associated with Gini coefficient and to a lesser extent with birth rate of the patients' residential country. Also, fetal outcome, such as SGA, was mainly associated with the maternal condition

and to a minor extent with country characteristics.

Maternal outcome and socioeconomic influences

Previous studies have shown that HDI is a strong predictor of maternal and fetal mortality rate in the global population. [14] Inequality of socioeconomic determinants within a country further increases the rate of maternal death. [15] A lower educational level and lower HDI have been reported to be associated with maternal adverse outcome. [16] Less educated women, for instance, have an increased risk of presenting to an emergency department in a severe condition. [17] This may be related to several issues: women from emerging countries tend to have a later presentation to a medical centre, which is probably associated with limited knowledge and awareness of risks and lack of money, and also to factors like a less well-developed infrastructure, longer travel time and perhaps less availability of skilled medical staff. To what extent these correlations can be extrapolated to women with pre-existent cardiac disease, and whether they need to be taken into account while analysing multinational registry data, has not been determined until now.

Although the number of maternal deaths was too low to allow for statistical analysis, the risk of a cardiac event (combined endpoint) was indeed associated with income inequality (expressed as the Gini coefficient) in a country. Also, a higher country birth rate correlated with a higher frequency of HF. These socioeconomic parameters need to be considered when interpreting data from registries; however, we feel that the number of factors actually showing a relationship to pregnancy outcome in these high-risk patients is actually relatively small compared with their impact in the general pregnant population. In fact, the most important determinant of pregnancy outcome was the underlying medical condition.

This cohort consists of a rather large subgroup of women with a cardiac condition considered modified WHO group 3 and 4. Category 4 involves women who should rather be advised to avoid pregnancy. However, in the end, the woman will decide herself whether she will proceed to try and get pregnant, and of course clinical care will not be denied to this group of women. Whether this group involves women who were not appropriately counselled about their risks following the latest guidelines may also be subject to further discussion. The fact that a greater part of women from less well-developed countries were in a higher modified WHO category (3 or 4) has undoubtedly influenced the outcome of our study. While the underlying disease is a given fact, availability of good preconception and perinatal and maternal care certainly deserves attention. It is part of the United Nations Millennium Goals, and this study emphasises the need for improvement of care.

Fetal outcome

With regard to fetal and obstetric outcome, previous reports showed that a higher income inequality (Gini coefficient) and educational level, rather than household income, seem to be associated with intrauterine growth but not with shorter gestational age at delivery. [18, 19] The exact underlying mechanism is difficult to determine. A recent large prospective cohort study of pregnant women showed that women from low socioeconomic subgroups have higher placental resistance indices, which may be explained by smoking. This association may contribute to a higher incidence of pregnancy complications and even stillbirth. [20, 21]

In our cohort of women with cardiac disease, country characteristics did not significantly influence the SGA rate, while maternal condition expressed as NYHA class and modified WHO classification did influence the frequency of SGA. In women with reduced cardiac function, an abnormal uteroplacental

flow is present, which is an important predictor of adverse obstetric and fetal outcome, [22] and this may explain the association in this study.

Research and clinical implications

The results imply that inter-regional differences need to be acknowledged also in research, but that the maternal condition seems to outweigh the influence of socioeconomic factors on reported cardiac and fetal outcome. A clear association between socioeconomic factors and events was present in univariable analysis, but it largely disappeared after correction for maternal condition. Thus, the higher event rate in emerging compared with advanced countries is mainly based on a worse prepregnancy condition of patients. Also, the need for multilevel modelling in this analysis was lost after adding the patient and country characteristics.

Data on cultural background were lacking, but would be very interesting to study. Differences in pregnancy outcome between emerging and advanced countries may be related to, for instance, religion. Women may have a strong feeling that their fate is predetermined and therefore less sensible to a doctor's advice. However, this hypothesis is rather philosophical and needs further investigation to determine whether this indeed influences pregnancy outcome.

Reducing adverse pregnancy outcome in any region, but particularly in remote areas, is an important goal as formulated by WHO. While this goal resulted in major declines in maternal death rates globally, this trend has definitely not been observed in maternal death due to cardiac disease. [23] Creating awareness in young women with cardiac disease about the potential high risks of pregnancy should be part of standard care and preferably initiated at a young age. The fifth millennium goal of the WHO is reduction of maternal mortality by means of increasing the number of women receiving at least four antenatal care visits and the number of births attended by skilled staff. [24]

An increase in the number of women receiving this level of care and a decline in maternal death rate has been observed in the past 10–15 years, but about 50% of women still do not receive the recommended minimum of four antenatal visits. Also, a well-developed infrastructure for cardiovascular health screening is warranted to ensure early diagnosis and management. [25] Improvements in these medical resources may also reduce the burden of adverse events in pregnant women with cardiac disease.

Other global observational studies, for instance those dealing with factors influencing secondary cardiovascular prevention, did find related socioeconomic factors. One study pointed out that the country-level socioeconomic factors explained two-thirds of the variation in preventive drug use compared with only a third explained by individual factors (such as smoking, gender and education). [26] Although these results are not in line with our findings, this knowledge needs to be appreciated for our population as well; it does show the between-country differences in (level of) healthcare availability.

Limitations

While ROPAC provides a unique view on global pregnancy outcome, including women from 39 countries, the current distribution of countries was within a range of medium to very high HDI. However, the range of country-specific characteristics was sufficient to illustrate the differences between more developed countries and those with poorer resources. Including patients from countries categorised with a low HDI may strengthen this study, but it is hard to achieve with limited availability of organised/specialised medical care in these countries. In previous studies, ethnicity was shown to influence maternal outcome. [27] In particular, non-Hispanic black women seem to have an increased risk of pregnancy-related mortality. ROPAC did not include demographic socioeco-

nomic data at a patient level, which is why we performed the analysis at a country level. If the socioeconomic data (income, education, social status and employment, among others) were available at patient level, it may have been possible to find stronger relationships. Since we performed the statistical analyses at three levels (patient, within centre, within country), we believe that meaningful conclusions can be drawn from our data. In future registries, it would be desirable to collect more socioeconomic data on a patient level.

The majority of the participating centres were university or tertiary centres (86%). Unfortunately, only 75% responded to the question whether they were a university, community or private clinic, which is why we did not include this information in the statistical analysis. However, it is likely that our data are derived from women cared for in larger centres with a specialised department for pregnancy.

ROPAC included 6 months' follow-up post partum. However, due to large differences in follow-up availability between countries, it was decided not to include these results to this analysis. Follow-up at 1 week was available in all patients. For future research, inclusion of long-term follow-up would be favourable. Finally, the number of pregnancies complicated by fetal and neonatal mortality was relatively low, which hampered statistical modelling, and conclusions should be interpreted carefully. This study aimed to comment on associations, rather than causal relations. It should be interpreted as a hypothesis-generating study and may be a starting point for future research studying, for instance, socioeconomic factors on a patient level.

Conclusion

Socioeconomic factors were partly explainable for differences in pregnancy outcome in women with cardiac disease, but the main denominator was the

individual's condition, at least in countries with a medium to very high HDI. Raising awareness and improving access to medical resources as advocated by WHO will help to improve the outcome for pregnant women, hopefully also for women with heart disease.

References

1. Kassebaum, N. J. *et al.* Global, regional, and national levels and causes of maternal mortality during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384,** 980–1004 (2014).
2. Collaborators, G. B. D. M. M. Global, regional, and national levels of maternal mortality, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388,** 1775–1812 (2016).
3. Roos-Hesselink, J. W. *et al.* Outcome of pregnancy in patients with structural or ischaemic heart disease: results of a registry of the European Society of Cardiology. *European Heart Journal* **34,** 657–665 (2013).
4. Van Hagen, I. M. *et al.* Pregnancy in Women With a Mechanical Heart Valve: Data of the European Society of Cardiology Registry of Pregnancy and Cardiac Disease (ROPAC). *Circulation* **132,** 132–42 (2015).
5. Van Hagen, I. M. *et al.* Global cardiac risk assessment in the Registry Of Pregnancy And Cardiac disease: results of a registry from the European Society of Cardiology. *Eur J Heart Fail* **18,** 523–33 (2016).
6. Sliwa, K. *et al.* Spectrum of cardiac disease in maternity in a low-resource cohort in South Africa. *Heart* **100,** 1967–74 (2014).
7. Soma-Pillay, P., Seabe, J. & Sliwa, K. The importance of cardiovascular pathology contributing to maternal death: Confidential Enquiry into Maternal Deaths in South Africa, 2011-2013. *Cardiovasc J Afr* **27,** 60–5 (2016).
8. Van Hagen, I. M. *et al.* Incidence and predictors of obstetric and fetal complications in women with structural heart disease. *Heart* **103,** 1610–1618 (2017).
9. Thorne, S. *et al.* Pregnancy and contraception in heart disease and pulmonary arterial hypertension. *J Fam Plann Reprod Health Care* **32,** 75–81 (2006).
10. European Society of, G. *et al.* ESC Guidelines on the management of cardiovascular diseases during pregnancy: the Task Force on the Management of Cardiovascular Diseases during Pregnancy of the European Society of Cardiology (ESC). *Eur Heart J* **32,** 3147–97 (2011).
11. Jessup, M. *et al.* 2009 focused update: ACCF/AHA Guidelines for the Diagnosis and Management of Heart Failure in Adults: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines: developed in collaboration with the International Society for Heart and Lung Transplantation. *Circulation* **119,** 1977–2016 (2009).

12. Nakagawa, S. & Schielzeth, H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* **4,** 133–142 (2013).

13. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67,** 1–48 (2015).

14. Lee, K. S., Park, S. C., Khoshnood, B., Hsieh, H. L. & Mittendorf, R. Human development index as a predictor of infant and maternal mortality rates. *Journal of Pediatrics* **131,** 430–433 (1997).

15. Ruiz, J. I. *et al.* Inequality as a Powerful Predictor of Infant and Maternal Mortality around the World. *Plos One* **10** (2015).

16. Sheldon, W. R. *et al.* Postpartum haemorrhage management, risks, and maternal outcomes: findings from the World Health Organization Multicountry Survey on Maternal and Newborn Health. *Bjog-an International Journal of Obstetrics and Gynaecology* **121,** 5–13 (2014).

17. Tuncalp, O. *et al.* Education and severe maternal outcomes in developing countries: a multicountry cross- sectional survey. *Bjog-an International Journal of Obstetrics and Gynaecology* **121,** 57–65 (2014).

18. Fujiwara, T., Ito, J. & Kawachi, I. Income Inequality, Parental Socioeconomic Status, and Birth Outcomes in Japan. *American Journal of Epidemiology* **177,** 1042–1052 (2013).

19. Nkansah-Amankra, S., Dhawain, A., Hussey, J. R. & Luchok, K. J. Maternal Social Support and Neighborhood Income Inequality as Predictors of Low Birth Weight and Preterm Birth Outcome Disparities: Analysis of South Carolina Pregnancy Risk Assessment and Monitoring System Survey, 2000-2003. *Maternal and Child Health Journal* **14,** 774–785 (2010).

20. Bouthoorn, S. H. *et al.* Socioeconomic inequalities in placental vascular resistance: a prospective cohort study. *Fertility and Sterility* **101,** 1367–U37 (2014).

21. Singh, T., Leslie, K., Bhide, A., D'Antonio, F. & Thilaganathan, B. Role of Second-Trimester Uterine Artery Doppler in Assessing Stillbirth Risk. *Obstetrics and Gynecology* **119,** 256–261 (2012).

22. Pieper, P. G. *et al.* Uteroplacental Blood Flow, Cardiac Function, and Pregnancy Outcome in Women With Congenital Heart Disease. *Circulation* **128,** 2478–2487 (2013).

23. Cantwell, R. *et al.* Saving Mothers' Lives: Reviewing maternal deaths to make motherhood safer: 2006-2008. The Eighth Report of the Confidential Enquiries into Maternal Deaths in the United Kingdom. *BJOG* **118 Suppl 1,** 1–203 (2011).

24. Nations, U. *The Millennium development goals report 2015* Report (2015).

25. Sliwa, K., Acquah, L., Gersh, B. J. & Mocumbi, A. O. Impact of Socioeconomic Status, Ethnicity, and Urbanization on Risk Factor Profiles of Cardiovascular Disease in Africa. *Circulation* **133,** 1199–1208 (2016).

26. Yusuf, S. *et al.* Use of secondary prevention drugs for cardiovascular disease in the community in high-income, middle-income, and low-income countries (the PURE Study): a prospective epidemiological survey. *Lancet* **378,** 1231–1243 (2011).

27. Creanga, A. A. *et al.* Race, Ethnicity, and Nativity Differentials in Pregnancy-Related Mortality in the United States 1993-2006. *Obstetrics and Gynecology* **120,** 261–268 (2012).

Supplemental Material

S1. Supplemental Tables

Supplemental Table 1: Socioeconomic indices per country

	n	Human Development Index (scale 0-1)	Gini Coefficient (scale 0-1)	Health Expenditure (% of GDP)	Schooling (years)	Gross Domestic Product per captita (U.S. dollar)	Birth rate (n/1000)	Hospital beds (n/1000)
Argentina	10	0.806	0.43	8.1	10.0	na	18	4.5
Australia	19	0.92	0.33	9.0	12.5	45477	14	3.9
Austria	83	0.834	0.28	10.6	8.9	47416	9	7.6
Azerbaijan	10	0.723	0.34	5.2	10.5	17142	19	4.6
Belgium	125	0.866	0.27	10.6	10.5	43059	12	6.5
Bosnia Herzegovina*	2	na	0.36	10.2	7.2	10116	9	3.5
Brazil*	4	0.752	0.53	8.9	7.3	15788	15	2.3
Bulgaria*	6	0.775	0.36	7.3	10.6	16574	10	6.4
Canada	56	0.894	0.32	11.2	12.3	44318	11	2.7
Czech republic	14	0.844	0.26	7.4	12.1	30044	10	6.8
Egypt	573	0.617	0.31	4.9	5.3	10383	28	1.7
France	58	0.878	0.31	11.6	10.9	39210	13	6.4
Georgia*	4	0.713	0.41	9.9	11.9	8508	14	2.9
Germany	229	0.892	0.29	11.1	12.6	44185	8	8.2
Greece	27	0.833	0.34	10.8	9.9	26753	10	4.8
Hungary	44	0.816	0.29	7.7	11.2	42715	9	7.2
Ireland*	8	0.881	0.3	9.4	11.7	47560	16	2.9
Israel	61	0.878	0.37	7.7	12.6	33382	21	3.4
Italy	238	0.853	0.33	9.5	9.7	35762	9	3.4
Japan	33	0.863	0.34	9.3	11.2	36698	8	13.7
Lithuania	60	0.847	0.35	6.6	12.3	26511	10	7.0
Macedonia*	5	0.708	0.44	6.6	7.9	12752	11	4.5
Malta	19	0.807	0.18	8.7	9.5	29526	10	4.4
Netherlands	299	0.899	0.28	12	11.6	47955	11	4.7
Norway	28	0.94	0.25	9.1	12.7	66766	12	3.3
Poland	113	0.836	0.3	6.7	11.9	24494	10	6.5
Portugal	13	0.808	0.34	10.4	8.0	27930	9	3.4
Romania*	2	0.771	0.27	5.8	10.4	19577	10	6.1
Russian Federation	108	0.793	0.42	6.2	11.7	22523	13	9.7
Serbia & Montenegro*	4	0.757	0.3	10.4	9.2	13668	9	5.4
Slovenia	128	0.876	0.25	9.1	11.8	29098	11	4.6
South Africa	62	0.658	0.63	8.5	9.8	12867	21	2.8
Spain	221	0.862	0.34	9.4	9.5	32842	10	3.1
Sweden	33	0.898	0.27	9.4	11.8	45067	12	2.7
Switzerland	45	0.895	0.29	10.9	11.5	59351	10	5.0
Turkey*	7	0.704	0.4	6.7	6.4	19351	17	2.5
United Arab Emirates	31	0.801	0.575	3.3	10.2	64112	16	1.9
UK	120	0.887	0.35	9.3	12.8	39111	13	2.9
USA	64	0.911	0.39	17.9	13.0	52980	13	2.9

na = notavailable
* Countries with n<10, excluded from analyses

Supplemental Table 2: Univariable and multivariable analyses of patient and country characteristics with heart failure

	Univariable		Multivariable	
Variable	Estimate	p-value	Estimate	p-value
Age	0.024	0.027	0.016	0.188
Nulliparity	-0.069	0.595		
NYHA I	NA		NA	
NYHA II	0.888	<0.001	0.305	<0.001
NYHA III	2.516	<0.001	1.211	<0.001
NYHA IV	3.771	0.001	2.264	0.048
WHO I	NA		NA	
WHO II	0.162	0.589	0.127	0.689
WHO II–III	0.913	<0.001	0.818	0.001
WHO III	1.142	<0.001	1.064	<0.001
WHO IV	2.219	<0.001	1.577	<0.001
Signs of heart failure	1.712	<0.001	1.018	<0.001
Gini	0.667	<0.001	0.464	<0.001
Health expenditure	-0.453	0.058		
Schooling	-0.144	0.709		
GDP	-0.338	0.197		
Birth rate	1.244	<0.001	0.509	0.001
Hospital beds	0.424	0.083		

Data are clustered within hospitals within countries. The categorical variable NYHA classification and WHO are tested against the reference category I.
GDP, gross domestic product, NYHA, New York Heart Association.

Supplemental Table 3: Univariable and multivariable analyses of patient and country characteristics with fetal/neonatal mortality

	Univariable		Multivariable	
Variable	Estimate	p-value	Estimate	p-value
Age	0.006	0.799		
Nulliparity	0.047	0.868		
NYHA I	NA		NA	
NYHA II	-0.284	0.400	-0.417	0.222
NYHA III	0.078	0.900	-0.132	0.846
NYHA IV	2.090	0.075	1.943	0.106
WHO I	NA		NA	
WHO II	-0.243	0.741	0.189	0.806
WHO II–III	0.739	0.152	0.984	0.080
WHO III	1.243	0.017	1.404	0.013
WHO IV	1.079	0.044	1.142	0.062
Signs of heart failure	0.087	0.822		
Gini	0.230	0.150		
Health expenditure	-0.376	0.164		
Schooling	-0.676	<0.001	0.325	0.265
GDP	-0.873	<0.001	-1.244	0.001
Birth rate	0.697	<0.001	-0.361	0.266
Hospital beds	-0.773	0.005	-0.608	0.057

Data are clustered within hospitals within countries. The categorical variable NYHA classification and WHO are tested against the reference category I.
GDP, gross domestic product, NYHA, New York Heart Association.

Supplemental Table 4: Univariable and multivariable analyses of patient and country characteristics with small-for-gestational-age

	Univariable		Multivariable	
Variable	Estimate	p-value	Estimate	p-value
Age	0.005	0.657		
Nulliparity	-0.069	0.602		
NYHA I	NA		NA	
NYHA II	0.522	0.001	0.426	0.012
NYHA III	1.127	0.001	0.882	0.023
NYHA IV	1.680	0.06	1.340	0.154
WHO I	NA		NA	
WHO II	0.094	0.689	0.136	0.567
WHO II–III	0.270	0.187	0.288	0.168
WHO III	0.710	0.002	0.678	0.004
WHO IV	0.952	<0.001	0.024	0.294
Signs of heart failure	0.642	0.004	0.091	0.928
Gini	0.135	0.090		
Health expenditure	-0.063	0.523		
Schooling	0.069	0.563		
GDP	-0.074	0.494		
Birth rate	-0.060	0.606	0.509	0.001
Hospital beds	0.030	0.749		

Data are clustered within hospitals within countries. The categorical variable NYHA classification and WHO are tested against the reference category I.
GDP, gross domestic product, NYHA, New York Heart Association.

S2. Definitions

Data for country characteristics were selected for 2013 where possible. If not or scarcely available, another year was chosen based on availability.

Birth rate per 1,000	"Crude birth rate indicates the number of live births occurring during the year, per 1,000 population estimated at midyear. Subtracting the crude death rate from the crude birth rate provides the rate of natural increase, which is equal to the rate of population change in the absence of migration." Source: http://data.worldbank.org/indicator/SP.DYN.CBRT.IN/countries The majority of data were selected for 2011.
GDP per capita, PPP ($)	"GDP per capita based on purchasing power parity (PPP). PPP GDP is gross domestic product converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GDP as the U.S. dollar has in the United States. GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current international dollars based on the 2011 ICP round." Source: http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD The majority of data were selected for 2013.
Gini index (or coefficient)	"The Gini Index measures the extent to which the distribution of income or consumption expenditure among individuals or households within an economy deviates from a perfectly equal distribution. A Gini Index of 0 represents perfect equality, an index of 100 implies perfect inequality. The Gini index is equal to the Gini coefficient multiplied by 100."

	Sources: http://www.oecd.org/social/income-distribution-database.htm http://data.worldbank.org/indicator/SI.POV.GINI http://hdr.undp.org/en/content/income-gini-coefficient The majority of data were selected for 2012.
Health expenditure	"Total health expenditure is the sum of public and private health expenditure. It covers the provision of health services (preventive and curative), family planning activities, nutrition activities, and emergency aid designated for health but does not include provision of water and sanitation." Source: http://hdr.undp.org/en/content/expenditure-health-total-gdp The majority of data were selected for 2011.
Hospital beds per 1,000	"Hospital beds include inpatient beds available in public, private, general, and specialized hospitals and rehabilitation centers. In most cases beds for both acute and chronic care are included." Source: http://data.worldbank.org/indicator/SH.MED.BEDS.ZS The majority of data were selected for 2011.

Human development index	"The HDI was created to emphasize that people and their capabilities should be the ultimate criteria for assessing the development of a country, not economic growth alone. It measures the average achievements in a country in three basic dimensions of human development: a long and healthy life, access to knowledge and a decent standard of living. The HDI is the geometric mean of normalized indices measuring achievements in each dimension. The HDI is the geometric mean of the three dimension indices and embodies imperfect substitutability across all HDI dimensions." (female HDI: Very High - 0.874; High - 0.71; Medium - 0.565; Low - 0.446) Source: http://hdr.undp.org/en/content/hdi-female The majority of data were selected for 2013.
Schooling	"Mean years of schooling: average number of years of education received by people ages 25 and older, converted from education attainment levels using official durations of each level." (very high human development: 11.7 years; low human development: 4.2 years) Source: http://hdr.undp.org/en/content/mean-years-schooling-femalesaged-25-years-and-above-years The majority of data were selected for 2013.

Part IV: Longitudinal modelling in practice

Chapter 6

Prognostic value of serial galectin-3 measurements in patients with acute heart failure

van Vark LC, Lesman-Leegte I, **Baart SJ**, Postmus D, Pinto YM, de Boer RA, Asselbergs FW, Boersma H, Hillege HL, Akkerhuis KM, for the TRIUMPH investigators

J Am Heart Assoc. 2017; 6(12)

Abstract

Background: Several clinical studies have evaluated the association between galectin-3 levels and outcome in patients with heart failure (HF). However, little is known about the predictive value of repeated galectin-3 measurements. This study evaluates the prognostic value of repeated time-dependent galectin-3 measurements in acute HF patients.

Methods and Results: In the TRIUMPH (Translational Initiative on Unique and Novel Strategies for Management of Patients with Heart Failure) clinical cohort study, 496 acute HF patients were enrolled in 14 hospitals in The Netherlands, between 2009 and 2014. Repeated blood samples (7) were drawn during 1-year follow-up. Associations between repeated biomarker measurements and the primary end point were assessed using a joint model. Median age was 74 years and 37% were women. The primary end point, composite of all-cause mortality and HF rehospitalization, was reached in 188 patients (40%), during a median follow-up of 325 days (interquartile range 85–401). The median baseline galectin-3 level was 24 ng/mL (interquartile range 18–34). The mean number of galectin-3 measurements available per patient was 4.3. After adjustment for clinical factors and N-terminal pro-brain natriuretic peptide, there was a weak association between baseline galectin-3 and risk of the primary end point. When repeated measurements were taken into account, the adjusted hazard ratio per 1 SD increase of the galectin-3 level (on the log2 scale) at any time point increased to 1.67 (95% confidence interval, 1.24–2.23, P<0.001). After additional adjustment for repeated N-terminal pro-brain natriuretic peptide measurements, the association remained statistically significant.

Conclusions: Repeated galectin-3 measurements appeared to be a strong predictor of outcome in acute HF patients, independent of N-terminal pro-brain natriuretic peptide. Hence, galectin-3 may be helpful in clinical practice for prognostication and treatment monitoring.

Introduction

Most studies on serum biomarkers in heart failure (HF) populations conducted so far have related adverse outcome during follow-up with a single measurement at baseline. [1–3] Although this approach has demonstrated the prognostic value of a variety of biomarkers, among which are the well-known natriuretic peptides, [4] it does not explore the biological variation within patients with evolving disease. In fact, HF is a highly variable, heterogeneous, and progressive condition. [5] Thus, repeated biomarker measurements may be required to more accurately reflect this dynamic and progressive nature of the underlying pathophysiologic processes, such as mechanical overload, atherosclerosis, inflammation, and cardiac fibrosis. Therefore, we expect that risk models that account for repeated measurements may more adequately reflect the current status of the patient compared with models that only use single measurements. The TRIUMPH (Translational Initiative on Unique and Novel Strategies for Management of Patients with Heart Failure) study was designed to identify and validate novel biomarkers to improve prognostication in HF. [6] TRIUMPH was designed as a translational study program, combining biological discovery of novel biomarkers, technologic advances, and clinical validation in patients presenting with acute HF. In the clinical validation study, both the novel and established HF biomarkers were evaluated for their prognostic properties using a unique design of 7 planned repeated measurements during 1-year follow-up. Based on previous clinical and epidemiological studies, galectin-3 was earmarked as a biomarker with high potential for improving prognostication. Galectin-3 is a member of a large family of β-galactosidebinding animal lectins. [7] Galectin-3 expression has been detected in macrophages, neutrophils, eosinophils, and mast cells. In response to a variety of mechanical and neurohormonal stimuli, macrophages secrete galectin-3. [8] Galectin-3 stimulates additional macrophages, pericytes, myofibroblasts,

and fibroblasts, which are all involved in the initiation and progression of tissue scarring. Consequently, galectin-3 appears to be involved in cardiac fibrosis. In addition, galectin-3 plays an important role in the inflammatory response, which is an important step in the process of cardiac remodeling. [9–11] Galectin-3 is expressed in numerous tissues such as heart, kidney, lung, uterus, and colon. [12] The level of galectin-3 expression is relatively low in heart tissue under normal conditions, but may increase substantially under pathophysiological circumstances. [13] Several clinical studies have evaluated the prognostic value of galectin-3. Higher levels of galectin-3 have been associated with an increased risk of incident HF and all-cause mortality in the general population. [14, 15] Furthermore, single galectin-3 levels have shown to be an independent risk factor of mortality in both stable and acute HF patients, although it still remains uncertain whether galectin-3 confers independent prognostic information when added to N-terminal pro-brain natriuretic peptide (NT-proBNP). [2, 3, 16–18] A few studies have been performed to assess the prognostic value of galectin-3 when measured multiple times. The change in galectin-3 level over time was predictive of outcome. [19–21] However, given the dynamic and progressive nature of HF, the number of galectin-3 measurements needed for adequate estimation of the true galectin-3 level is expected to be high. Therefore, in the present study, we assessed the independent association between the estimated instantaneous galectin-3 level, using frequently measured galectin-3 levels, and the incidence of all-cause mortality and HF readmission during 1-year follow-up in the 496 patients with acute HF who compose the TRIUMPH clinical cohort.

Methods

Objective and Study Design

TRIUMPH was designed as a translational bench-to-bedside study program encompassing the entire spectrum of biomarker discovery to clinical validation. [6] The clinical validation study was an observational prospective study enrolling patients admitted with acute HF in 14 hospitals in The Netherlands, between September 2009 and December 2013. This cohort study was designed to validate the clinical value of biomarkers successfully passing the bio-informatics and early-validation stages of TRIUMPH, and to further evaluate more established biomarkers of HF. There was a particular interest in the change in biomarker levels over time. The study was approved by the medical ethics committee at all participating centers.

Patient Selection

Patients $\geq$18 years of age were eligible for enrollment if they were hospitalized with decompensation of known chronic HF or newly diagnosed HF. Furthermore, 3 other criteria had to be met: (1) natriuretic peptide levels had to be elevated to $\geq$3 times the upper limit of normal, (2) there had to be evidence of sustained systolic or diastolic left ventricular dysfunction, and (3) patients had to be treated with intravenous diuretics. Patients with HF precipitated by a noncardiac condition, by severe valvular dysfunction without sustained left ventricular dysfunction, or by an acute ST-segment elevation myocardial infarction were excluded. Furthermore, patients scheduled for a coronary revascularization procedure, on a waiting list for a heart transplantation, with severe renal failure for which dialysis was needed, or with a coexistent condition with a life expectancy $<$1 year could not participate. All study participants provided

written informed consent.

Patient Management

Patient management was at the discretion of the treating physician, and in accordance with the guidelines of the European Society of Cardiology. [22] Importantly, the biomarker data that were generated in the context of this observational study were not used for treatment decisions.

Study Procedures

During hospitalization, blood samples were obtained at admission (day 1), once during days 2 to 4 and, subsequently, on the day of discharge. Afterwards, repeated blood samples were also obtained at outpatient follow-up visits, which were planned at 2 to 4 weeks, 3 months, 6 months, and 9 to 12 months after discharge. The baseline blood sample was defined as the first sample obtained after inclusion, up to a maximum of 2 days after inclusion. At each visit, HF symptoms were assessed using the New York Heart Association classification. Medication use was determined at discharge using 3 categories: (1) use of an angiotensin-converting enzyme inhibitor or an angiotensin II receptor antagonist or both, (2) use of a β-blocker, and (3) use of diuretics. Patients underwent physical examination and systematic measurements of weight, blood pressure, and heart rate.

Blood Collection

Nonfasting blood samples were obtained by venipuncture and transported to the clinical chemistry laboratory of each participating hospital for further processing according to a standardized protocol. The collected material was centrifuged at 1700g/relative centrifugal force, after which citrate-, EDTA- , heparin-, and

trasylol-plasma was separated, as well as blood serum. Buffy coats were collected from EDTA tubes to enable analysis of genetic factors. Dimethylsulfoxide was added to an additional EDTA tube for cryopreservation of blood cells. All blood aliquots were subsequently stored at a temperature of -80°C within 2 hours after venipuncture.

Galectin-3 Measurements

Serum and heparin-plasma were transported under controlled conditions to a central laboratory (Future Diagnostics Solutions B.V.) for batch analysis of galectin-3 and NTproBNP levels. Galectin-3 concentrations were determined in serum, using the BGM Galectin-3 Test as instructed by the manufacturer (BG Medicine, Inc, Waltham, MA). NTproBNP concentrations were determined in heparin plasma using the Elecsys NT-proBNP assay on a Cobas 8000 analyzer (Roche Diagnostics Limited, Rotkreuz, Switzerland). Analysts were blinded for patient characteristics and end points.

End Points

Information on vital status and hospital readmissions was obtained until at least 9 months with a maximum of 400 days after the index hospitalization. We approached the civil registry, screened all medical records, and asked patients for information during their follow-up visits.

The primary end point is the composite of all-cause mortality and readmission for HF. Readmission for HF was defined as an unplanned rehospitalization because of decompensation of HF, with at least 2 of the following 3 criteria being present: elevated natriuretic peptide levels $\geq$3 times the upper limit of normal, symptoms of cardiac decompensation (rales, edema, or elevated central venous pressure), and treatment with intravenous diuretics. Secondary end points

included the individual components of the primary end point and cardiovascular mortality. An event adjudication committee, blinded for biomarker information, was established for reviewing and adjudication of end points.

Statistical Analysis

The distributions of continuous variables, including biomarker levels, were evaluated for normality by visual examination of the histogram and Kolmogorov–Smirnov tests. Variables with a normal distribution are presented as mean± SD, whereas the median and interquartile range (IQR) are presented in case of non-normality. Categorical variables are presented as counts and percentages. Galectin-3 and NT-proBNP levels had a non-normal distribution and were therefore log-transformed for further analysis.

Patients were classified according to the quartiles of the galectin-3 distribution, and differences in baseline characteristics between these quartiles were evaluated by χ^2 tests (categorical variables), analysis of variance, or Kruskal–Wallis tests, as appropriate.

We applied Cox proportional hazards models to evaluate the association of baseline galectin-3 levels with the study end points. Subjects were censored at the time of occurrence of the end point under investigation, death, and at the scheduled end of follow-up. No deviations of the proportional hazards assumption were found by inspecting log minus log plots of the survival functions. We performed univariate analyses to obtain the crude estimates of the effect of baseline galectin-3 level (model 1), analyses that were adjusted for age and sex only (model 2), and analyses that were additionally adjusted for systolic blood pressure, diabetes mellitus, left ventricular ejection fraction, previous hospitalization for HF during the past 6 months, ischemic HF, body mass index, estimated glomerular filtration rate, and baseline NT-proBNP level (model 3).

The results are presented as adjusted hazard ratios (HR) per 1 SD increase of the biomarker level (on the log2 scale) with 95% confidence intervals (CI). We calculated the estimated glomerular filtration rate using the Modification of Diet in Renal Disease equation. [23]

Joint models were fitted to assess the association between estimated instantaneous biomarker levels, calculated using the repeated biomarker levels, and the specified study end points. A joint model combines a mixed-effects linear regression model for the serial measurements with a Cox proportional hazards model for the risk of the specified study end points. [24] We used cubic splines, with knots set at 1 week and 1 month after initial hospitalization. For the analyses with the repeated galectin-3 measurements, we used similar univariate and multivariate models as mentioned above (models 1, 2, and 3), except for model 3 in which we added medication use at discharge to the mixed-effects linear regression model. We also tested whether the instantaneous slope of the galectin-3 trajectories itself, when added to model 3, was an independent predictor. Finally, we combined the repeated measurements of galectin-3 and NT-proBNP to assess their respective independent prognostic value. Taking into account the limitations of the R packages for Joint Modeling, we were able to combine the estimated galectin-3 trajectory (using a mixed-effects linear regression model) and the estimated NT-proBNP trajectory (using a time-dependent Cox proportional hazards model) in 1 joint model. Since the model did not converge when we adjusted for all the covariates in model 3, baseline systolic blood pressure had to be left out in this final model (model 4). Diagnostics and sensitivity analyses were performed to evaluate the joint models. To account for the correlation structure between serial biomarker measurements collected from the same patient, we obtained the SD from the total variance of a random intercepts linear mixed model fitted on the post-discharge data. The

final results are presented as adjusted HR per 1 SD increase of the biomarker level (on the log2 scale) at any point in time with 95% CI. The TRIUMPH sample size was chosen to achieve a power of 80% (1-β=0.8) to detect an odds ratio of at least 2.0 (α=0.05, 2-sided test) for a biomarker value above the 75% percentile of its distribution comparing end point cases with non-cases. The incidence of the primary end point was initially estimated at 25% to 30%, based on observations in historical HF populations. Then, 780 patients are required. During the course of the study, based on evolving evidence, the estimated incidence was adjusted to 30% to 35%, and the sample size was eventually determined at 490 patients. TRIUMPH enrolled 496 patients, and 40% reached the primary end point. Data on covariates were complete in 93% of patients, except for left ventricular ejection fraction, which was complete in 78%. Single imputation was applied to account for missing values of covariates. Data are imputed using predictive mean matching for continuous variables, logistic regression for binary variables, and polytomous regression for unordered categorical data. Baseline covariates used in the full model and survival information were used in the imputation. The software used was R package MICE (https://cran.r-project.org/web/packages/mice/mice.pdf). A sensitivity analyses was performed on the full model for the primary end point on the complete cases.

Results

Patients

A total of 496 patients were enrolled in the TRIUMPH clinical cohort. Three patients withdrew their informed consent. Eighteen patients were withdrawn from statistical analyses because of inclusion violation. These patients had

Table 1: Baseline Parameters According to Overall Sample in Study Population (N=475)

Variables	Overall Sample
Demographic characteristics, median (IQR) or %	
Age, y	74 (65–80)
Female	37
White	95
Measurements at baseline, median (IQR) or %	
Body mass index, kg/m2	28 (25–31)
Systolic blood pressure, mm Hg	125 (110–147)
Diastolic blood pressure, mm Hg	74 (65–85)
Heart rate, bpm	85 (72–100)
eGFR, mL/min per 1.73 m2	46 (34–62)
Left ventricular ejection fraction, %	30 (21–41)
NYHA classification	
II	17
III	55
IV	27
Medical history, %	
Newly diagnosed heart failure	36
Heart failure with reduced ejection fraction	83
Previous heart failure admission within 6 mo	20
Ischemic heart failure	49
Myocardial infarction	40
Hypertension	51
Atrial fibrillation	42
Diabetes mellitus	36
Stroke	17
Medication use at discharge, %	
ACE-I and/or ARB	78
β-Blocker	78
Diuretics	93
Biomarkers, median (IQR)	
Galectin-3, ng/mL	24 (18–34)
NT-proBNP, pg/mL	4152 (2089–9387)

ACE-I indicates angiotensin-converting enzyme inhibitor; ARB, angiotensin II receptor antagonist; bpm, beats per minute; eGFR, estimated glomerular filtration rate; IQR, interquartile range; NT-proBNP, N-terminal pro-brain natriuretic peptide; NYHA, New York Heart Association.

no evidence of sustained systolic or diastolic left ventricular dysfunction on echocardiography. Accordingly, 475 patients compose the analysis set. Their median age was 74 years (IQR 65–80) and 37% were women (Table 1). Me-

Table 2: Baseline Parameters According to Quartiles of Galectin-3 Level

Variables	Quartile 1	Quartile 2	Quartile 3	Quartile 4	P value*
Demographic characteristics, median (IQR) or %					
Age, y	70	73	76	75	0.010
Female	45	31	33	38	0.13
White	92	93	97	96	0.27
Measurements at baseline, median (IQR) or %					
Body mass index, kg/m2	27	27	29	29	0.035
Systolic blood pressure, mm Hg	130	125	125	122	0.29
Diastolic blood pressure, mm Hg	80	73	74	70	<0.001
Heart rate, bpm	94	85	84	80	0.002
eGFR, mL/min per 1.73 m2	63	55	42	32	<0.001
Left ventricular ejection fraction, %	30	30	34	31	0.020
NYHA classification					
II	23	18	11	14	0.12
III	50	51	63	60	
IV	27	28	25	26	
Medical history, %					
Newly diagnosed heart failure	57	40	26	21	<0.001
Heart failure with reduced ejection fraction	88	88	76	81	0.080
Previous heart failure admission within 6 mo	8	17	24	29	<0.001
Ischemic heart failure	40	45	55	56	0.036
Myocardial infarction	28	32	54	48	<0.001
Hypertension	40	50	56	60	0.016
Atrial fibrillation	32	44	45	46	0.089
Diabetes mellitus	20	32	41	50	<0.001
Stroke	14	14	17	22	0.29
Biomarkers, median					
NT-proBNP, pg/mL	3180	3970	4372	7544	<0.001
Galectin-3, ng/mL	16	21	28	40	<0.001

bpm indicates beats per minute; eGFR, estimated glomerular filtration rate; IQR, interquartile range; NT-proBNP, N-terminal pro-brain natriuretic peptide; NYHA, New York Heart Assocation.
*P value for differences between groups.

dian systolic blood pressure was 125 mm Hg (IQR 110–147) and median left ventricular ejection fraction was 30% (IQR 21–41). At discharge 78% used an angiotensin-converting enzyme inhibitor or an angiotensin II receptor antagonist or both, 78% used a β-blocker, and 93% used diuretics. Median baseline galectin-3 level was 24 ng/mL (IQR 18–34) and NTproBNP was 4152 pg/mL (IQR 2089–9387). Table 2 shows the baseline characteristics of patients in different quartiles of galectin-3 level. Patients in quartiles with a higher galectin-3 level were older and had a worse kidney function. In the higher galectin-3 quartiles, more patients had a history of myocardial infarction and diabetes mellitus, had ischemic HF, and had been admitted to the hospital for HF during the past 6 months. In the lower galectin-3 quartiles, more patients had newly diagnosed HF during the initial hospitalization.

Baseline Galectin-3 Levels and the Incidence of Study End Points

During the median follow-up of 325 days (IQR 85–401), 188 patients (40%) reached the primary composite end point of all-cause death (n=113) or readmission for HF (n=123). This corresponds with an incidence rate of 55.9 per 100 patient-years for the primary end point. In the highest quartile of baseline galectin-3, 65 patients (59%) reached the primary end point compared with 27 patients (24%) in the lowest quartile. The number of events in the highest quartile compared with the lowest quartile of galectin-3 was also higher for all-cause mortality (n=44 [40%] and n=14 [13%], respectively) and readmission for HF (n=44 [40%] and n=19 [17%], respectively).

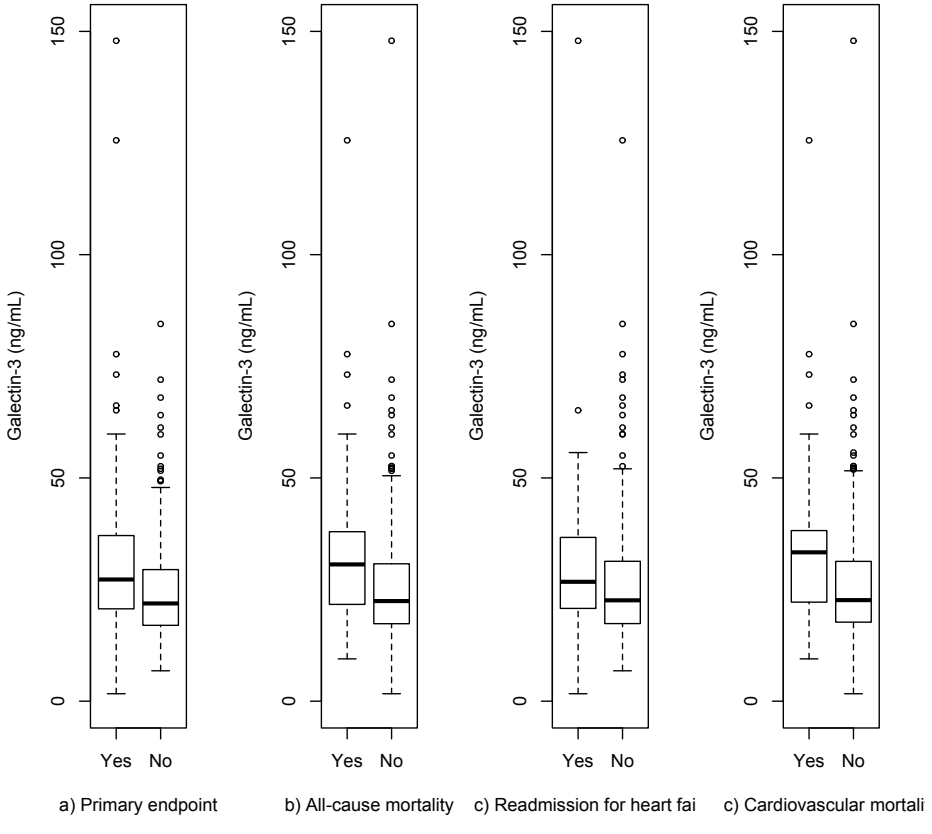


Figure 1: Distributions of baseline galectin-3 levels within the subpopulations of patients who had an event and those who did not experience an event for: (A) the primary end point; (B) the single end point of all-cause mortality; (C) the single end point of readmission for heart failure; and (D) the single end point of cardiovascular mortality.

Repeatedly Measured Galectin-3 Levels and the Incidence of Study End Points

On average, galectin-3 was available 4.3 times during follow-up. The mean galectin-3 level during follow-up was 23.8 ng/mL; an increase of 1 SD galectin-3

Table 3: Hazard Ratios for Different End Points Per 1 SD Increase of the Baseline Galectin-3 Level (on the log2 Scale).

	M - SD* 15.9	M* 24.7	M + SD* 38.2	HR† (95% CI)	P-value
Primary end point					
Model 1				1.50 (1.30-1.75)	<0.001
Model 2				1.49 (1.28-1.73)	<0.001
Model 3				1.12 (0.93-1.36)	0.241
Number of events/patients: 188/475					
All-cause mortality					
Model 1				1.54 (1.29-1.85)	<0.001
Model 2				1.52 (1.26-1.83)	<0.001
Model 3				1.26 (1.01-1.59)	0.044
Number of events/patients: 113/475					
HF hospitalization					
Model 1				1.47 (1.22-1.76)	<0.001
Model 2				1.47 (1.23-1.76)	<0.001
Model 3				1.05 (0.82-1.33)	0.720
Number of events/patients: 123/475					
Cardiovascular mortality					
Model 1				1.60 (1.28-1.99)	<0.001
Model 2				1.57 (1.26-1.97)	<0.001
Model 3				1.24 (0.93-1.67)	0.147
Number of events/patients: 77/475					

Model 1 unadjusted; model 2 adjusted for age and sex; model 3 adjusted for age, sex, systolic blood pressure, diabetes mellitus, LVEF, previous hospitalization for heart failure during the past 6 mo, ischemic heart failure, body mass index, eGFR, and baseline NT-proBNP.
CI indicates confidence interval; eGFR, estimated glomerular filtration rate; HF, heart failure; HR, hazard ratio; LVEF, left ventricular ejection fraction; M, mean; NT-proBNP, N-terminal pro-brain natriuretic peptide.
*Mean ± 1 SD of the patient-specific geometric mean galectin-3 value at baseline (presented on the linear scale).
†Hazard ratios are related to a 1-SD increase of galectin-3 (on the log scale) at baseline.

level on the log2 scale from the mean was 13 ng/mL. A decrease of 1 SD galectin-3 level on the log2 scale was 8 ng/mL. After adjustment for age and sex (model 2), the HR per SD increase of the galectin-3 level (on the log2 scale) at any point in time was 2.09 (95% CI, 1.71–2.56) for the primary end point. After adjustment for the broader range of potential confounders including medication

use at discharge and baseline NT-proBNP level (model 3), the association remained highly statistically significant with a HR of 1.67 (95% CI, 1.24–2.23) (Table 4). Results were similar for the secondary end points. The instantaneous slope of the galectin-3 level trajectories itself was not an independent predictor of the primary end point.

Table 4: Hazard Ratios for Different End Points Per 1 SD Increase of the Galectin-3 Level (on the log2 Scale) at Any Point in Time, Using a Joint Model

	Mean Value*			Instantaneous Level†	
	M - SD	M	M + SD	HR (95% CI)	P-value
Primary end point	15.4	23.8	36.6		
Model 1				2.07 (1.71-2.53)	<0.001
Model 2				2.09 (1.71-2.56)	<0.001
Model 3				1.67 (1.24-2.23)	<0.001
All-cause mortality	15.4	23.8	36.9		
Model 1				2.41 (1.83-3.15)	<0.001
Model 2				2.36 (1.78-3.08)	<0.001
Model 3				2.14 (1.47-3.16)	<0.001
HF hospitalization	15.4	23.8	36.6		
Model 1				1.87 (1.47-2.39)	<0.001
Model 2				1.92 (1.48-2.46)	<0.001
Model 3				1.41 (1.02-1.93)	0.035
Cardiovascular mortality	15.4	23.8	36.9		
Model 1				2.46 (1.79-3.34)	<0.001
Model 2				2.43 (1.76-3.35)	<0.001
Model 3				2.22 (1.48-3.36)	<0.001

Model 1 unadjusted; model 2 adjusted for age and sex; model 3 adjusted for age, sex, systolic blood pressure, diabetes mellitus, LVEF, previous hospitalization for HF during the past 6 mo, ischemic HF, body mass index, eGFR, medication use at hospital discharge (ACE-I and/or ARB, β-blocker, and diuretics) and baseline NT-proBNP level.

ACE-I indicates angiotensin converting enzyme inhibitor; ARB, angiotensin II receptor antagonist; CI, confidence interval; eGFR, estimated glomerular filtration rate; HF, heart failure; HR, hazard ratio; LVEF, left ventricular ejection fraction; M, mean; NT-proBNP, N-terminal pro-brain natriuretic peptide.

*Mean ± 1 SD of the patient-specific geometric mean galectin-3 value during follow-up (presented on the linear scale).

†Hazard ratios are related to a 1-SD increase of galectin-3 (on the log scale) at any point in time.

After adjustment for repeated NT-proBNP measurements (model 4), the

association between repeated galectin-3 levels and adverse outcome remained statistically significant with a HR of 1.54 (95% CI, 1.16–2.05) for the primary end point corresponding with 1 SD increase of galectin-3 level (on the log2 scale) at any point in time (Table 5). The HR corresponding with a 1-SD increase of NT-proBNP level (on the log2 scale) at any point in time was 2.10 (95% CI, 1.63–2.74) after adjustment for repeated galectin-3 levels.

Table 5: Hazard Ratios for Different End Points Per 1-SD Increase of Galectin-3 Level or NT-proBNP Level (on the log2 Scale) at Any Point in Time Using Repeated Galectin-3 and NT-proBNP Measurements in a Joint Model

	Mean Value*			Instantaneous Level†	
	M - SD	M	M + SD	HR (95% CI)	P-value
Primary end point					
Galectin-3	15.4	23.8	36.6	1.54 (1.16-2.05)	0.003
NT-proBNP	742	2445	8062	2.10 (1.63-2.74)	<0.001
All-cause mortality					
Galectin-3	15.4	23.8	36.9	1.77 (1.22-2.52)	<0.001
NT-proBNP	739	2480	8321	2.68 (1.90-3.86)	<0.001
HF hospitalization					
Galectin-3	15.4	23.8	36.6	1.29 (0.92-1.81)	0.160
NT-proBNP	742	2445	8062	1.71 (1.27-2.25)	<0.001
Cardiovascular mortality					
Galectin-3	15.4	23.8	36.9	1.89 (1.25-2.85)	0.002
NT-proBNP	739	2480	8321	2.62 (1.70-4.27)	<0.001

Model 4 adjusted for age, sex, diabetes mellitus, LVEF, previous hospitalization for heart failure during the past 6 mo, ischemic heart failure, body mass index, eGFR, medication use at hospital discharge (ACE-I and/or ARB, β-blocker, and diuretics) and baseline NTproBNP level.
ACE-I indicates angiotensinconverting enzyme inhibitor; ARB, angiotensin II receptor antagonist; CI, confidence interval; eGFR, estimated glomerular filtration rate; HF, heart failure; HR, hazard ratio; LVEF, left ventricular ejection fraction; M, mean; NT-proBNP, N-terminal pro-brain natriuretic peptide.
*Mean $\pm$ 1 SD of the patient-specific geometric mean biomarker level during follow-up (presented on the linear scale).
†Hazard ratios are related to a 1-SD increase of biomarker level (on the log scale) at any point in time.

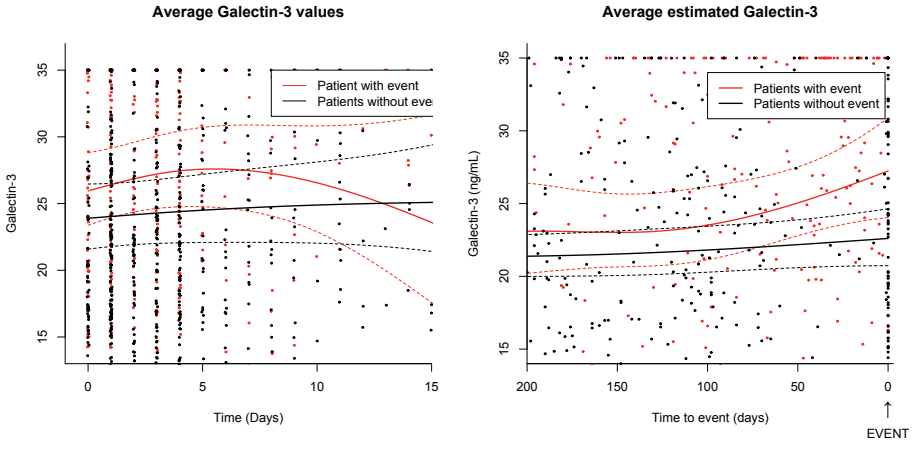
Figure 2A shows the average estimated galectin-3 level in patients with and without the primary end point according to model 3 and the individual galectin-

3 measurements. During hospitalization the average galectin-3 level remains steady for patients who remained free of the primary end point. For patients who reached the primary end point during follow-up, the average estimated galectin-3 level decreased slightly after the initial hospitalization. Apparently, throughout follow-up, patients who reached the primary end point had, on average, higher levels than their counterparts who remained free of the primary end point. Furthermore, the average estimated galectin-3 levels appeared to elevate several weeks before the time of the primary end point (Figure 2B).

Discussion

This study clearly demonstrates that, in patients admitted with acute HF, repeated galectin-3 measurements are a strong and independent predictor of the composite end point of all-cause mortality or readmission for HF during 1-year follow-up. Our results illustrate that repeated measurements of galectin-3 offer incremental prognostic value to (repeatedly measured) NT-proBNP, which is considered the criterion standard biomarker in HF patients.

Our observation that baseline galectin-3 level was associated with mortality confirms earlier findings both in acute and stable HF patients. [2, 3, 25, 26] Similar to previous studies, the association between baseline galectin-3 level and mortality attenuated after adjustment for established risk factors, including kidney function and NT-proBNP level. [16, 17, 27] The association between baseline galectin-3 level and readmission for HF was less apparent. However, the decision to hospitalize a patient for decompensation of HF may be influenced by several subjective patient- and physician-related factors that are unlikely to have an association with the galectin-3 level. Furthermore, several risk factors such as kidney function, diabetes mellitus, and NT-proBNP level influence this decision and are related to galectin-3. Therefore, the association between



(a) (b)

Figure 2: A, Average estimated galectin-3 pattern during initial hospitalization for decompensated heart failure for patients with and without the primary end point. The figure includes the individual galectin-3 measurements for patients with and without the primary end point. B, Average estimated galectin-3 pattern before the primary end point or end of follow-up for patients with and without the primary end point. The figure includes the individual galectin-3 measurements for patients with and without the primary end point. The average estimated galectin-3 levels are adjusted for age, sex, systolic blood pressure, diabetes mellitus, LVEF, previous hospitalization for heart failure during the past 6 months, ischemic heart failure, body mass index, eGFR, medication use at hospital discharge (ACE-I and/or ARB, β-blocker, and diuretics), and baseline NT-proBNP (model 3). ACE-I indicates angiotensinconverting enzyme inhibitor; ARB, angiotensin II receptor antagonist; eGFR, estimated glomerular filtration rate; LVEF, left ventricular ejection fraction; NT-proBNP, N-terminal pro-brain natriuretic peptide.

baseline galectin-3 level and HF readmission attenuated after adjustment for these risk factors. Since the primary end point is a composite of all-cause mortality and readmission for HF, the relationship between the galectin-3 level and the mortality end points per se are stronger compared with the primary end point.

Repeated galectin-3 measurements were strongly and independently related to the primary end point, as well as its separate components. Repeated measure-

ments take into account the dynamic and continuous change in galectin-3 level over time, which better reflects the true nature of the underlying pathophysiology in HF. In this study, the number of galectin-3 measurements per patient was high and therefore the repeated galectin-3 measurements could be used to estimate instantaneous galectin-3 levels (ie, the estimated galectin-3 level at any point in time during the follow-up period). When compared with baseline galectin-3 levels, the estimated instantaneous galectin-3 levels identified patients at an even higher risk for reaching an end point. The estimated instantaneous galectin-3 level more accurately approximates the true galectin-3 level and therefore reflects the actual condition of the patient at that point in time during follow-up. This is expected to be important since HF is a dynamic and often progressive disease in which inflammation, cardiac fibrosis, and remodeling are ongoing processes that cannot be captured in a single biomarker assessment at 1 point in time. [5] Furthermore, baseline galectin-3 measurements were all taken during hospitalization for galectin-3, in contrast to natriuretic peptides, does not respond to volume overload and unloading directly, which occurs during hospitalization. [28] As galectin-3 is involved in the process of myocardial fibrosis, it is more likely that galectin-3 is of more prognostic value when patients enter a more chronic phase of HF. [11]

Interestingly, the slope of the galectin-3 trajectory did not add prognostic information to the estimated instantaneous galectin-3 level. An explanation could be that galectin-3 is helpful in identifying high-risk patients when their galectin-3 level rises above a certain threshold. The change in galectin-3 level before reaching this threshold is not essential for risk stratification. However, to be able to estimate whether a patient's galectin-3 level rises above the threshold, repeated measurements are required. A few studies have been conducted on the prognostic value of multiple galectin-3 measurements in acute and stable

HF patients. [19, 20] These studies showed that change in galectin-3 level is associated with mortality. A possible explanation as to why in the present study slope of the galectin-3 trajectory did not add further prognostic information might be that the number of galectin-3 measurements during follow-up was substantially higher in our study, which allowed us to estimate an instantaneous slope of the galectin-3 trajectory, rather than the slope of the difference ("delta") between the level at baseline and that at a fixed point in time.

The statistical method (Joint Model) used to estimate the trajectory of the galectin-3 level takes into account the continuous changes in biomarker levels and adequately analyzes the relation between these biomarker trajectories and different end points considering the changing population because of censoring at the time of occurrence of an end point. Previous studies presented changes in biomarker level as a "delta" between just 2 measurements that are separated in time. If >2 samples are taken into account, patients have often been categorized according to the number of high or low biomarker levels. Obviously, both approaches do not fully capture the true biomarker pattern of the dynamic disease. Additionally, the power to predict adverse outcome is reduced.

Galectin-3 measurements conferred additional and independent prognostic information to that offered by baseline as well as repeated NT-proBNP measurements. The fact that NT-proBNP and galectin-3 reflect different underlying pathophysiological processes in HF may be the most important reason for this observation. Galectin-3 is a marker of cardiac fibrosis, inflammation, and remodeling, whereas NT-proBNP is a marker of volume overload. [13, 29] As such, galectin-3 might be a marker that more directly reflects the pathophysiological processes that lead to adverse cardiac remodeling and deterioration of cardiac function, whereas NT-proBNP reflects the volume overload resulting from the actual (left) ventricular dysfunction. In this way, the galectin-3 and NT-proBNP

level provide complementary information on the pathophysiological state, as well as with respect to the assessment of prognosis. With respect to prognostication in HF, the results of the present study, therefore, not only provide evidence for the use of repeated galectin-3 measurements, but also for the combined use with (repeatedly measured) NT-proBNP.

Although this study is a large multicenter prospective observational study, it seems that the studied population is not completely representative for the average HF population. The mean age in our study population is 74 years and the women are underrepresented. Moreover, only 18% of the included HF patients have a preserved ejection fraction. De Boer et al. [30] showed that galectin-3 levels did not differ between HF patients with a reduced and preserved ejection fraction and the predictive value of galectin-3 was stronger in patients with a preserved ejection fraction. By underrepresenting the HF patients with a preserved ejection fraction in our study, we possibly underestimated the prognostic value of galectin-3.

Future studies should evaluate the value of repeated galectin-3 measurements when used to guide treatment decisions. It may be hypothesized that treatment is to be intensified in patients with high galectin-3 levels or unfavorable galectin-3 patterns. On the other hand, repeated galectin-3 measurements might be helpful to identify patients who are more likely to respond to certain treatments. [31] Furthermore, it remains to be addressed whether galectin-3 may be targeted by specific antigalectin-3 therapies. Additional studies should also determine the number of galectin-3 measurements needed for optimal prognostication and therapy monitoring. The frequency by which galectin-3 levels should be measured may not be identical for each patient, but depends on the clinical condition of the patient, the treatment given, the galectin-3 level, and the progression of galectin-3 levels during follow-up.

Conclusion

The TRIUMPH study clearly demonstrates that repeated measurements of galectin-3 are a strong and independent predictor of adverse outcome in patients following admission for acute HF. The estimated instantaneous galectin-3 level identified patients at a higher risk of reaching adverse events than baseline galectin-3 levels alone. In addition, repeated galectin-3 measurements offer incremental prognostic value to that conferred by other known risk factors and, importantly, repeated measurements of NT-proBNP. These results suggest that repeated galectin-3 measurements in addition to NT-proBNP measurements may be helpful in clinical practice to identify HF patients who are at increased risk of adverse outcome.

References

1. Lok, D. J. *et al.* Prognostic value of galectin-3, a novel marker of fibrosis, in patients with chronic heart failure: data from the DEAL-HF study. *Clin Res Cardiol* **99,** 323–8 (2010).
2. Ueland, T. *et al.* Galectin-3 in heart failure: high levels are associated with all-cause mortality. *Int J Cardiol* **150,** 361–4 (2011).
3. Van Kimmenade, R. R. *et al.* Utility of amino-terminal pro-brain natriuretic peptide, galectin-3, and apelin for the evaluation of patients with acute heart failure. *J Am Coll Cardiol* **48,** 1217–24 (2006).
4. Januzzi J. L., J. *et al.* Utility of amino-terminal pro-brain natriuretic peptide testing for prediction of 1-year mortality in patients with dyspnea treated in the emergency department. *Arch Intern Med* **166,** 315–20 (2006).
5. Braunwald, E. Heart failure. *JACC Heart Fail* **1,** 1–20 (2013).
6. *TRIAL ID: NTR1893. TRanslational Initiative on Unique and novel strategies for Management of Patients with Heart failure (TRIUMPH)*
7. Barondes, S. H. *et al.* Galectins: a family of animal beta-galactoside-binding lectins. *Cell* **76,** 597–8 (1994).
8. Barondes, S. H., Cooper, D. N., Gitt, M. A. & Leffler, H. Galectins. Structure and function of a large family of animal lectins. *J Biol Chem* **269,** 20807–10 (1994).
9. Henderson, N. C. & Sethi, T. The regulation of inflammation by galectin-3. *Immunol Rev* **230,** 160–71 (2009).

10. Frangogiannis, N. G. The immune system and cardiac repair. *Pharmacol Res* **58,** 88–111 (2008).

11. Liu, Y. H. *et al.* N-acetyl-seryl-aspartyl-lysyl-proline prevents cardiac remodeling and dysfunction induced by galectin-3, a mammalian adhesion/growth-regulatory lectin. *Am J Physiol Heart Circ Physiol* **296,** H404–12 (2009).

12. Hughes, R. C. The galectin family of mammalian carbohydrate-binding molecules. *Biochem Soc Trans* **25,** 1194–8 (1997).

13. De Boer, R. A., Yu, L. & van Veldhuisen, D. J. Galectin-3 in cardiac remodeling and heart failure. *Curr Heart Fail Rep* **7,** 1–8 (2010).

14. Ho, J. E. *et al.* Galectin-3, a marker of cardiac fibrosis, predicts incident heart failure in the community. *J Am Coll Cardiol* **60,** 1249–56 (2012).

15. De Boer, R. A. *et al.* The fibrosis marker galectin-3 and outcome in the general population. *J Intern Med* **272,** 55–64 (2012).

16. Felker, G. M. *et al.* Galectin-3 in ambulatory patients with heart failure: results from the HF-ACTION study. *Circ Heart Fail* **5,** 72–8 (2012).

17. Bayes-Genis, A. *et al.* Head-to-head comparison of 2 myocardial fibrosis biomarkers for long-term heart failure risk stratification: ST2 versus galectin-3. *J Am Coll Cardiol* **63,** 158–66 (2014).

18. Chen, A., Hou, W., Zhang, Y., Chen, Y. & He, B. Prognostic value of serum galectin-3 in patients with heart failure: a meta-analysis. *Int J Cardiol* **182,** 168–70 (2015).

19. Anand, I. S. *et al.* Baseline and serial measurements of galectin-3 in patients with heart failure: relationship to prognosis and effect of treatment with valsartan in the Val-HeFT. *Eur J Heart Fail* **15,** 511–8 (2013).

20. Van der Velde, A. R. *et al.* Prognostic value of changes in galectin-3 levels over time in patients with heart failure: data from CORONA and COACH. *Circ Heart Fail* **6,** 219–26 (2013).

21. Motiwala, S. R. *et al.* Serial measurement of galectin-3 in patients with chronic heart failure: results from the ProBNP Outpatient Tailored Chronic Heart Failure Therapy (PROTECT) study. *Eur J Heart Fail* **15,** 1157–63 (2013).

22. McMurray, J. J. *et al.* ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. *Eur Heart J* **33,** 1787–847 (2012).

23. Stevens, L. A., Coresh, J., Greene, T. & Levey, A. S. Assessing kidney function–measured and estimated glomerular filtration rate. *N Engl J Med* **354,** 2473–83 (2006).

24. Rizopoulos, D. The R Package JMbayes for Fitting Joint Models for Longitudinal and Time-to-Event Data using MCMC. http://arxiv.org/abs/1404.7625 (2014).

25. Lok, D. J. *et al.* Incremental prognostic power of novel biomarkers (growth-differentiation factor-15, high-sensitivity C-reactive protein, galectin-3, and high-sensitivity troponin-T) in patients with advanced chronic heart failure. *Am J Cardiol* **112,** 831–7 (2013).

26. Zhang, Y. *et al.* The utility of galectin-3 for predicting cause-specific death in hospitalized patients with heart failure. *J Card Fail* **21,** 51–9 (2015).

27. Gullestad, L. *et al.* The predictive value of galectin-3 for mortality and cardiovascular events in the Controlled Rosuvastatin Multinational Trial in Heart Failure (CORONA). *Am Heart J* **164,** 878–83 (2012).

28. Milting, H. *et al.* Plasma biomarkers of myocardial fibrosis and remodeling in terminal heart failure patients supported by mechanical circulatory support devices. *J Heart Lung Transplant* **27,** 589–96 (2008).

29. Kim, H. N. & Januzzi J. L., J. Natriuretic peptide testing in heart failure. *Circulation* **123,** 2015–9 (2011).

30. De Boer, R. A. *et al.* Predictive value of plasma galectin-3 levels in heart failure with reduced and preserved ejection fraction. *Ann Med* **43,** 60–8 (2011).

31. De Boer, R. A. *et al.* Galectin-3: A Modifiable Risk Factor in Heart Failure. *Cardiovasc Drugs Ther* (2014).

Supplemental Material

S1. TRIUMPH Investigators

List of Site Investigators: Hans L. Hillege, MD, PhD (UMCG Groningen, The Netherlands); Ivonne Lesman-Leegte, PhD (UMCG Groningen, The Netherlands); Michiel J. Nagelsmit, MD, PhD (Treant Zorggroep Scheperziekenhuis Emmen, The Netherlands); Elly M.C.J. Wajon, MD (Medisch Spectrum Twente Enschede, The Netherlands); Hans-Peter Brunner-La Rocca, MD, PhD (Maastricht UMC, The Netherlands); Gerard C.M. Linssen, MD, PhD (Twenteborg ziekenhuis Almelo, The Netherlands); Willem F. Terpstra, MD, PhD (Slingeland Ziekenhuis Doetinchem, The Netherlands); K. Martijn Akkerhuis, MD, PhD (Erasmus MC Rotterdam, The Netherlands); Laura C. van Vark, MD (Erasmus MC Rotterdam, The Netherlands); Eric Boersma, PhD (Erasmus MC Rotterdam, The Netherlands); Anho H. Liem, MD, PhD (Admiraal de Ruyter Ziekenhuis Goes, The Netherlands); Alexander J. Wardeh, MD, PhD (Haaglanden MC Den Haag, The Netherlands); Addy J.M. van Miltenburg, MD, PhD (Fransiscus Gasthuis en Vlietland Rotterdam, The Netherlands); Ewout J. van den Bos, MD, PhD (Albert Schweitzer Ziekenhuis Dordrecht, The Netherlands); Stijn P.J. de Ridder (St. Anna Ziekenhuis Geldrop, The Netherlands); Folkert W. Asselbergs, MD, PhD (UMC Utrecht, The Netherlands); Yigal M. Pinto, MD, PhD (AMC Amsterdam, The Netherlands).

Chapter 7

Prognostic value of serial ST2 measurements in patients with acute heart failure

van Vark LC, Lesman-Leegte I, **Baart SJ**, Postmus D, Pinto YM, de Boer RA, Asselbergs FW, Boersma H, Hillege, HL, Akkerhuis, KM, for the TRIUMPH investigators

J Am Coll Cardiol 2017; 70:2378–88

Abstract

Background: Several clinical studies have evaluated the association between ST2 and outcome in patients with heart failure (HF). However, little is known about the predictive value of frequently measured ST2 levels in patients with acute HF.

Objectives: This study sought to describe the prognostic value of baseline and repeated ST2 measurements in patients with acute HF.

Methods: In the TRIUMPH (Translational Initiative on Unique and novel strategies for Management of Patients with Heart failure) clinical cohort study, 496 patients with acute HF were enrolled in 14 hospitals in the Netherlands between 2009 and 2014. Repeated blood samples (7) were drawn during 1-year follow-up. ST2 and N-terminal pro–B-type natriuretic peptide (NT-proBNP) levels were measured in a central laboratory. The primary endpoint was the composite of all-cause mortality and HF rehospitalization. Associations between repeated biomarker measurements and the primary endpoint were assessed using a joint model.

Results: Median age was 74 years, and 37% of patients were women. The primary endpoint was reached in 188 patients (40%) during a median follow-up of 325 days (interquartile range: 85 to 401). The median baseline ST2 level was 71 ng/ml (interquartile range: 46 to 102). After adjustment for clinical factors and NT-proBNP, baseline ST2 was associated with an increased risk of the primary endpoint, and the hazard ratio per 1 SD increase of the baseline ST2 level (on the log2 scale) was 1.30 (95%confidence interval: 1.08 to 1.56; p = 0.005). When repeated measurements were taken into account, the adjusted hazard ratio per 1 SD increase of the ST2 level (on the log2 scale) during follow-up increased to 1.85 (95% confidence interval: 1.02 to 3.33; p = 0.044), adjusted for clinical factors and repeated measurements of NT-proBNP. Furthermore, ST2 levels appeared to elevate several weeks before the time of the primary endpoint.

Conclusions: Repeated ST2 measurements appeared to be a strong predictor of outcome in patients with acute HF, independent of repeatedly measured NT-proBNP. Hence ST2 may be helpful in clinical practice for prognostication and treatment monitoring.

(TRanslational Initiative on Unique and novel strategies for Management of Patients with Heart failure [TRIUMPH]; NTR1893)

Introduction

Heart failure (HF) is a major cause of morbidity and mortality in the Western World. [1] Improvements in treatment and patient management are needed because most patients with HF die despite evidence-based treatment. Serum biomarkers may play an important role in bridging the gap between the assessment of HF and the occurrence of adverse outcomes, and they may expose novel, potentially modifiable disease pathways.

Most studies on the prognostic value of biomarkers of HF conducted so far have related adverse outcome during follow-up with a single measurement at baseline. [2–4] This approach does not explore the biological variation that exists within patients with a highly variable, heterogeneous, and progressive condition such as HF. [5] Thus, repeated biomarker measurements may be required to reflect more accurately the dynamic and progressive nature of the underlying pathophysiological processes, such as mechanical overload, cardiac fibrosis, and inflammation, and therefore may be more suitable for prognostication and therapy monitoring.

ST2 is an interleukin-1 (IL-1) receptor family member with membrane-bound (ST2L) and soluble (sST2) isoforms. An IL-1–related protein, called IL-33, was identified as a functional ligand for ST2L. [6] IL-33/ST2L signaling protects the myocardium against hypertrophy and cardiac fibrosis following pressure overload. [7] Soluble ST2, which is the form measured by current assays, acts as a decoy receptor for IL-33 and prevents the IL-33/ST2L interaction and the subsequent cardioprotective cascade of events. The major source of ST2 is currently not fully established. For a long time, the source of circulating sST2 in cardiac disease was presumed to be myocardial, following in vitro data that sST2 has been shown to be secreted by cardiomyocytes when the cells are subjected to biomechanical overload. [8] Accordingly, serum ST2 levels correlate

strongly with serum levels of natriuretic peptides. [9] More recent work, however, suggests that in human cardiac disease, the vascular endothelial cells may be the predominant source of sST2, rather than the human myocardium. [10]

In clinical studies, single ST2 levels have shown to be a risk factor for mortality in patients with both stable and acute HF, independent of N-terminal pro–B-type (NT-proBNP). [2, 11, 12] A recent meta-analysis supports the use of ST2 in patients with stable chronic HF for risk stratification. [12] Furthermore, several studies have evaluated the prognostic value of multiple ST2 measurements. [9, 13–15] It is known that ST2 levels in patients with acute HF are significantly higher than in patients with chronic HF and fall rapidly over days to weeks during HF treatment. [13] This lack of reduction in ST2 level during acute HF treatment is predictive of mortality. In addition, persistently high levels of ST2 were associated with increased mortality risk. [16] Only a few studies, most in patients with chronic systolic HF, have evaluated the prognostic value of the change in ST2 levels, in which the ST2 level was measured with an interval of at least 1 month. [14, 15] Increases in ST2 levels from baseline to 12 months were associated with a significant increased risk for all-cause mortality. On the contrary, the CORONA study (Controlled Rosuvastatin Multinational Trial in Heart Failure) showed that change in ST2 levels from baseline to 3 months was not associated with mortality. [17] The RELAX-AHF (Efficacy and Safety of Relaxin for the Treatment of Acute Heart Failure) trial showed that serial sST2 measurements combined in a multimarker approach are useful for prognostication in patients with acute HF. [18]

Given the dynamic and progressive nature of HF and the pathophysiology of ST2, we hypothesized that in patients admitted with acute HF, frequently measured ST2 levels during follow-up will add incremental prognostic information to that conferred by repeated measurements of NT-proBNP. In the American Heart

Association/American College of Cardiology guidelines for management of heart failure, ST2 is considered useful for prognostication and therapy monitoring, but more research is required to support this suggestion. [19] Therefore, in the present TRIUMPH study (TRanslational Initiative on Unique and novel strategies for Management of Patients with Heart failure [TRIUMPH]: NTR1893), we assessed the association between frequently measured ST2 independent of frequently measured NT-proBNP and the incidence of all-cause mortality and HF readmission during 1-year follow-up in 496 patients admitted with acute HF.

Methods

Objective and Study Design

TRIUMPH was designed as a translational bench-to-bedside study program encompassing the entire spectrum of biomarker discovery to clinical validation. The clinical validation study was an observational prospective study enrolling patients admitted with acute HF in 14 hospitals in the Netherlands between September 2009 and December 2013. This cohort study was designed to validate the clinical value of biomarkers successfully passing the bioinformatics and early validation stages of TRIUMPH, as well as to evaluate more established biomarkers of HF further. There was a particular interest in the change in biomarker levels over time, as well as in the analyses and prognostic significance of repeated biomarker sampling during the follow-up of patients with HF. The study was approved by the medical ethics committees at all participating centers.

Patient Selection

Patients ≥18 years of age were eligible for enrollment if they were hospitalized with decompensation of known chronic HF or newly diagnosed HF. Furthermore, 3 other criteria had to be met: 1) natriuretic peptide levels had to be elevated to ≥3 times the upper limit of normal; 2) there had to be evidence of sustained systolic or diastolic left ventricular dysfunction; and 3) patients had to be treated with intravenous diuretics. Patients with HF that was precipitated by a noncardiac condition, by severe valvular dysfunction without sustained left ventricular dysfunction, or by an acute ST-segment elevation myocardial infarction were excluded. Furthermore, patients scheduled for a coronary revascularization procedure, on a waiting list for heart transplantation, with severe renal failure for which dialysis was needed, or with a coexisting condition with a life expectancy <1 year could not participate. All study participants provided written informed consent.

Patient Management

Patient management was at the discretion of the treating physician and was provided in accordance with the guidelines of the European Society of Cardiology. [20] Importantly, the biomarker data that were generated in the context of this observational study were not used for treatment decisions.

Study Procedures

During hospitalization, blood samples were obtained at admission (day 1), once during days 2 to 4, and subsequently on the day of discharge. Afterward, repeated blood samples were also obtained at outpatient follow-up visits, which were planned at 2 to 4 weeks, 3 months, 6 months, and 9 to 12 months after discharge. The baseline blood sample was defined as the first sample obtained

after inclusion, up to a maximum of 2 days after inclusion. At each visit, HF symptoms were assessed using the New York Heart Association functional classification. Medication use was determined at discharge by using 3 categories: 1) use of an angiotensin-converting enzyme inhibitor or an angiotensin II receptor antagonist, or both; 2) use of a beta-blocker; or 3) use of diuretics. Patients underwent physical examination, and weight, blood pressure, and heart rate were systematically measured.

Blood Collection

Nonfasting blood samples were obtained by venipuncture and transported to the clinical chemistry laboratory of each participating hospital for further processing according to a standardized protocol. The collected material was centrifuged at 1,700 G/relative centrifugal force, and then heparin plasma and blood serum were separated. All blood aliquots were subsequently stored at a temperature of -80°C within 2 h after venipuncture.

ST2 Measurements

Serum samples and heparin plasma samples were transported under controlled conditions to a central laboratory (Future Diagnostics Solutions B.V., Wijchen, the Netherlands) for batch analysis of ST2 and NT-proBNP levels. ST2 concentrations were determined in serum in single measurements by using a quantitative sandwich monoclonal enzyme-linked immunosorbent assay (Presage ST2 Assay, Critical Diagnostics, Inc., San Diego, California). In our hands the average coefficient of variation for interassay variation was 4.9%, in line with the average interassay coefficient of variation of 5.2% reported by the manufacturer. NT-proBNP concentrations were determined in heparin plasma by using the Elecsys NT-proBNP electrochemiluminescent sandwich immunoassay on a Cobas 8000

analyzer (Roche Diagnostics, Ltd., Rotkreuz, Switzerland). Analysts were blinded to patients' characteristics and endpoints.

ST2 Pattern

Post hoc analyses were performed to identify ST2 patterns in patients with and without the primary endpoint. Two investigators, blinded to baseline patients' characteristics and clinical outcomes data, individually analyzed the ST2 pattern. ST2 patterns were classified as follows: 1) "U-shaped," if the ST2 level initially decreased and later increased; 2) "J-shaped," if the ST2 level initially decreased and did not increase later; 3) "not interpretable," if fewer than 3 ST2 measurements were available or 3 ST2 measurements were close together; or 4) "other," if a different ST2 pattern was identified. If there was disagreement, a consensus was reached in a separate session.

Endpoints

Information on vital status and hospital readmissions was obtained until at least 9 months with a maximum of 400 days after the index hospitalization. We approached the civil registry, screened all medical records, and asked patients for information during their follow-up visits. The primary endpoint is the composite of all-cause mortality and readmission for HF. Readmission for HF was defined as an unplanned rehospitalization resulting from decompensation of HF, with at least 2 of the following 3 criteria being present: elevated natriuretic peptide levels $\geq$3 times the upper limit of normal; symptoms of cardiac decompensation (rales, edema, or elevated central venous pressure); and treatment with intravenous diuretics. Secondary endpoints included the individual components of the primary endpoint and cardiovascular mortality. An event adjudication committee, blinded to biomarker information, was established

Table 1: Baseline Characteristics According to Overall Sample (n = 475) and Quartiles of Baseline ST2 Level (n = 386)

	Overall Sample	Q1	Q2	Q3	Q4	p-value*
Demographic characteristics						
Age, y	74 (65-80)	72	75	73	74	0.427
Female	37	45	37	38	34	0.434
Caucasian	95	91	95	95	95	0.541
Measurements at baseline						
Body mass index, kg/m^2	28 (25-31)	28	28	28	27	0.768
Systolic blood pressure, mm Hg	125 (110-147)	128	135	124	124	0.534
Diastolic blood pressure, mm Hg	74 (65-85)	75	76	72	74	0.513
Heart rate, bpm	85 (72-100)	85	86	84	84	0.503
eGFR, mL/min per 1.73 m^2	46 (34-62)	51	49	44	40	0.002
Left ventricular ejection fraction, %	30 (21-41)	34	30	30	29	0.204
NYHA classification						0.378
II	17	20	16	16	11	
III	55	53	58	63	53	
IV	27	27	25	20	34	
Medical history						
Newly diagnosed heart failure	36	43	40	37	27	0.088
Heart failure with reduced ejection fraction	83	78	85	79	87	0.434
Previous heart failure admission within 6 mo	20	20	18	15	27	0.245
Ischemic heart failure	49	43	44	47	53	0.498
Myocardial infarction	40	35	31	43	50	0.034
Hypertension	51	55	55	46	48	0.470
Atrial fibrillation	42	38	45	43	46	0.640
Diabetes mellitus	36	32	32	41	39	0.439
Stroke	17	13	16	16	19	0.718
Biomarkers						
ST2, ng/mL	71 (46-102)	37	59	89	132	
NT-proBNP, pg/mL	4,152 (2,089-9,387)	2,347	3,970	4,871	5,692	<0.001
Endpoints						
Primary endpoint	40	23	34	44	52	<0.001
All-cause mortality	24	7	20	26	32	<0.001
HF hospitalization	26	20	27	33	34	0.150
Cardiovascular mortality	16	2	15	17	23	<0.001

Values are median (interquartile range) or %.
*p value for differences between quartiles of baseline ST2 level.
eGFR, estimated glomerular filtration rate; HF, heart failure; NT-pro-BNP, N-terminal pro–B-type natriuretic peptide; NYHA, New York Heart Association; Q, quartile.

for reviewing and adjudication of endpoints.

Statistical Analysis

The distributions of continuous variables were evaluated for normality by visual examination of the histogram and Kolmogorov-Smirnov tests. Variables with a normal distribution are presented as mean ± SD, whereas the median and interquartile range (IQR) are presented in case of non-normality. Categorical variables are presented as counts and percentages. ST2 and NT-proBNP levels had a non-normal distribution and were therefore log-transformed for further analyses.

Patients were classified according to the quartiles of the ST2 distribution, and differences in baseline characteristics between these quartiles were evaluated by chi-square tests (categorical variables), analysis of variance, or Kruskal-Wallis tests, as appropriate.

We applied Cox proportional hazards models to evaluate the association of baseline ST2 levels with the study endpoints. Subjects were censored at the time of occurrence of the endpoint under investigation, death, and at the scheduled end of follow-up. No deviations of the proportional hazards assumption were found by inspecting log minus log plots of the survival functions. We performed univariate analyses to obtain the crude estimates of the effect of baseline ST2 level (model 1), analyses that were adjusted for age and sex only (model 2), and analyses that were additionally adjusted for systolic blood pressure, diabetes mellitus, left ventricular ejection fraction (LVEF), previous hospitalization for HF during the last 6 months, ischemic HF, body mass index, estimated glomerular filtration rate (eGFR), and baseline NT-proBNP level (model 3). The results are presented as adjusted hazard ratios (HRs) per 1 SD increase of the biomarker level (on the log2 scale) with 95% confidence intervals (CIs).

We calculated the eGFR using the Modification of Diet in Renal Disease equation. [21]

Joint models were fitted to assess the association between estimated instantaneous biomarker levels during follow-up, calculated using the repeated time-dependent biomarker levels, and the specified study endpoints. A joint model combines a mixed effects linear regression model for the serial measurements with a Cox proportional hazards model for the risk of the specified study endpoints. [22] We used cubic splines, with knots set at 1 week and 1 month after initial hospitalization, for the mixed model. For the analyses with the repeated ST2 measurements, we performed univariate analyses (model 1). We combined repeated measurements of ST2 and NTproBNP in 1 joint model to assess their independent prognostic value and adjusted for age and sex (model 2). We additionally adjusted for systolic blood pressure, diabetes mellitus, LVEF, previous hospitalization for HF during the last 6 months, ischemic HF, body mass index, eGFR, and use of medication at hospital discharge (angiotensin-converting enzyme inhibitor and/or angiotensin II receptor antagonist, beta-blocker, diuretics) (model 3). We also tested whether the slope of the ST2 trajectories itself, when added to model 3, was an independent predictor. Diagnostics and sensitivity analyses were performed to evaluate the joint models. The final results are presented as adjusted HRs per 1 SD increase of the biomarker level (on the log2 scale) at any point in time with 95% CIs. Data on covariates were complete in 93% of patients, except for LVEF, which was complete in 78%. Single imputation was applied to account for missing values of covariates.

Statistical Package for Social Sciences, version 21.0 software (SPSS, IBM Corp., Armonk, New York) was used for descriptive data analysis. R statistical software (version 2.15.0, R Foundation, Vienna, Austria) was used for advanced statistical analyses of the longitudinal biomarker data and study endpoints

(packages JMBayes and JM). All statistical tests were 2-tailed, and p values <0.05 were considered statistically significant.

Table 2: Hazard Ratios for Different End Points Per 1 SD Increase of the Baseline ST2 Level (on the log2 Scale)

		Baseline Level*	
	N	HR (95% CI)	P-value
Primary end point			
Model 1†		1.49 (1.26-1.77)	<0.001
Model 2		1.48 (1.25-1.76)	<0.001
Model 3		1.30 (1.08-1.56)	0.005
Number of events/patients	188/475		
All-cause mortality			
Model 1		1.80 (1.41-2.29)	<0.001
Model 2		1.77 (1.39-2.27)	<0.001
Model 3		1.43 (1.11-1.86)	0.006
Number of events/patients	113/475		
HF hospitalization			
Model 1		1.33 (1.09-1.61)	0.005
Model 2		1.33 (1.09-1.61)	0.005
Model 3		1.16 (0.94-1.43)	0.159
Number of events/patients	123/475		
Cardiovascular mortality			
Model 1		2.01 (1.49-2.72)	<0.001
Model 2		1.98 (1.46-2.67)	<0.001
Model 3		1.63 (1.19-2.23)	0.002
Number of events/patients	77/475		

Mean ± 1 SD of the patient-specific geometric mean ST2 value at baseline (presented on the linear scale): 70.0 (40.7 ± 120.3).
*Hazard ratios are related to a 1 SD increase of ST2 (on the log scale) at baseline.
†Model 1 unadjusted; model 2 adjusted for age and sex; model 3 adjusted for age, sex, systolic blood pressure, diabetes mellitus, left ventricular ejection fraction, previous hospitalization for HF during the last 6 months, ischemic heart failure, body mass index, estimated glomerular filtration rate, and baseline NT-proBNP.
CI, confidence interval; HF, heart failure; HR, hazard ratio; other abbreviations as in Table 1.

Results

Patients

A total of 496 patients were enrolled in the TRIUMPH clinical cohort. Three patients withdrew their informed consent. Eighteen patients were withdrawn from statistical analyses because of inclusion violation. These patients had no evidence of sustained systolic or diastolic left ventricular dysfunction on echocardiography. Accordingly, 475 patients comprised the analysis set. Their median age was 74 years (IQR: 65 to 80 years), and 37% were women (Table 1). Median systolic blood pressure was 125 mm Hg (IQR: 110 to 147 mm Hg), and median LVEF was 30% (IQR: 21% to 42%). Most patients had HF with a reduced ejection fraction (83%). The median baseline ST2 level was 71 ng/ml (IQR: 46 to 102 ng/ml), and that of NT-proBNP was 4,152 pg/ml (IQR: 2,089 to 9,387 pg/ml). Additionally, Table 1 shows the baseline characteristics of patients in different quartiles of ST2 level. Patients in quartiles with a higher ST2 level had worse kidney function, and more patients had a history of myocardial infarction.

Baseline ST2 Levels and the Incidence of Study End Points

During the median follow-up of 325 days (IQR: 85 to 401 days), 188 patients (40%) reached the primary endpoint of all-cause death (n = 113) or readmission for HF (n = 123). This corresponds with an incidence rate of 55.9 per 100 patient-years for the primary endpoint. Baseline ST2 levels were available in 386 patients. In the highest quartile of baseline ST2, 50 patients (52%) reached the primary endpoint compared with 22 patients (23%) in the lowest quartile of ST2. All-cause mortality was also higher in the highest ST2 quartile compared with the lowest ST2 quartile: 31 (32%) and 7 (7%), respectively. This was

similar for cardiovascular mortality: 22 (23%) and 2 (2%), respectively (Table 1).

The baseline ST2 level was associated with an increased risk of all the predefined study endpoints (Table 2). With respect to the primary endpoint, all-cause mortality and cardiovascular mortality, these associations remained statistically significant after adjustment for all selected potential confounders, including baseline NT-proBNP level (model 3).

Prognostic Value of Repeated ST2 Measurements

The average number of ST2 measurements per patient during follow-up was 3.9 and 4.1 for NT-pro-BNP. After adjustment for repeated measurements of NT-pro-BNP, age, and sex (model 2), the HR for the primary endpoint corresponding to a 1 SD increase of ST2 level (on the log2 scale) during follow-up was 3.54 (95% CI: 2.07 to 7.32; p<0.001). After adjustment for the broader range of potential confounders including repeated measurements of NT-proBNP (model 3), the association remained statistically significant, with an HR corresponding to a 1 SD increase of ST2 level (on the log2 scale) during follow-up of 1.85 (95% CI: 1.02 to 3.33; p=0.044). The HR corresponding to a 1 SD increase of NT-proBNP level (on the log2 scale) during follow-up for the primary endpoint was 2.13 (95% CI: 1.35 to 3.88; p<0.001) adjusted for model 3 and repeated measurements of ST2 (Table 3). The HRs for all-cause and cardiovascular mortality corresponding to a 1 SD increase of ST2 level (on the log2 scale) during follow-up after adjustment for all covariates and repeated measurements of NT-proBNP (model 3) were highly statistically significant: 4.36 (95% CI: 2.31 to 8.92; p<0.001) and 3.98 (95% CI: 2.15 to 7.94; p<0.001), respectively. The slope of the ST2 level trajectories itself was not an independent predictor of the primary endpoint.

Table 3: Hazard Ratios for Different End Points Per 1 SD Increase of ST2 Level or NT-proBNP Level (on the log2 Scale) at Any Point in Time Using Repeated ST2 and NT-proBNP Measurements in a Joint Model

		Mean Value†			Instantaneous Level‡	
	Model*	M - SD	M	M + SD	HR (95% CI)	p value
Primary end point						
ST2 (crude)	1	24.2	41.4	70.9	2.78 (2.16-3.64)	<0.001
ST2	2	24.2	41.4	70.9	3.54 (2.07-7.32)	<0.001
NT-proBNP	2	517	1,776	6,093	1.67 (1.20-2.34)	0.002
ST2	3	24.2	41.4	70.9	1.85 (1.02-3.33)	0.044
NT-proBNP	3	517	1,776	6,093	2.13 (1.35-3.88)	<0.001
All-cause mortality						
ST2 (crude)	1	24.8	42.6	73.3	4.45 (3.12-6.39)	<0.001
ST2	2	24.8	42.6	73.3	4.19 (2.31-8.79)	<0.001
NT-proBNP	2	545	1,874	6,447	1.85 (1.22-2.83)	0.002
ST2	3	24.8	42.6	73.3	4.36 (2.31-38.92	<0.001
NT-proBNP	3	545	1,874	6,447	2.48 (1.35-6.10)	0.004
HF hospitalization						
ST2 (crude)	1	24.2	41.4	70.9	2.24 (1.68-3.01)	<0.001
ST2	2	24.2	41.4	70.9	1.80 (1.27-2.56)	<0.001
NT-proBNP	2	517	1,776	6,093	1.62 (1.18-2.19)	0.002
ST2	3	24.2	41.4	70.9	1.10 (0.64-1.83)	0.690
NT-proBNP	3	517	1,776	6,093	1.47 (0.92-2.45)	0.096
Cardiovascular mortality						
ST2 (crude)	1	24.8	42.6	73.3	5.27 (3.31-8.31)	<0.001
ST2	2	24.8	42.6	73.3	4.55 (2.47-8.37)	<0.001
NT-proBNP	2	545	1,874	6,447	1.66 (1.05-2.67)	0.022
ST2	3	24.8	42.6	73.3	3.98 (2.15-7.94)	<0.001
NT-proBNP	3	545	1,874	6,447	1.85 (1.02-3.45)	0.046

*Model 1 unadjusted; model 2 adjusted for repeated measurements of NT-proBNP or ST2, age, and sex; model 3 adjusted for repeated measurements of NT-proBNP or ST2, age, sex, systolic blood pressure, diabetes mellitus, left ventricular ejection fraction, previous hospitalization for HF during the last 6 months, ischemic HF, body mass index, eGFR, and use of medication at hospital discharge (angiotensin-converting enzyme inhibitor and/or angiotensin II receptor antagonist, beta-blocker, diuretics).
†Mean ±1 SD of the patient-specific geometric mean biomarker level during follow-up (presented on the linear scale).
‡Hazard ratios are related to a 1 SD increase of biomarker level (on the log scale) at any point in time.
Abbreviations as in Tables 1 and 2.

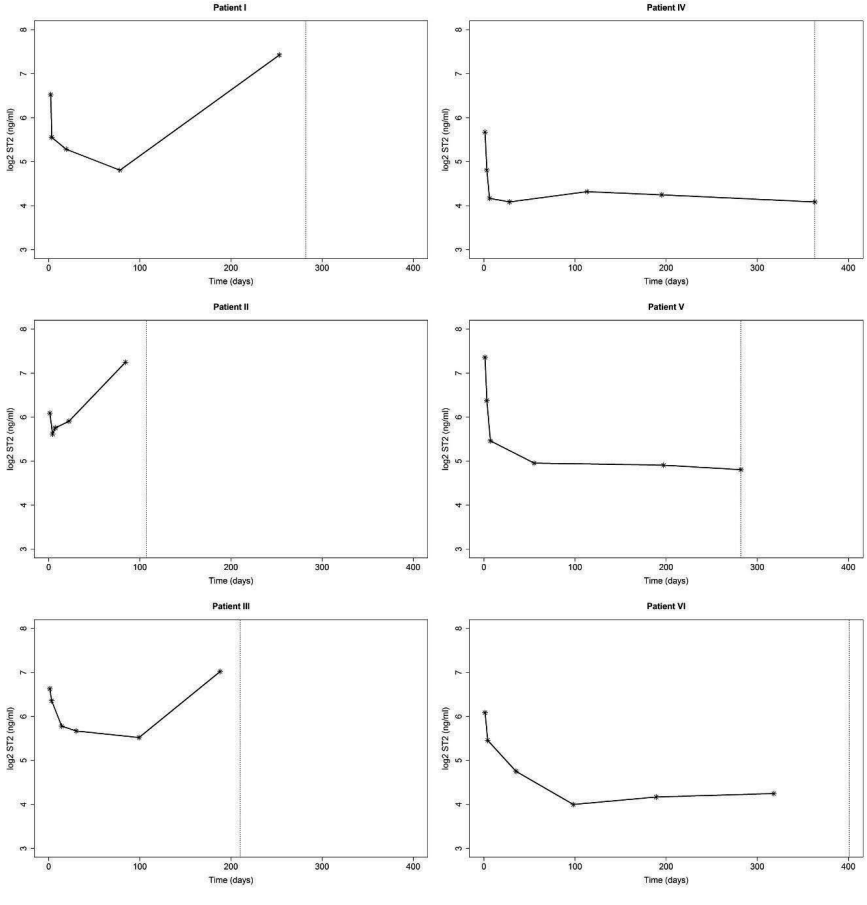
Figure 1 shows the measured ST2 levels of 3 individuals who had a U-shaped ST2 pattern and of 3 individuals who had a J-shaped pattern. Of the patients who reached the primary endpoint, 56% had a U-shaped ST2 pattern preceding the occurrence of the endpoint event, as illustrated in Figure 1 patients I, II, and III. Figure 1 patients IV, V, and VI are examples of J-shaped ST2 patterns in patients who did not reach the primary endpoint. When a J-shaped ST2 pattern was present during follow-up, 82% of the patients remained event free.

Figure 2 shows the average estimated biomarker level and the individual biomarker measurements in patients with and without the primary endpoint adjusted according to model 3. During initial hospitalization, when all patients were treated for decompensated HF, the average estimated ST2 level decreased (Figure 2A). Following initial hospitalization, the average estimated ST2 levels in patients who reached the primary endpoint were higher than in their counterparts who remained free of the primary endpoint. Furthermore, the average estimated ST2 levels increased several weeks before the time of the primary endpoint. The shape of the average estimated NT-proBNP pattern following initial hospitalization was comparable to that of the average estimated ST2 pattern (Figure 2B).

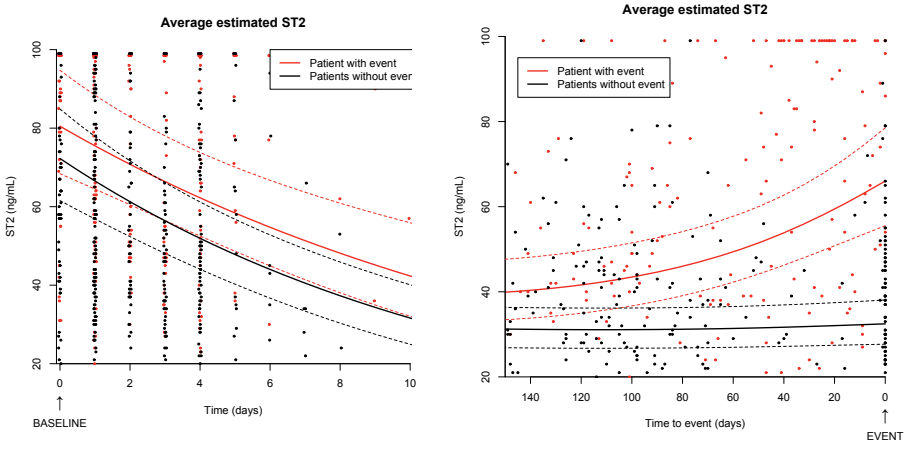
Discussion

This study clearly demonstrates that baseline ST2 levels, and especially repeated ST2 measurements, are a strong and independent predictor of the composite endpoint of all-cause mortality or readmission for HF during 1-year follow-up in patients admitted with acute HF. Our results support the concept that serial measurements of ST2 offer substantial incremental prognostic value to (repeatedly measured) NT-proBNP, which is still considered the gold standard biomarker in HF.

Figure 1: Examples of the ST2 Pattern During Follow-Up in Different Patients



The ST2 level of 6 patients during follow-up. The vertical dotted line represents the time of occurrence of the primary endpoint or the scheduled end of follow-up. Patients I, II, and III demonstrate a U-shaped ST2 pattern and reach the primary endpoint. Patients IV, V, and VI demonstrate a J-shaped ST2 pattern and remained event free during follow-up.



(a) (b)

Figure 2: (a) Average estimated ST2 pattern during initial hospitalization for decompensated heart failure for patients with and without the primary endpoint. (b) Average estimated N-terminal pro–B-type natriuretic peptide (NT-proBNP) pattern before the primary endpoint or at the end of follow-up for patients with and without the primary endpoint. The average estimated ST2 and NT-proBNP levels are adjusted for age, sex, systolic blood pressure, diabetes mellitus, left ventricular ejection fraction, previous hospitalization for heart failure during the last 6 months, ischemic heart failure, body mass index, estimated glomerular filtration rate, and use of medication at hospital discharge (angiotensin-converting enzyme inhibitor and/or angiotensin II receptor antagonist, beta-blocker, diuretics) (model 3).

The TRIUMPH study was designed to identify and validate novel biomarkers to improve prognostication in HF. TRIUMPH was designed as a translational study program combining biological discovery of novel biomarkers, technological advances, and clinical validation in patients presenting with acute HF. In the clinical validation study, the biomarkers were evaluated for their prognostic properties by using a unique design of repeated measurements during 1-year follow-up. Within TRIUMPH, ST2 was labeled as a biomarker with high potential for improving prognostication.

It has been established that ST2 levels in patients with acutely decompen-

sated HF are useful for prognostication. [3, 23, 24] Our observation that baseline ST2 level was significantly associated with all of the predefined study endpoints confirms this. In line with previous studies, the association between baseline ST2 level and readmission for HF is weaker than the association between baseline ST2 and the mortality endpoints when adjusted for all potential confounders and baseline NT-proBNP.

Repeated ST2 measurements were strongly related to the primary endpoint, as well as its separate components. The association between repeated ST2 level and the primary endpoint was highly significant and considerably stronger than the association between baseline ST2 level and the primary endpoint. Repeated measurements take into account the dynamic and continuous change in ST2 level over time that may better reflect the true changes that occur in the underlying pathophysiological processes in the individual patient with HF. In this study, repeated ST2 measurements were used to estimate the instantaneous ST2 levels (i.e., the estimated ST2 level at any point in time during the follow-up period). These estimated instantaneous ST2 levels were strongly associated with the occurrence of the predefined endpoints, most likely because the level of the estimated ST2 level is close to the true ST2 level and therefore reflects the true cardiac condition of the patient at that point in time during follow-up. This is important because HF is a dynamic and often progressive disease in which inflammation, cardiac fibrosis, and remodeling are ongoing processes that cannot be captured in a single biomarker assessment at 1 point in time. [5]

Another finding of the present study is that the estimated average ST2 levels increase in patients before the primary endpoint is reached, whereas the average estimated ST2 level in patients without the primary endpoint during follow-up stabilizes. The slope of the ST2 trajectory itself did not add significant prognostic information to the estimated instantaneous ST2 level. An explanation

for this finding could be that the distribution of the biomarker measurements is not ideal for assessment of the instantaneous slope. To clarify these findings, a post hoc analysis was performed to define the ST2 pattern in individual patients. This analysis demonstrated that almost twice as many patients who reached the primary endpoint during follow-up had a so-called U-shaped ST2 pattern, compared with patients without an event. Furthermore, when a J-shaped ST2 pattern was identified, 82% of these patients remained event free during 1 year of follow-up. Although we acknowledge that the classification of the ST2 pattern may be affected by subjectivity and that one should be careful about drawing conclusions from this post hoc analyses, these findings suggest that the progression of ST2 levels may be important for the evaluation of an HF patient. The increase or stabilization of the ST2 level may be a useful variable in daily practice not only for stratifying patients in high-risk and low-risk categories but even more so for acting on an anticipated cardiac deterioration of a patient when ST2 levels rise during outpatient clinic follow-up visits.

Another important finding of the present study is that repeated ST2 measurements conferred independent prognostic information in addition to that offered by repeated NT-proBNP measurements. The finding that NT-proBNP and ST2 levels reflect different underlying pathophysiological processes in HF may be the most important reason for this observation. NT-proBNP is a marker of volume overload. [25] ST2 responds to mechanical overload as well, but it is also a marker of cardiac fibrosis, inflammation, and remodeling. [8] In this way, ST2 and NT-proBNP levels provide complementary information on the pathophysiological state, as well as information relevant to the assessment of prognosis. With respect to prognostication in HF, the results of the present study therefore provide evidence not only for the use of repeated ST2 measurements, but also for the combined use with (repeatedly measured) NT-proBNP

levels.

This study combined repeated ST2 measurements with repeated NT-proBNP measurements in patients with acute HF and therefore adds important evidence to the statement in the AHA/ACC guidelines for management of HF that ST2 is considered useful for prognostication and therapy monitoring, in addition to the use of NT-proBNP. [19]

Future studies should assess the value of repeated ST2 measurements when used to guide treatment decisions. It may be hypothesized that treatment should be intensified in patients with high ST2 levels or unfavorable (increasing) ST2 patterns. Moreover, repeated ST2 measurements may be helpful to identify patients who are more likely to respond to certain treatments. Additional studies should also determine the number of ST2 measurements needed for optimal prognostication and therapy monitoring. The frequency by which ST2 levels should be measured may not be identical for each patient, but they may depend on the clinical condition of the patient, the treatment given, the ST2 level, and the progression of ST2 levels during follow-up. On the basis of these factors, an individual survival curve could be plotted, which should be used for planning of the next ST2 measurement. Because of the significantly lower biological variability of ST2 compared with NT-proBNP in patients with stable HF, it has been suggested that ST2 may be a better biomarker for monitoring patients with HF. [26]

Study Limitations

Although this study is a large multicenter prospective observational study, it seems that the studied population is not completely representable for the average HF population. The mean age in our study population is 74 years, and women are underrepresented. Moreover, only 17% of the included patients

with HF have a preserved ejection fraction. Future studies need to investigate whether similar results are found in a population that represents more women, different age groups, and HF patients with a preserved ejection fraction.

Conclusion

The TRIUMPH study clearly demonstrates that repeated measurements of ST2 are a strong and independent predictor of adverse outcome in patients following admission for acute HF. The repeated ST2 measurements identified patients at a substantially higher risk of adverse events than did baseline ST2 levels alone. In addition, repeated ST2 measurements offer incremental prognostic value to that conferred by other known risk factors and, importantly, repeated measurements of NT-proBNP. These results suggest that repeated ST2 measurements in addition to NT-proBNP measurements may be helpful in clinical practice to identify patients with HF who are at increased risk of adverse outcomes.

References

1. Go, A. S. *et al.* Heart disease and stroke statistics–2014 update: a report from the American Heart Association. *Circulation* **129,** e28–e292 (2014).
2. Manzano-Fernandez, S., Mueller, T., Pascual-Figal, D., Truong, Q. A. & Januzzi, J. L. Usefulness of Soluble Concentrations of Interleukin Family Member ST2 as Predictor of Mortality in Patients With Acutely Decompensated Heart Failure Relative to Left Ventricular Ejection Fraction. *American Journal of Cardiology* **107,** 259–267 (2011).
3. Lassus, J. *et al.* Incremental value of biomarkers to clinical variables for mortality prediction in acutely decompensated heart failure: the Multinational Observational Cohort on Acute Heart Failure (MOCA) study. *Int J Cardiol* **168,** 2186–94 (2013).
4. Januzzi J. L., J. *et al.* Utility of amino-terminal pro-brain natriuretic peptide testing for prediction of 1-year mortality in patients with dyspnea treated in the emergency department. *Arch Intern Med* **166,** 315–20 (2006).
5. Braunwald, E. Heart failure. *JACC Heart Fail* **1,** 1–20 (2013).
6. Schmitz, J. *et al.* IL-33, an interleukin-1-like cytokine that signals via the IL-1 receptor-related protein ST2 and induces T helper type 2-associated cytokines. *Immunity* **23,** 479–90 (2005).

7. Sanada, S. *et al.* IL-33 and ST2 comprise a critical biomechanically induced and cardioprotective signaling system. *J Clin Invest* **117,** 1538–49 (2007).

8. Weinberg, E. O. *et al.* Expression and regulation of ST2, an interleukin-1 receptor family member, in cardiomyocytes and myocardial infarction. *Circulation* **106,** 2961–6 (2002).

9. Weinberg, E. O. *et al.* Identification of serum soluble ST2 receptor as a novel heart failure biomarker. *Circulation* **107,** 721–726 (2003).

10. Bartunek, J. *et al.* Nonmyocardial production of ST2 protein in human hypertrophy and failure is related to diastolic load. *J Am Coll Cardiol* **52,** 2166–74 (2008).

11. Ky, B. *et al.* High-sensitivity ST2 for prediction of adverse outcomes in chronic heart failure. *Circ Heart Fail* **4,** 180–7 (2011).

12. Aimo, A. *et al.* Prognostic Value of Soluble Suppression of Tumorigenicity-2 in Chronic Heart Failure: A Meta-Analysis. *JACC Heart Fail* **5,** 280–286 (2017).

13. Boisot, S. *et al.* Serial sampling of ST2 predicts 90-day mortality following destabilized heart failure. *J Card Fail* **14,** 732–8 (2008).

14. Gaggin, H. K. *et al.* Head-to-head comparison of serial soluble ST2, growth differentiation factor-15, and highly-sensitive troponin T measurements in patients with chronic heart failure. *JACC Heart Fail* **2,** 65–72 (2014).

15. Anand, I. S., Rector, T. S., Kuskowski, M., Snider, J. & Cohn, J. N. Prognostic value of soluble ST2 in the Valsartan Heart Failure Trial. *Circ Heart Fail* **7,** 418–26 (2014).

16. Tang, W. H. *et al.* Prognostic Value of Baseline and Changes in Circulating Soluble ST2 Levels and the Effects of Nesiritide in Acute Decompensated Heart Failure. *JACC Heart Fail* **4,** 68–77 (2016).

17. Broch, K. *et al.* Soluble ST2 is associated with adverse outcome in patients with heart failure of ischaemic aetiology. *European Journal of Heart Failure* **14,** 268–277 (2012).

18. Demissei, B. G. *et al.* A multimarker multi-time point-based risk stratification strategy in acute heart failure: results from the RELAX-AHF trial. *Eur J Heart Fail* **19,** 1001–1010 (2017).

19. Yancy, C. W. *et al.* 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol* **62,** e147–239 (2013).

20. McMurray, J. J. *et al.* ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. *Eur Heart J* **33,** 1787–847 (2012).

21. Stevens, L. A., Coresh, J., Greene, T. & Levey, A. S. Assessing kidney function–measured and estimated glomerular filtration rate. *N Engl J Med* **354,** 2473–83 (2006).

22. Rizopoulos, D. *The R Package JMbayes for Fitting Joint Models for Longitudinal and Time-to-Event Data using MCMC* Web Page. 30 April 2014.

23. Rehman, S. U., Mueller, T. & Januzzi J. L., J. Characteristics of the novel interleukin family biomarker ST2 in patients with acute heart failure. *J Am Coll Cardiol* **52,** 1458–65 (2008).

24. Shah, R. V., Chen-Tournoux, A. A., Picard, M. H., van Kimmenade, R. R. & Januzzi, J. L. Serum levels of the interleukin-1 receptor family member ST2, cardiac structure and function, and long-term mortality in patients with acute dyspnea. *Circ Heart Fail* **2,** 311–9 (2009).

25. Hunt, P. J. *et al.* Immunoreactive amino-terminal pro-brain natriuretic peptide (NT-PROBNP): a new marker of cardiac impairment. *Clin Endocrinol (Oxf)* **47,** 287–96 (1997).

26. Piper, S., deCourcey, J., Sherwood, R., Amin-Youssef, G. & McDonagh, T. Biologic Variability of Soluble ST2 in Patients With Stable Chronic Heart Failure and Implications for Monitoring. *Am J Cardiol* **118,** 95–8 (2016).

Chapter 8

Parent reports of health-related quality of life and heart failure severity score independently predict outcome in children with dilated cardiomyopathy

den Boer SL, **Baart SJ**, van der Meulen MH, van Iperen GG, Backx AP, ten Harkel AD, Rammeloo LA, du Marchie Sarvaas GJ, Tanke RB, Helbing WA, Utens EM, Dalinghaus M

Cardiology in the Young 2017; 27:1194-1202

Abstract

Background: Dilated cardiomyopathy in children causes heart failure and has a poor prognosis. Health-related quality of life in this patient group is unknown. Moreover, results may provide detailed information of parents' sense of their child's functioning. We hypothesised that health-related quality of life, as rated by parents, and the paediatric heart failure score, as assessed by physicians, have both predictive value on outcome.

Methods and Results: In this prospective study, health-related quality of life was assessed by parent reports: the Infant Toddler Quality of Life questionnaire (0–4 years) or Child Health Questionnaire-Parent Form 50 (4–18 years) at 3–6-month intervals. We included 90 children (median age 3.8 years, interquartile range (IQR) 0.9–12.3) whose parents completed 515 questionnaires. At the same visit, physicians completed the New York University Pediatric Heart Failure Index. Compared with Dutch normative data, quality of life was severely impaired at diagnosis (0–4 years: 7/10 subscales and 4–18 years: 8/11 subscales) and $\geq$1 year after diagnosis (3/10 and 6/11 subscales). Older children were more impaired ($p < 0.05$). After a median follow-up of 3 years (IQR 2–4), 15 patients underwent transplantation. Using multivariable time-dependent Cox regression, "physical functioning" subscale and the Heart Failure Index were independently predictive of the risk of death and heart transplantation (hazard ratio 1.24 per 10% decrease of predicted, 95% confidence interval (CI) 1.06–1.47 and hazard ratio 1.38 per unit, 95% CI 1.19–1.61, respectively).

Conclusions: Physical impairment rated by parents and heart failure severity assessed by physicians independently predicted the risk of death or heart transplantation in children with dilated cardiomyopathy.

Introduction

Dilated cardiomyopathy in children causes heart failure and may have a poor prognosis. After diagnosis, the 1-year transplant-free survival rate has been reported to be between 69 and 82% and the 5-year transplant-free survival rate between 54 and 72%. [1, 2] Around 35% of the children, however, develop chronic dilated cardiomyopathy and around 35% recover, with the highest recovery rates seen in children aged 1–6 years at diagnosis. [1] To assess the impact of disease on patient life, functional status assignment by a physician and patient-reported health-related quality of life have been used, and may contain important prognostic information. In adults, the NYHA Classification is used to categorise heart failure functional class and has been strongly associated with outcome; [3, 4] furthermore, in adults with heart failure, health- related quality of life is affected as compared with healthy, age-matched controls, but also as compared with other chronically ill patients. [5–7] In addition, health-related quality of life has been shown to be an independent predictor for mortality. [8] In children with heart failure secondary to dilated cardiomyopathy, such data are largely lacking. To assess functional class, the New York University Pediatric Heart Failure Index has been developed. [9]

This score, however, has not been related to clinical outcome in dilated cardiomyopathy yet. In children, the effect of dilated cardiomyopathy on health-related quality of life is largely unknown. An explorative study investigating parent-reported health-related quality of life in children visiting the paediatric cardiology clinic for various diseases reported on a small subgroup of 17 children with cardiomyopathy. [10] Using the Child Health Questionnaire-Parent Form 50, cardiomyopathy patients scored worse compared with all other patients attending the cardiology clinic on “physical functioning”, “general health perceptions”, and “parental impact – emotional”.

The use of health-related quality-of-life questionnaires in children enables a structural assessment of patients' physical and psychosocial functioning reported by parents. As parents "know their child best", we hypothesised that parents' assessment of their child's health-related quality of life, on an internationally validated questionnaire, provides valuable information about a child's functioning, which may have prognostic value; furthermore, we hypothesised that physicians' assessment of heart failure severity, using a validated heart failure severity score, also provides prognostic information.

The present study had two aims. First, to evaluate health-related quality of life in children with dilated cardiomyopathy. Second, to assess the predictive value of health-related quality-of-life subscales and the heart failure severity score on the risk of death and heart transplantation at diagnosis and during follow-up.

Materials and Methods

The institutional review boards of the seven participating centres approved the study protocol. Parents and children ≥12 years of age gave their written informed consent.

From 1 October, 2010 until 1 March, 2015, all eligible children were asked to participate in this prospective study. Children were either included at dilated cardiomyopathy diagnosis or were followed-up for a previously diagnosed dilated cardiomyopathy in one of the participating tertiary paediatric cardiology centres. Dilated cardiomyopathy was defined as fractional shortening ≤25% and left ventricular end-diastolic dimension z-score >2 for body surface area. Dilated cardiomyopathy could be idiopathic or secondary to other causes. Patients with CHD, neuromuscular disease, or with parents who were unable to read the Dutch language were excluded.

Study entry was defined as the first time that a health-related quality-of-life questionnaire was completed. Patients were seen at 3–6-month intervals. At each visit, parents were asked to complete a health-related quality-of-life questionnaire, and during the same visit the paediatric cardiologist completed the New York University Pediatric Heart Failure Index. [9] This index assesses heart failure severity on the basis of symptoms and medications used. The score ranges from 0 to 30; a higher score represents more severe heart failure. Demographics were recorded, and the socioeconomic status was determined using parents' occupation and categorised into the following: low, elementary occupations; middle, middle-level occupations; or high, high-level scientific occupations, according to the Dutch classification system. [11] The highest occupation of either parent was recorded. Follow-up ended either at 15 September, 2015 or when a patient reached the age of 18 years or at the combined primary end point of death and heart transplantation.

Health-related quality-of-life questionnaires

Health-related quality of life was assessed by age-specific questionnaires: the Infant Toddler Quality of Life questionnaire for patients aged 0–4 years and the Child Health Questionnaire-Parent Form 50 for patients aged 4–18 years. Both questionnaires consisted of subscales (Table 2a and 2b). Subscale scores ranged from 0 to 100, with a higher score representing better quality of life. Normative data from Dutch healthy children are available for both questionnaires. [12, 13] Health-related quality of life was evaluated on two different time points in the disease course. First, in patients at dilated cardiomyopathy diagnosis and second in patients after 1 year or more since diagnosis. These time points were chosen, because event rates in paediatric dilated cardiomyopathy differ markedly between the 1st year of diagnosis and from 1 year after diagnosis onwards, [14]

and health-related quality of life may differ according to parents and patients who need to cope with a recent diagnosis, compared with patients who have been diagnosed a long time ago. To compare both age groups (0–4 and 4–18 years) and to predict outcome, individual subscale scores were transformed to percentage of predicted using the mean of the corresponding normal population. Using this transformation, only scores on comparable subscales from both questionnaires were combined – that is, "physical functioning", "bodily pain", "general behaviour", "general health perception", "parental impact – time", "parental impact – emotional", and "family cohesion".

Statistical Analysis

The distribution of continuous variables was tested using the Kolmogorov–Smirnov test. Almost all health-related quality-of-life subscales were non-normally distributed, and are therefore reported as medians and interquartile ranges (IQR). The medians of patients were compared with normal values using the one-sample Wilcoxon Signed Rank Test. To compare age groups, medians – as percentage of predicted – were compared using the Mann– Whitney U-test. Using univariable time-dependent Cox regression analysis, we assessed the predictive value of the health-related quality-of-life subscales – as percentage of predicted – and the New York University Pediatric Heart Failure Index at the end point. For this analysis, data of all visits were included (n = 515 in 90 different patients). The maximum number of covariates used in the multivariable time-dependent Cox regression analysis was the number of events divided by 10. Proportional hazard assumptions were tested and were not violated. The hazard ratios of the health-related quality-of-life subscales were calculated per 10% of the predicted values (10 units of the original scale). For readability, hazard ratios of health-related quality-of-life subscales were transformed to

values >1.00, using the following formula: 1/hazard ratio. For descriptive data analyses, we used IBM SPSS Statistics for Windows, version 21.0 (IBM Corp., Armonk, New York, United States of America). For advanced statistical analyses of repeated measurements and survival data, R environment was used (R version 3.1.1, 2014-07-10). Testing was performed using two-sided tests, and statistical significance was defined as $p < 0.05$.

Table 1: Cross-sectional characteristics of children with dilated cardiomyopathy at study entry (n= 90), diagnosis (n= 46), and ≥1 year after diagnosis (n=77)

	All patients, study entry (n=90)	At diagnosis (n=46)	≥1 year after diagnosis* (n=77)
Gender, male (n(%))	48 (53)	24 (52)	39 (51)
Age (years)	3.8 (0.9-12.3)	1.3 (0.4-7.0)	5.2 (1.8-12.7)
Time since DCM diagnosis (years)	0.5 (0.1-3.4)	0.1 (0.1-0.3)	1.5 (1.1-3.7)
Socioeconomic status (n(%))†			
Low	18 (22)	8 (19)	14 (19)
Middle	26 (32)	12 (29)	23 (32)
High	37 (46)	22 (52)	35 (49)
NYU PHFI	8 (6-11)	9 (6-11)	7 (4-9)
Follow-up time since first questionnaire (years)	2.8 (1.5-3.8)	2.5 (1.6-3.6)	3.0 (2.1-4.0)
Number of questionnaires per patient	6 (4-7)		
Number of ITQoL; Number of CHQ PF 50		33; 13	36;41

CHQ PF 50=Child Health Questionnaire-Parent Form 50; DCM=dilated cardiomyopathy; ITQoL=Infant Toddler Quality of Life questionnaire; NYU PHFI = New York University Pediatric Heart Failure Index

Continuous variables are represented as medians (interquartile range)

*In total, 34 children were also represented in the group "at diagnosis"

†Socioeconomic status was missing in nine cases

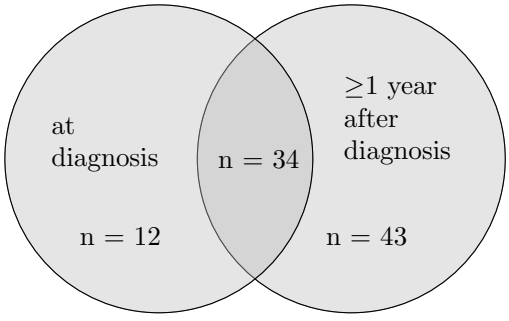


Figure 1: Venn diagram of patients included in this study. At diagnosis, parents of 46 children completed a health-related quality-of-life (HRQoL) questionnaire; ≥1 year after diagnosis, parents of 77 children completed a HRQoL questionnaire.

Results

We included 90 children in our study (median age 3.8 years, IQR 0.9–12.3, Table 1). Parents reported their child's health-related quality of life at several time points during follow-up. At the same visit, the physician scored heart failure severity on the New York University Pediatric Heart Failure Index. In total, 515 health-related quality-of-life questionnaires were completed over 4.5 years, 226 Infant Toddler Quality of Life questionnaires, and 312 Child Health Questionnaire-Parent Form 50, with a median of 6/patient (range 1–13). Accordingly, 498 New York University Pediatric Heart Failure Index ratings were completed, and 3.3% were missing. To analyse health-related quality of life at two time points in the disease, we describe the results of two cross-sectional groups – n=46 questionnaires of children included at dilated cardiomyopathy diagnosis and n=77 children whose parents completed a questionnaire at least 1 year after diagnosis; a total of 34 children were represented in both groups (Figure 1).

Table 2a. Health-related quality of life by parent reports: results of infants and toddlers, 0–4 years old, with dilated cardiomyopathy at diagnosis and ≥1 year after diagnosis.

ITGoL subscales	At diagnosis (n=33)	≥1 year after diagnosis (n=36)	Norm (n=410)
Physical functioning	**90** (77-100)†	98 (79-100)	97.2±9.8
Growth and development	**75** (66-85)‡	**79** (73-91)*	86.5±10.6
Bodily pain	**67** (35-83)‡	**75** (58-90)†	83.8±16.8
Temperament and moods	**69** (60-76)†	79 (67-86)	77.2±10.5
General behaviour	81 (67-89)	**78** (70-91)*	72.8±12.7
Getting along	69 (62-80)	**78** (69-86)*	71.4±8.8
General health perceptions	**39** (23-52)‡	**40** (33-59)‡	79.0±14.5
Parental impact - emotional	**71** (57-89)‡	89 (82-96)	92.1±10.5
Parental impact - time	**76** (67-86)‡	93 (82-100)	93.0±11.0
Family cohesion	**85** (85-100)*	85 (60-100)	75.3±18.8

ITQoL = Infant Toddler Quality of Life questionnaire
Higher scores represent better functioning. Patient values are presented as medians (interquartile range) and norm values as mean ± SD
p-value for comparison with age-specific norm values. Bold values are significantly different from norm values
*p-value <0.05; †p-value <0.01; ‡p-value <0.001

Health-related quality-of-life results at diagnosis

Of the 90 children, 46 were newly diagnosed with dilated cardiomyopathy (median age 1.3 year, IQR 0.4–7.0). Their first questionnaire was completed at a median of 1.4 months after diagnosis (IQR 1.1–3.1). In all, 33 children were between 0 and 4 years of age (Table 2a) and 13 children were between 4 and 18 years of age (Table 2b).

Comparison with the norm. At diagnosis, results of almost all subscales on both age-specific questionnaires were significantly lower compared with the normal population (Infant Toddler Quality of Life: 7/10, and Child Health Questionnaire-Parent Form 50: 8/11). Parents of children aged 0–4 years showed the largest difference compared with the normal population on "general health perception". Notably, better "family cohesion" was reported in this

Table 2b. Health-related quality of life by parent reports: results in children aged 4–18 years with dilated cardiomyopathy at diagnosis and ≥1 year after diagnosis.

CHQ PF50 subscales	At diagnosis (n=13)	≥1 year after diagnosis (n=41)	Norm (n=353)
Physical functioning	**50** (39-69)†	**83** (61-100)‡	99.1±4.3
Role functioning - emotional	61 (25-100)	100 (78-100)	97.9± 7.2
Role functioning - physical	**33** (33-67)†	**100** (67-100)*	95.8±15.6
Bodily pain	**50** (20-65)†	80 (60-100)	85.7±17.2
General behaviour	81 (68-85)	77 (66-85)	78.5±13.1
Mental health	**65** (58-78)†	**75** (65-90)*	81.4±12.1
Self-esteem	**58** (54-79)*	71 (58-83)†	79.2±11.0
General health perceptions	**60** (38-69)†	**43** (31-56)‡	82.9±13.4
Parental impact - emotional	**42** (17-75)†	**67** (58-83)‡	86.3±15.2
Parental impact - time	**44** (28-61)†	89 (67-100)	94.0±13.0
Family cohesion	60 (60-96)	**60** (60-85)*	72.2±19.4

CHQ PF50 = Child Health Questionnaire-Parent Form 50
Higher scores represent better functioning. Patient values are presented as medians (interquartile range) and norm values as mean ± SD
p-value for comparison with age-specific norm values. Bold values are significantly different from norm values
*p-value <0.05; †p-value <0.01; ‡p-value <0.001

age group (Table 2a). Parents of children aged 4–18 years showed the largest differences on "physical functioning", "role functioning – physical", "parental impact – emotional", and "parental impact – time" (Table 2b).

Comparison between age groups. At the time of diagnosis, we found that parents of older children (4–18 years) scored significantly worse than parents of young children (0–4 years) on the subscales "physical functioning", "parental impact – emotional", and "parental impact – time" (Figure 2).

Health-related quality-of-life results ≥1 year after diagnosis

Parents of 77 children completed a questionnaire at least 1 year after diagnosis, at a median time of 1.5 years after diagnosis (range 1–16 years, Table 1). Between age groups, the time since diagnosis was significantly different – patients aged

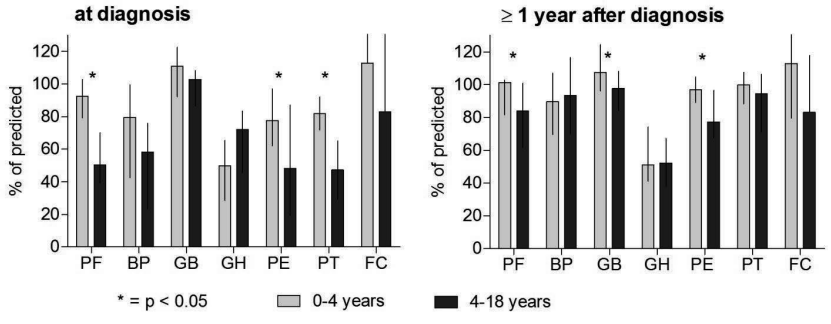


Figure 2: Differences between age groups. (a) Differences in health-related quality of life (HRQoL) at diagnosis between 33 infants and toddlers (0–4 years old) and 13 children (4–18 years old); results are percentages of predicted values. (b) Differences in HRQoL ≥1 year after diagnosis between 36 infants and toddlers (0–4 years old) and 41 children (4–18 years old); results are percentages of predicted values. BP = bodily pain; FC = family cohesion; GB = general behaviour; GH = general health perceptions; PE = parental impact-emotional; PF = physical functioning; PT = parental impact-time; * = $p<0.05$ between age groups.

0–4 years were at 1.2 years after diagnosis (IQR 1.0–1.6), whereas patients aged 4–18 years were at 3.4 years after diagnosis (IQR 1.3–7.8, p=0.004).

Comparison with the norm. Parents of children aged 4–18 years scored lower on more than half of the subscales (6/11), with the largest difference compared with the normal population on "general health perceptions". In contrast, parents of children aged 0–4 years had lower scores on three subscales – that is, "growth and development", "bodily pain", and "general health perceptions". Parents of young children with dilated cardiomyopathy scored their children better than the normal population on "general behaviour" and "getting along". The other subscales were comparable with the normal group.

Comparison between age groups. At least 1 year after diagnosis, we found that parents of older children scored their children significantly worse than younger children on "physical functioning", "general behaviour", and "parental

impact – emotional" (Figure 2). Notably, parents of young children scored their children higher than the normal population on "general behaviour", and parents of older children scored them comparable with the normal population.

Cardiac outcome and follow-up

In children included at diagnosis, n = 46, the median New York University Pediatric Heart Failure Index was 9 (IQR 6–11). For children who subsequently reached an end point (n = 4), the median New York University Pediatric Heart Failure Index was 11 (IQR 9–14) compared with 9 (IQR 6–11) for those without an end point. At least 1 year after diagnosis, the median New York University Pediatric Heart Failure Index was 7 (IQR 4–9). For children who subsequently reached an end point (n = 15), the median New York University Pediatric Heart Failure Index was 11 (IQR 8–12) compared with 6 (IQR 3–9) for those without an end point (n = 62). The median follow-up time since the first questionnaire to the end of the study or an end point was 2.8 years (IQR 1.5–3.8). During the study, 15 patients reached an end point – all were transplanted (1.3 years (IQR 0.9–2.2) since completing the first questionnaire; 3.2 years (IQR 2.5–6.2) since diagnosis). All 15 children are included in the cross-sectional group >1 year after diagnosis (n = 77). In the group of newly diagnosed children (n = 46), four children reached an end point – all after 1 year since diagnosis. Of these 15 children, 87% had a Class I and 13% a Class IIa indication for heart transplantation at the time of listing and at the time of transplantation. At the time of listing, 20% had Stage D heart failure, and all of them were dependent on inotropes. At the time of transplantation, 40% had Stage D heart failure – four patients were on mechanical circulatory support and two patients were dependent on inotropes.

Table 3. Results of univariable and multivariable time dependent Cox regression analyses

Model	Variables	Coefficient	HR	(95% CI)	p-value
Univariable					
	NYU PHFI (per unit)	0.40	1.49	(1.32-1.67)	< 0.001
	QoL subscales (per 10 % of predicted)*				
	Physical functioning	-0.42	1.53	(1.38–1.69)	<0.001
	Bodily pain	-0.38	1.46	(1.26–1.68)	<0.001
	General behaviour	0.01	0.99	(0.75–1.30)	0.95
	General health perceptions	-0.68	1.97	(0.98–4.00)	0.06
	Parental impact – emotional	-0.39	1.48	(1.32–1.68)	<0.001
	Parental impact – time	-0.35	1.42	(1.29–1.58)	<0.001
	Family cohesion	-0.12	1.13	(0.91–1.41)	0.27
Multivariable					
	Physical functioning*	-0.22	1.24	(1.06–1.47)	0.01
	NYU PHFI (per unit)	0.32	1.38	(1.19–1.61)	<0.001

CI=confidence interval; HRQoL =health-related quality of life; NYU PHFI= New York University Pediatric Heart Failure Index
*For readability, 1/HR are presented

Predictors for outcome

For predicting the risk of death and transplantation, all available measurements were used – that is, 515 health-related quality-of-life questionnaires and 498 New York University Pediatric Heart Failure Index results in 90 different patients including 15 end points. Using univariable time-dependent Cox regression, the subscales "physical functioning", "bodily pain", "parental impact – emotional", "parental impact – time", and the New York University Pediatric Heart Failure Index were each significant predictors for the risk of death and heart transplantation. For the multivariable model, "physical functioning" was used as it reflects the child's actual physical ability and had the highest hazard ratio in univariable analysis. The multivariable model showed that "physical func-

tioning" and the New York University Pediatric Heart Failure Index were both independently predictive of the risk of death and heart transplantation (Table 3). A decrease in physical functioning by 10% of the predicted value resulted in a hazard ratio of 1.24 (95% confidence interval (CI) 1.06–1.47), indicating a 24% higher risk for a patient with a score of 80% versus a patient with a score of 90% of the predicted value; one point higher score on the New York University Pediatric Heart Failure Index resulted in a 38% higher risk of death and heart transplantation (hazard ratio 1.38, 95% CI 1.19–1.61).

Discussion

This is the first study that systematically investigated health-related quality of life and the New York University Pediatric Heart Failure Index in a relatively large cohort of children with dilated cardiomyopathy. It clearly demonstrates that health-related quality of life is severely impaired, and that parent reported "physical functioning" and the New York University Pediatric Heart Failure Index as assessed by the physician are independently predictive of the risk for death and heart transplantation.

At diagnosis, patients of both age groups scored worse on physical, psychosocial, and parental impact subscales compared with normal values. Older children scored significantly worse than younger children. More than 1 year after diagnosis, health-related quality of life was still impaired, but to a lesser extent than at diagnosis, and again was more impaired in older than in younger children.

The differences between age groups may have several explanations. First, impairments may be more obvious in older than in younger children because their daily-life activities and range of skills are more diverse. Moreover, older children are normally more independent, but when they become ill, parents

need to accept their caretaking role and be more in control again, which may be disruptive for family routines. In contrast, parents of young children are used to an active caregiving role during daily life, whether their children are healthy or diseased. This shift in the locus of control has previously been described in older children with chronic illnesses. [15] Second, older children are cognitively able to realise and experience the impact of the disease themselves, as demonstrated by the lower scores on "mental health" and "self-esteem". Thus, parents of older children have to cope with more physical and psychosocial impact than parents of young children. [16] This effect was demonstrated by the larger effect on parental impact in patients aged 4–18 years, both at diagnosis and ≥1 year after diagnosis. Third, ≥1 year after diagnosis, older patients had dilated cardiomyopathy for a longer period and may have been "growing into deficit". This phenomenon has been described in children with other diseases, and means that psychological problems on higher cognitive functions, such as emotion regulation, may develop over time, because these functions need to mature. [17] Finally, it may also be related to the severity of heart failure. Of the 77 children studied ≥1 year after diagnosis, 15 reached the end point, of whom 10 were >4 years old. Furthermore, highest recovery rates have been described in children aged 1–6 years; [1] thus, the group with younger children may include more children who eventually recover. Considering these results, patients at highest risk for psychological problems – that is, those at diagnosis and older children with chronic disease (≥1 year after diagnosis) – may benefit most from timely referral to a psychosocial support team. As we described two cross-sections in which we had no complete cases, we cannot draw firm conclusions about the development of health-related quality of life from diagnosis to >1 year after diagnosis. Nevertheless, we speculate that health-related quality of life improves after the 1st year of diagnosis. Our data clearly showed the severe

impairment at diagnosis. Scores on several subscales were also impaired ≥1 year after diagnosis, but then the difference from the norm was less extreme, and especially in the young age-group several subscales were comparable with the norm. This improvement was not explained by the number of children who reached an end point, because all 15 children with adverse outcome were represented in the group ≥1 year after diagnosis. This indicates that parents may be adapting to the knowledge that their child has dilated cardiomyopathy and may rate their child's disabilities with different intensity. This phenomenon may be explained by response shift, which means that parents change their internal standards towards health-related quality of life in case of chronical illness. [18] This has also been described in children with sequelae of complex CHD who rate their health-related quality of life on some subscales as normal as compared with healthy controls. [19] Another factor, which may contribute to the improvement of health-related quality-of-life scores in the young age group is a high recovery rate. Previously, we reported a recovery rate of 69% in 1–6-year-olds at a median time of 1 year after diagnosis. [1] We suspect that the improvement in clinical condition accompanying this recovery is also reflected in the health-related quality-of-life scores in the young age group.

Previous studies in adults with heart failure have shown that self-reported health-related quality of life was predictive of mortality. [8] As far as we know, this is the first study in children with dilated cardiomyopathy showing that health-related quality of life, as reported by parents, was predictive of the risk of death and heart transplantation. Moreover, we demonstrated for the first time that the New York University Pediatric Heart Failure Index, as assessed by the physician, was predictive of the risk of death and heart transplantation. Earlier reports in adults and children have shown that the presence of congestive heart failure and higher NYHA functional class were related to adverse outcomes. [2,

20] The direct association between NYHA and physical health-related quality of life is a limitation for the use of both markers in the prediction of outcome. [8] The New York University Pediatric Heart Failure Index may be a more discriminative measure of functional status in children, because it is a 30-points index focussing on heart failure symptoms and medication use, rather than patients' physical functioning. [9] In this study, we demonstrated in multivariable analysis that both the New York University Pediatric Heart Failure Index as well as the health-related quality-of-life parameter "physical functioning" independently predicted outcome. We obtained health-related quality of life and the New York University Pediatric Heart Failure Index frequently during follow-up and found that their predictive values were constant over time. Therefore, these two predictors can be used from diagnosis onwards and during follow-up in paediatric dilated cardiomyopathy.

In the present study, no deaths occurred and all end points were reached more than 1 year after diagnosis. This is in line with our previous report, indicating a conservative approach to listing for transplantation. [1] We have shown a low transplantation rate in the 1st year after diagnosis without an increase in mortality as compared with other cohorts. In the next few years, transplantation rates were comparable with other cohorts. Listing strategies in general followed the American Heart Association guidelines. [21] According to these recommendations, 83% had a Class I and 13% a Class IIa indication at listing for transplantation, underscoring the severity of disease in children who underwent transplantation.

The few studies that have been performed concerning health-related quality of life in children with dilated cardiomyopathy have included mainly small cohorts. [10, 22–24] The group of Menteer described reduced health-related quality of life in two small subgroups of children with heart failure (n = 15 and n

= 11), but used another health-related quality-of-life questionnaire, which limits comparison with our results. [22, 23] Walker et al performed an explorative study in the out-patient clinic and included a sub-group of 17 children with cardiomyopathy aged 5–17 years. [10] They found significantly lower scores on "physical functioning", "general health perception", and "parental impact – emotional", in line with our findings. They reported a significantly higher score on "family cohesion", which is in contrast with the results in the older age group of our cohort. Nevertheless, "family cohesion" was better in infants and toddlers in our study. Clinical experience shows that the seriousness of the disease may either "bring families closer together" or "tear them apart". Finally, the Pediatric Cardiomyopathy Registry reported limited results on the Child Health Questionnaire-Parent Form 50 in children with cardiomyopathy. [24] On average, they reported impaired health-related quality of life, with more physical problems than psychosocial problems, and suggested improvement over time in functional status. Finally, they suggested that poorer functional status might be a risk factor for subsequent death and heart transplantation. Our study adds to the existing data by clearly demonstrating the predictive value of functional status on outcome, by demonstrating improvement over time, but less in older children and by demonstrating the independent predictive value of a paediatric heart failure score on outcome.

Limitations

This study had some limitations. First, the number of events was only 15, limiting the number of variables in the multivariable analysis to only two. The "physical functioning" subscale was most relevant, but it would be interesting to test other significant subscales. Similarly, it would be worthwhile to study other variables, besides the New York University Pediatric Heart Failure Index,

such as biomarkers or echocardiographic parameters, but it requires a larger cohort with more end points. Second, the median follow-up time was almost 3 years. Therefore, the outcome results need to be interpreted at a mid-term follow-up time. Finally, the treating physicians who recorded the New York University Pediatric Heart Failure Index scores were not blinded to the results. However, these were not registered in the clinical file of the patients, and were not a part of the clinical evaluation and treatment decisions. Therefore, it is unlikely that this has caused bias in eligibility for transplantation decisions.

Conclusions

In children with dilated cardiomyopathy, health-related quality of life is severely impaired at diagnosis and ≥1 year after diagnosis. Children ≥4 years of age had lower health-related quality of life than children <4 years of age. "Physical functioning" as reported by parents and heart failure severity using the New York University Pediatric Heart Failure Index are independent predictors for death and heart transplantation. Our findings corroborate the use of such parameters in, composite, end points in future studies in paediatric dilated cardiomyopathy.

References

1. Den Boer, S. L. *et al.* Management of children with dilated cardiomyopathy in The Netherlands: Implications of a low early transplantation rate. *The Journal of Heart and Lung Transplantation* **34,** 963–969 (2015).
2. Towbin, J. A. *et al.* Incidence, Causes, and Outcomes of Dilated Cardiomyopathy in Children. *JAMA* **296,** 1867–1876 (2006).
3. Gradman, A. *et al.* Predictors of total mortality and sudden death in mild to moderate heart failure. Captopril-Digoxin Study Group. *J Am Coll Cardiol* **14,** 564–70, 564–70 (1989).
4. Dolgin, M., New York Heart, A. & Criteria, C. *Nomenclature and criteria for diagnosis of diseases of the heart and great vessels* ISBN: 0316605387 9780316605380 (Little Brown, Boston; New York, 1994).

5. Hoekstra, T., Lesman-Leegte, I., van Veldhuisen, D. J., Sanderman, R. & Jaarsma, T. Quality of life is impaired similarly in heart failure patients with preserved and reduced ejection fraction. *European Journal of Heart Failure* **13,** 1013–1018 (2011).
6. Lesman-Leegte, I. *et al.* Quality of Life and Depressive Symptoms in the Elderly: A Comparison Between Patients With Heart Failure and Age- and Gender-Matched Community Controls. *Journal of Cardiac Failure* **15,** 17–23 (2009).
7. Juenger, J. *et al.* Health related quality of life in patients with congestive heart failure: comparison with other chronic diseases and relation to functional variables. *Heart* **87,** 235–41 (2002).
8. Hoekstra, T. *et al.* Quality of life and survival in patients with heart failure. *European Journal of Heart Failure* **15,** 94–102 (2013).
9. Connolly, D., Rutkowski, M., Auslender, M. & Artman, M. The New York University Pediatric Heart Failure Index: A new method of quantifying chronic heart failure severity in children. *The Journal of Pediatrics* **138,** 644–648 (2001).
10. Walker, R. E., Gauvreau, K. & Jenkins, K. J. Health-related quality of life in children attending a cardiology clinic. *Pediatric Cardiology* **25,** 40–8 (2004).
11. Statistics Netherlands [Centraal Bureau voor Statistiek] Den Haag/ Heerlen, T. N. *Dutch Standard Classification of Occupations [Standaard Beroepenclassificatie]*
12. Raat, H., Bonsel, G. J., Essink-Bot, M.-L., Landgraf, J. M. & Gemke, R. J. B. J. Reliability and validity of comprehensive health status measures in children: The Child Health Questionnaire in relation to the Health Utilities Index. *Journal of Clinical Epidemiology* **55,** 67–76 (2002).
13. Raat, H., Landgraf, J. M., Oostenbrink, R., Moll, H. A. & Essink-Bot, M. L. Reliability and validity of the Infant and Toddler Quality of Life Questionnaire (ITQOL) in a general population and respiratory disease sample. *Quality of Life Research* **16,** 445–60 (2007).
14. Alexander, P. M. A. *et al.* Long-Term Outcomes of Dilated Cardiomyopathy Diagnosed During Childhood: Results From a National Population-Based Study of Childhood Cardiomyopathy. *Circulation* **128,** 2039–2046 (2013).
15. Tong, E. M. & Kools, S. Health care transitions for adolescents with congenital heart disease: patient and family perspectives. *Nursing Clinics of North America* **39,** 727–740 (2004).
16. Uzark, K. & Jones, K. Parenting stress and children with heart disease. *Journal of Pediatric Health Care* **17,** 163–168 (2003).
17. Aarsen, F. K. *et al.* Functional outcome after low-grade astrocytoma treatment in childhood. *Cancer* **106,** 396–402 (2006).
18. Rapkin, B. D. & Schwartz, C. E. Toward a theoretical model of quality-of-life appraisal: Implications of findings from studies of response shift. *Health Qual Life Outcomes* **2,** 14 (2004).
19. Dulfer, K. *et al.* Does functional health status predict health-related quality of life in children after Fontan operation? *Cardiol Young* **26,** 459–68 (2016).

20. Nohria, A. *et al.* Clinical assessment identifies hemodynamic profiles that predict outcomes in patients admitted with heart failure. *Journal of the American College of Cardiology* **41,** 1797–1804 (2003).

21. Canter, C. E. *et al.* Indications for heart transplantation in pediatric heart disease: a scientific statement from the American Heart Association Council on Cardiovascular Disease in the Young; the Councils on Clinical Cardiology, Cardiovascular Nursing, and Cardiovascular Surgery and Anesthesia; and the Quality of Care and Outcomes Research Interdisciplinary Working Group. *Circulation* **115,** 658–76 (2007).

22. Menteer, J., Beas, V., Chang, J., Reed, K. & Gold, J. Mood and Health-Related Quality of Life Among Pediatric Patients With Heart Failure. *Pediatric Cardiology,* 1–7 (2012).

23. Sabati, A. A. *et al.* Survival and quality of life for children with end-stage heart failure who are not candidates for cardiac transplant. *The Journal of Heart and Lung Transplantation* **34** (7 2015).

24. Wilkinson, J. D. *et al.* The Pediatric Cardiomyopathy Registry and Heart Failure: Key Results from the First 15 Years. *Heart Failure Clinics* **6,** 401–413 (2010).

Chapter 9

Electrical conduction dynamics after transcatheter aortic valve implantation

van Gils L, **Baart SJ**, Kroon H, Rahhab Z, El Faquir N, Rodriguez Olivares R, Aga Y, Maugenest A, Theuns DA, Boersma H, Szili Torok T, de Jaegere PPT, van Mieghem NM

Europace - 2018; 20:1981-1988

Abstract

Aim: To correlate dynamics in electrical conduction after transcatheter aortic valve implantation (TAVI) with need for permanent pacemaker implantation (PPM) and assess implications for early discharge.

Methods and Results: Daily electrocardiograms after TAVI were analysed for rhythm and conduction times and were correlated with PPM. Transcatheter aortic valve implantation was performed in 291 consecutive patients with three contemporary transcatheter heart valve designs: Medtronic CoreValve (n = 111), Edwards Sapien XT (n = 29) and Sapien 3 (n = 72), and Boston Lotus (n = 79). We considered two cohorts: (A) Patients with normal baseline conduction; and (B) patients with pre-existent conduction disturbances. Based on QRS dynamics, three patterns were discerned: stable normal QRS duration, transient QRS prolongation, and persistent QRS prolongation. In Cohort B, QRS dynamics did not correlate with PPM. In contrast, in Cohort A, QRS dynamics and PPM appeared highly correlated. Neither patients with stable normal QRS duration (0/47), nor patients with transient QRS prolongation required PPM (0/26). All PPMs occurred in patients with persistent QRS prolongation until discharge (27/85). Persistent QRS prolongation was typically seen with Lotus and CoreValve, whereas stable normal QRS duration was typically seen with Sapien XT and Sapien 3.

Conclusions: Three distinct patterns of QRS dynamics can be discerned after TAVI and their predictive probabilities for PPM strongly relate to the baseline conduction status. Patients with normal conduction at baseline and stable QRS duration after TAVI are potentially eligible for early discharge.

Introduction

Transcatheter aortic valve implantation (TAVI) has evolved into an attractive, minimally invasive alternative to surgical aortic valve replacement for patients with severe aortic stenosis and intermediate or greater surgical risk. [1–3] Not only the procedure itself, but also hospital stay has shortened to extremes of same day discharge in some instances. [4] Electrical conduction disturbances and need for permanent pacemaker implantation (PPM) are frequent after TAVI, [5] and imposes an important obstacle for early discharge after TAVI.

Conduction disturbances are more common with self-expanding and mechanically expanded transcatheter heart valves (THVs) compared to balloon-expandable valves. [6, 7] Apart from THV design, several baseline predictors for post-procedural conduction disturbances have been identified (e.g. pre-existing conduction disturbances, excessive device oversizing relative to the annular root dimensions, and depth of implantation). [6, 8]

Newly acquired conduction disturbances do not always persist since half of patients who received a PPM are no longer pacemaker-dependent at long-term follow-up. [8, 9] Therefore, the decision for either safe early discharge or monitoring by telemetry and potential PPM implantation poses a current challenge.

Electrical conduction after TAVI may be dynamic and device dependent. Proper understanding of these dynamics is clinically relevant and may help guide patient management and facilitate early discharge. The electrocardiogram (ECG) immediately after TAVI already provides information to determine whether patients might be eligible for early discharge. [10] The purpose of this study was to assess conduction times (i.e. QRS-duration) during the entire admission after TAVI, in order to identify dynamic patterns and to correlate with need for permanent pacemaker dependency.

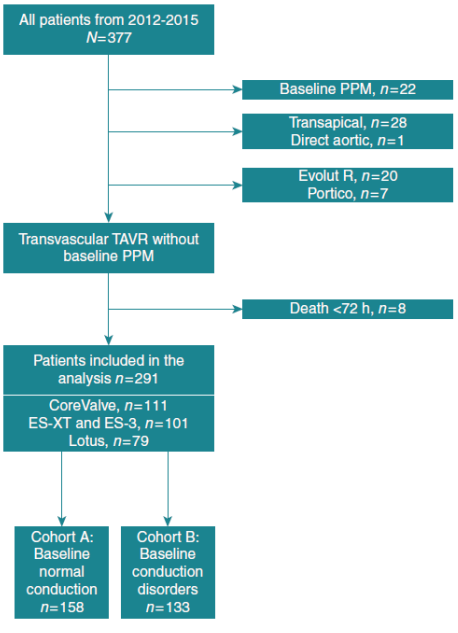


Figure 1: A flowchart of study inclusion. ES-XT, Edwards Sapien XT; ES-3, Edwards Sapien 3; PPM, permanent pacemaker implantation; TAVR, transcatheter aortic valve replacement.

Methods

All consecutive patients who underwent transarterial (transfemoral or transsubclavian) TAVI between January 2012 and December 2015 in our centre were entered in a prospective database. This study complied with the Declaration of Helsinki. All patients provided written informed consent for the procedure and data analysis for research purposes per Institutional Review Board approval. This study was not subject to the Dutch Medical Research Involving Human Subjects Act, which was confirmed by the local medical ethics committee of the Erasmus MC Rotterdam.

An overview of inclusion is illustrated in Figure 1. Patients who died within 72 h after the procedure were excluded from the analysis. Three THV-designs

were used: CoreValve (n=111) (Medtronic, Minneapolis, MN, USA), Sapien XT (n=29) and Sapien 3 (n=72) (Edwards Lifesciences, Irvine, CA, USA), and Lotus (n=79) (Boston Scientific Corporation, Marlborough, MA, USA).

For the purpose of this study, we considered two cohorts: Cohort A consisted of patients with untainted conduction, whereas Cohort B consisted of patients with pre-existent conduction disturbances (i.e. first or degree AVB, hemiblock, or bundle branch block).

Twelve-lead ECGs were collected prior to TAVI and daily afterwards up to discharge to a maximum of 14 days and at 1 month follow-up at the outpatient clinic visit. ECGs were interpreted by dedicated clinical researchers (L.V.G.,H.K., and Y.A.). If necessary, an experienced cardiologist (N.V.M.) was consulted for consensus. The ECGs were analysed for rhythm, conduction times, and the presence of AVB or bundle branch block. Conduction times were derived from digitalized ECGs with a chart speed of 25mm/s. Computer-calculated conduction times were used, since they show less variability compared with manual caliper methods. [11] Only ECGs without pacemaker intrusion were included in our analysis. In presence of multiple ECGs on the same day, the ECG with the longest calculated QRS-duration was selected.

QRS-prolongation of 20ms was considered a significant change. Three conduction patterns were discerned: (i) stable: QRS-duration after TAVI did not prolong by >20ms; (ii) transient: QRS duration after TAVI prolonged by >20ms but at the discharge ECG the QRS duration narrowed again within 20ms; (iii) persistent: QRS duration after TAVI prolonged by >20ms and the QRS duration at discharge persisted at least 20ms beyond baseline.

The primary outcomes of this study were: (i) New onset high degree atrioventricular block (AVB) and (ii) need for PPM. The decision for PPM was per treating physician's discretion, although agreed by an electrophysiologist and

in general in compliance with contemporary European Society of Cardiology guidelines on permanent pacemaker implantation. [12] Patients who receive a permanent pacemaker in our institution systematically visit the pacemaker technician at 10 days and 6 months after implantation at the outpatient clinic. At 10 days the anterograde and retrograde properties of the atrioventricular conduction are assessed with pacing manoeuvres, after which the pacemaker settings are adapted in order to prioritize the native conduction system. In patients without a total AVB at 10 days, the device is programmed in DDD-mode with an algorithm which prefers the native conduction system (i.e. paced AV-delay of 200ms and sensed AV-delay of 150ms and activation of the algorithm). In patients with a total AVB or a low Wenckebach point (<120 b.p.m.) the device is programmed in standard DDD-mode. A PPM interrogation at 6 months assesses pacing percentage. For the purpose of this study, patients with less than 20% ventricular pacing over 6 months of followup — which is suggestive for independency of ventricular pacing [13]—were then re-adjudicated by an electrophysiology expert (D.T.) for pacemaker dependence. Depending on whether there was normal intrinsic atrioventricular conduction pacing was labelled as 'dependent' or 'independent'.

Continuous variables were presented as mean ± standard deviation or median [interquartile range (IQR)]. The distribution of continuous variables was assessed for normality with histograms and the Shapiro–Wilk test. Comparisons between the ECG conduction patterns for repeatedly measured continuous variables were done using a repeated measures analysis of variance with post hoc Tukey adjustment. The assumption of homogeneity of variances was tested with the Levene's test. Categorical variables were expressed as absolute counts plus percentages and were compared by use of the Pearson χ^2 test and Z-test for proportions.

All statistical analyses were performed with SPSS version 21.0.1 (IBM Corp, Armonk, NY, USA). A two-sided P-value of <0.05 was considered statistically significant.

Results

A total of 291 consecutive patients underwent transfemoral (94%) or transsubclavian (6%), TAVI with CoreValve (38%), Sapien XT or Sapien 3 (35%), or Lotus (27%). Mean age was 79±8 years, 46% were female. Balloon predilatation was performed in 51% of patients. Half of the patients (54%) did not have pre-existent conduction disturbances (Cohort A) whereas the other half (46%) did (Cohort B). Baseline characteristics between Cohorts A and B were well balanced (Table 1). AV1B was the most common conduction disturbance (21%), followed by left anterior fascicular block (16%), left bundle branch block (LBBB) (12%) and right bundle branch block (11%).

Conduction related outcomes are displayed in Table 2. Intraprocedural LBBB and high degree AVB were common (67% and 33%, respectively). Delayed high degree AVB (i.e. high degree AVB that first presented after the patients had left the catheterization laboratory) occurred in 8% of patients. At 30 days, overall 23% required PPM; 17% in Cohort A compared with 29% in Cohort B (P=0.013). Indication for PPM was almost exclusively a high degree AVB (94%). None of the patients had a documented high degree AVB after discharge nor did any of them need a PPM during 1 year follow-up. In the majority, the percentage of pacing that was provided was either less than 10% or more than 90% of the time (Table 2, Supplementary material online, Figure S1). Pacemaker interrogations at 6 months follow-up were completed in 74% of patients (17 patients were lost to follow-up). The interrogation revealed <20% paced rhythm (i.e. pacemaker independent) in 39% of patients. The amount of

Table 1: Baseline characteristics

	Cohort A ($n = 158$)	Cohort B ($n = 133$)	Overall ($n = 291$)	P value
Age (years)	78 ± 8	80 ± 7	79 ± 8	0.116
Male gender	77 (49)	79 (59)	156 (54)	0.069
Body mass index (kg/m^2)	27 ± 5	27 ± 5	27 ± 5	0.860
Body surface area (m^2)	1.86 ± 0.20	1.86 ± 0.20	1.86 ± 0.20	0.966
Creatinine level (μmol/L)	99 (74–128)	94 (77–117)	96 (77–122)	0.463
Renal dialysis	7 (5)	4 (4)	11 (4)	0.759
DM	50 (32)	48 (36)	98 (34)	0.424
Hypertension	134 (85)	104 (78)	238 (82)	0.145
Log EuroSCORE (%)	11 (7–18)	12 (9–19)	12 (8–19)	0.102
NYHA class				
I	4 (3)	3 (2)	7 (3)	0.987
II	36 (24)	28 (23)	64 (23)	
III	95 (63)	80 (65)	175 (64)	
IV	16 (11)	12 (10)	28 (10)	
Atrial fibrillation	48 (30)	32 (24)	80 (28)	0.229
Conduction disturbance[a]				
None	158 (100)	0 (0)	158 (54)	NA
AV1B	0 (0)	61 (46)	61 (21)	
LBBB	0 (0)	35 (27)	35 (12)	
RBBB	0 (0)	33 (25)	33 (11)	
LAFB	0 (0)	47 (35)	47 (16)	
LPFB	0 (0)	6 (5)	6 (2)	
Access				
Transfemoral	152 (96)	122 (92)	274 (94)	0.105
Trans-subclavian	6 (4)	11 (8)	17 (6)	
Pre-dilatation	79 (50)	70 (53)	149 (51)	0.651
THV				
CoreValve	58 (37)	53 (40)	111 (38)	0.813
ES-XT or ES-3	55 (35)	46 (35)	101 (35)	
Lotus	45 (49)	34 (26)	79 (27)	

Categorical variables are displayed as counts (%). Continuous variables are displayed as mean ± SD or median (interquartile range).

AV1B, first degree atrioventricular block; DM, diabetes mellitus; ES-3, Edwards Sapien 3; ES-XT, Edwards Sapien XT; LAFB, left anterior fascicular block; LBBB, left bundle branch block; LPFB, left posterior fascicular block; NA, not applicable; NYHA, New York Heart Association; RBBB, right bundle branch block; SD, standard deviation; THV, transcatheter heart valve.

[a]Patients without conduction disturbances at baseline represent Cohort A and patients with conduction disturbances represent Cohort B.

ventricular pacing did not differ between Cohorts A and B (P=0.459).

In Cohort A, conduction patterns correlated with PPM, regardless of THV type (Figure 2). In Cohort B, high degree AVB and PPM appeared unrelated to QRS patterns (Figure 3).

Changes in QRS duration for the three discerned patterns in Cohort A are displayed in Supplementary material online, Table S1 and Figure S2. In patients with a stable QRS duration, the mean QRS duration at baseline was 95±11 ms. In patients with transient QRS prolongation, mean QRS duration at baseline was 100±9 ms and prolonged up to 144±15 ms, after which it narrowed to 103±10 ms at discharge. On aggregate, mean QRS duration at follow-up was 102±11 ms. In patients with persistent QRS prolongation, the mean QRS duration at baseline was 97±11 ms and prolonged up to 157±15 ms, and it remained broad (152±17) until discharge. At follow-up, the QRS-duration was still broad (137±23 ms). It took 1 day to reach the maximum QRS duration in the majority of patients (61%) irrespective of THV design or the transient or persistent nature [transient 0.5 (IQR 0–4) day; persistent 1 (IQR 0–3) day]. The transient pattern became apparent at 1 day post-TAVI (IQR 1–4), and the QRS interval normalized up to Day 6.

Implantation of a Lotus valve was typically followed by persistent QRS prolongation (69%), in contrast to stable normal QRS duration (7%) or transient QRS prolongation (24%). Conversely, half of the patients treated with the Sapien XT and Sapien 3 had a stable normal QRS duration (51%) compared with transient QRS prolongation (13%) and persistent QRS prolongation (36%).With CoreValve, persistent QRS prolongation was most common (57%) in contrast to stable normal QRS duration (28%) or transient QRS prolongation (16%).

In Cohort A, patients with a stable normal QRS duration and patients with transient QRS prolongation never required a PPM, although high degree

AVB appeared and resolved in 28% and 35%, respectively. In patients with a persistent QRS prolongation high degree AVB appeared in 52% (44/84) and 32% (27/84) required a PPM. At the 6 month pacemaker interrogation 41% was independent of their pacemaker.

Table 2: Conduction related and clinical outcomes

	Cohort A (N = 158)	Cohort B (N = 133)	Overall (N = 291)	P value
Number of days to discharge	8 (6–11)	8 (6–12)	8 (6–11)	0.905
30 day mortality	8/157 (5)	6/133 (5)	14/290 (5)	0.817
1 year mortality	23/151 (15)	19/121 (16)	42/272 (15)	0.915
Intra-procedural new LBBB	105/133 (74)	48/84 (57)[a]	153/217 (67)	0.007
Intra-procedural new AV3B	35/133 (25)	47/133 (43)	82/291 (33)	0.002
PPM at 30 days[b]				
All THVs	27/158 (17)	39/133 (29)	66/291 (23)	0.013
CoreValve	15/58 (26)	16/53 (30)	31/111 (28)	0.621
ES-XT or ES-3	5/55 (9)	11/46 (24)	16/101 (16)	0.042
Lotus	7/45 (16)	12/34 (35)	19/79 (24)	0.042
Indication for PPM[c]				
High degree AVB	26/27 (96)	36/39 (92)	62/66 (94)	0.094
Sick sinus syndrome	1/27 (4)	2/39 (5)	3/66 (5)	
Trifascicular block	0/27 (0)	1/39 (3)	1/66 (2)	
Number of days to PPM	6 (1–8)	4 (1–8)	5 (1–8)	0.377
Percentage pacing at 6 months interrogation[d]				
0	2/22 (9)	4/27 (15)	6/49 (12)	0.459
1–20	7/22 (32)	6/27 (22)	13/49 (27)	
21–99	6/22 (27)	12/27 (44)	18/49 (37)	
100	7/22 (32)	5/27 (19)	12/49 (25)	

Categorical variables are displayed as counts (n/N) (%). Continuous variables are displayed as median (interquartile range).
AV3B, third degree atrioventricular block; AVB, atrioventricular block; LBBB, left bundle branch block; PPM, permanent pacemaker implantation; THV, transcatheter heart valve.
[a]Patients with LBBB at baseline are herein excluded.
[b]N is total number of patients treated with that particular valve.
[c]N is total number of patients with a PPM.
[d]Pacemaker interrogations were performed in 49 of 66 patients (17 patients were lost to follow-up).

QRS-patterns determined at daily intervals and subsequent conduction related events are listed in see Supplementary material online, Figure S2. Patients

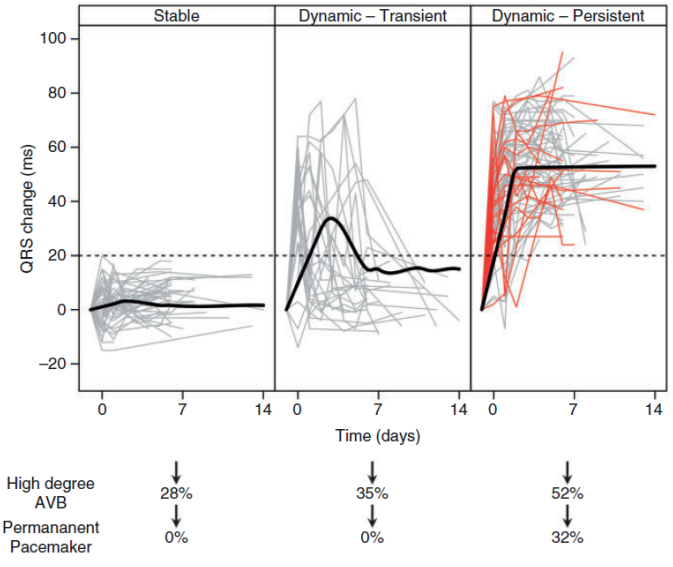


Figure 2: Dynamics of QRS-duration in Cohort A—patients with normal conduction at baseline—with associated high degree AVB and PPM rates. QRS-interval times are plotted as a difference from baseline up to 14 days after TAVI (depending on date of discharge). The grey lines represent individual patients who did not require PPM, red lines represent patients who required PPM, and the bold black line is a smoothed line that connects mean QRS durations to reflect the general trend of the groups. Graphs were created with R version 3.3.0 (R Core Team, Vienna, Austria; http://www.R-project.org). AVB, atrioventricular block.

with a stable normal QRS pattern 1 day post-TAVI have a low likelihood for high degree AVB or PPM with pacemaker dependency. Conversely, patients with a transient/persistent QRS prolongation may develop high degree AVB or pacemaker dependency up to Day 6.

Mortality at 30 days and 1 year were 5% and 15%, respectively, and were similar between Cohorts A and B. Supplementary material online, Table S2 shows an overview of death causes during 1 year follow-up. Notably, 1 year mortality in patients from Cohort A with stable conduction occurred in seven patients in total, and none were related to late conduction disorders. These

deaths were caused by cerebral stroke, hepatic failure, hospital-acquired pneumonia, sepsis, terminal heart failure, rectum carcinoma, and in one patient the cause was unknown.

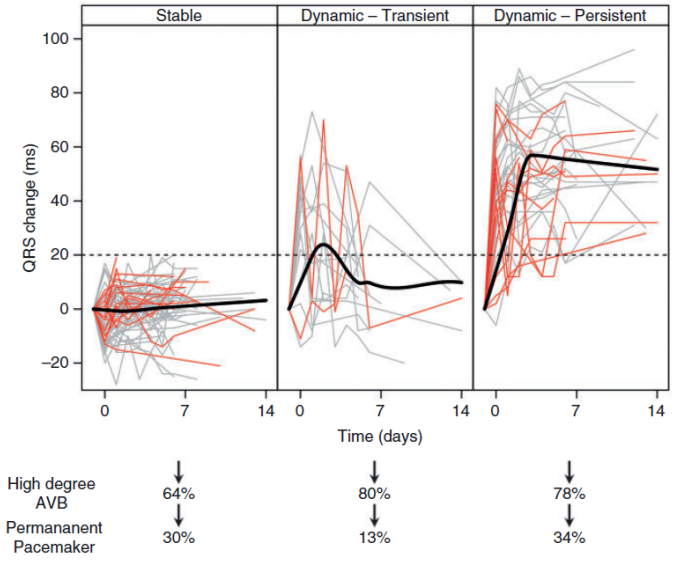


Figure 3: Dynamics of QRS-duration in Cohort B—patients with conduction disturbances at baseline—with associated high degree AVB and PPM rates. QRS-interval times are plotted as a difference from baseline up to 14 days after TAVI (depending on date of discharge). The grey lines represent individual patients who did not require PPM, red lines represent patients who required PPM, and the bold black line is a smoothed line that connects mean QRS durations to reflect the general trend of the groups. Graphs were created with R version 3.3.0 (R Core Team, Vienna, Austria; http://www.R-project.org). AVB, atrioventricular block.

Discussion

This study demonstrates that QRS dynamics following TAVI may have clinical implications. The main findings can be highlighted as follows: (i) three distinct patterns of QRS dynamics can be identified after TAVI: stable normal QRS-duration, transient QRS-prolongation, and persistent QRS-prolongation, (ii)

patients with newly acquired QRS prolongation after TAVI require longer telemetric monitoring than those with stable normal QRS duration and if persistent they have a high need for PPM, (iii) in patients with pre-existing conduction disturbances before TAVI, high degree AVB, and PPM occur irrespective of QRS dynamics, (iv) QRS prolongation typically peaks within 1 day after TAVI, (v) balloon-expandable TAVI is associated with more stable QRS duration, and (vi) up to 40% of patients are no longer pacemaker dependent during follow-up.

The vicinity of the atrioventricular His-bundle to the aortic valve contributes to the risk for conduction disturbances and PPM after aortic valve replacement in general and TAVI in particular. [5, 14] Past research has mainly focused on baseline predictors of conduction disturbances, [6, 8] enabling to identify those patients who are at high risk for PPM. In addition, choice of THV is a main contributor, as PPM rates are consistently higher with the self-expanding CoreValve, reported to be ∼30% [15–17] compared with 10–15% with the balloon-expandable Sapien XT and Sapien 315–17 and ∼30% with the mechanically expanded Lotus. [7]

Currently, daily practice is proceeding rapidly by discharging patients early at the expense of missing emerging conduction disorders. [4] In this study, QRS-prolongation typically peaked within 1 day after TAVI. Patients with normal baseline conduction and stable QRS duration or transient QRS prolongation never required PPM. When observed at daily intervals it appears that a stable QRS duration 1 day post-TAVI may justify safe early discharge. The opposite is true for patients with normal baseline conduction and persistent QRS prolongation, since they are at risk for high degree AVB, which impedes early discharge. Therefore, our data strongly recommends to keep these patients admitted on telemetric monitoring for a minimum of 6 days. We hypothesize that a persistent QRS-prolongation may identify more permanent and explicit

damage to the atrioventricular bundles.

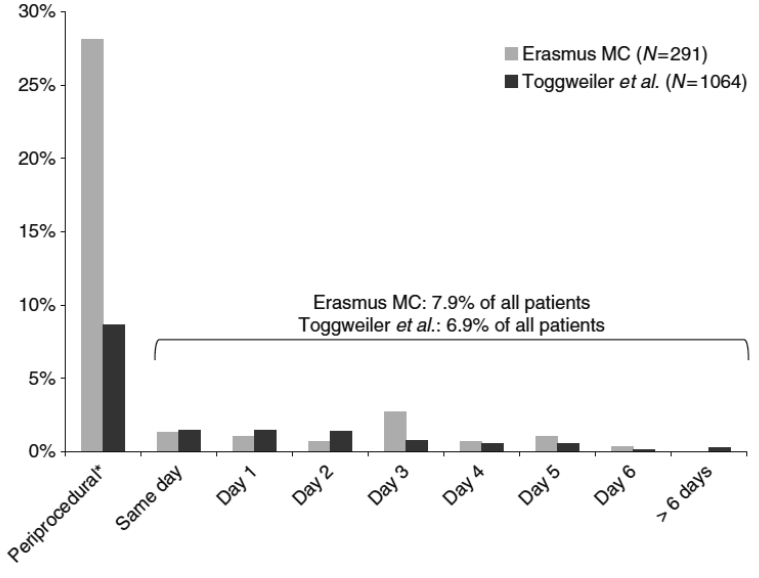


Figure 4: Overview of first presentation of high degree AVB in the present study and the study by Toggweiler et al. [10] Note that in this study, high degree AVB during the procedure was also noted, whereas the study by Toggweiler et al. only noted high degree AVB present at the moment the patient left the catheterization laboratory.

On the other hand, in patients with pre-existent conduction disturbances QRS prolongation can be deceiving, since patients with stable QRS duration and transient QRS prolongation also required PPM in this study. Therefore, patients with pre-existent conduction disorders have an eminent risk for PPM and thus warrant longer telemetric monitoring.

Recently, Toggweiler et al. [10] reported that ECGs after TAVI can be helpful to identify patients who need telemetric monitoring. The authors concluded that patients without conduction disorders or a stable ECG for 48 h after TAVI can be safely discharged. Our data may refine this concept by adding that patients with pre-existent conduction disturbances (i.e. prior to TAVI) may need longer clinical observation to rule out need for PPM.

In patients with normal conduction at baseline, QRS prolongation—either persistent or transient—was more frequent with CoreValve and Lotus. This is consistent with the higher reported PPM rates with these devices. It is intriguing that QRS-prolongation occurred less with balloon-expandable valves, suggesting patients treated with these devices are more suitable for early discharge (provided that other complications have been ruled out).

Whether TAVI patients remain pacemaker dependent is subject of ongoing debate. QRS prolongation after TAVI is sometimes transient as in more than one-third of the patients bundle branch blocks recover. [8] Pacemaker rhythm on follow-up ECGs can give a rough impression for at least partial pacemaker dependency. Data from the PARTNER-trial reported ventricular paced rhythm in 50% of patients 30 days after TAVI. [8] Moreover, at 1 year follow-up more than half of the patients appear no longer pacemaker dependent. [9] Outside the field of TAVI, the 'Inhibition of Unnecessary Right Ventricular Pacing With AVSH in ICDs'-study (INTRINSIC RV) aimed to inhibit the rate of unnecessary right ventricular pacing with implantable cardioverter-defibrillators. The authors reported that patients with less than 20% right ventricular pacing had 0% pacing when switched to VVI-mode. [13] Unnecessary pacing is mostly related to premature atrial and ventricular contractions or pacemaker dysfunction, [18] which implicates that 0% pacing is rare, even in patients with normal conduction. In this study, 41% of the patients with a permanent pacemaker were paced less than 20% during 6 months follow-up. Dedicated interrogation of these pacemakers revealed that none of these patients were truly pacemaker dependent at follow-up. This supports the theory that the threshold for PPM after TAVI may be (too) low. Future devices may enable continuous rhythm monitoring at home in the form of self-adhesive patches, with wireless transmission to the physician. [19] This may be a cost-effective way for safe early discharge and

avoiding needless PPM.

Our findings may have clinical implications. PPM in the elderly is not harmless, for early complications are common in patients above 75 years (5%). [20] Most frequent complications include lead dislodgement/loss of capture, pneumothorax, and infection. Moreover, unnecessary right ventricular pacing may contribute to heart failure.

The rate of delayed high degree AVB in this study was similar to what has been reported by Toggweiler et al. (Figure 4). In addition, newly acquired QRS prolongation that persisted for 6 days announced events that would oppose safe discharge in 12% of patients with normal QRS at baseline and underscores the importance of prolonged telemetric monitoring. Duration of telemetric monitoring and timing for safe discharge in patients with newly acquired or pre-existing conduction disorders requires further study and validation in a larger prospective cohort.

Limitations

This was a single-centre observational study and may suffer from inherent bias. Pre-existent conduction disorders were not equally distributed among different THVs and combinations of various conduction disorders were present. Our study represents a real world TAVI population and it therefore enhances generalizability. The decision to implant a PPM was at the treating physician's discretion but was almost exclusively high degree AVB and thus conform current international guidelines. [12] Electrocardiogram with analysable conduction times (i.e. not intruded by a ventricular pacing) was available for 66% of total hospitalized days and was thus incomplete, yet reflects retrospective analysis of current clinical practice. Our study represents the most elaborate sample of conduction times after TAVI reported to date. The missing ECGs were equally

distributed over time and among the different devices and conduction patterns, therefore, we believe the described patterns are valid.

Conclusions

Three distinct patterns of QRS dynamics can be discerned after TAVI and their predictive probabilities strongly relate to the baseline conduction status. Patients with normal conduction before and after TAVI do not develop need for PPM and may be pre-eminently eligible for early discharge. Patients with pre-existing or newly acquired QRS prolongation need prolonged telemetric monitoring because need for PPM is high.

References

1. Leon, M. B. *et al.* Transcatheter or Surgical Aortic-Valve Replacement in Intermediate-Risk Patients. *N Engl J Med* **374,** 1609–20 (2016).
2. Reardon, M. J., Van Mieghem, N. M. & Popma, J. J. Surgical or Transcatheter Aortic-Valve Replacement. *N Engl J Med* **377,** 197–198 (2017).
3. Baumgartner, H. *et al.* 2017 ESC/EACTS Guidelines for the management of valvular heart disease. *Eur Heart J* **38,** 2739–2791 (2017).
4. Durand, E. *et al.* Feasibility and safety of early discharge after transfemoral transcatheter aortic valve implantation with the Edwards SAPIEN-XT prosthesis. *Am J Cardiol* **115,** 1116–22 (2015).
5. Van der Boon, R. M. *et al.* New conduction abnormalities after TAVI–frequency and causes. *Nat Rev Cardiol* **9,** 454–63 (2012).
6. Siontis, G. C. *et al.* Predictors of permanent pacemaker implantation in patients with severe aortic stenosis undergoing TAVR: a meta-analysis. *J Am Coll Cardiol* **64,** 129–40 (2014).
7. Meredith Am, I. T. *et al.* Transcatheter aortic valve replacement for severe symptomatic aortic stenosis using a repositionable valve system: 30-day primary endpoint results from the REPRISE II study. *J Am Coll Cardiol* **64,** 1339–48 (2014).
8. Nazif, T. M. *et al.* Predictors and clinical outcomes of permanent pacemaker implantation after transcatheter aortic valve replacement: the PARTNER (Placement of AoRtic TraNscathetER Valves) trial and registry. *JACC Cardiovasc Interv* **8,** 60–9 (2015).

9. Van der Boon, R. M. *et al.* Pacemaker dependency after transcatheter aortic valve implantation with the self-expanding Medtronic CoreValve System. *Int J Cardiol* **168,** 1269–73 (2013).

10. Toggweiler, S. *et al.* The Electrocardiogram After Transcatheter Aortic Valve Replacement Determines the Risk for Post-Procedural High-Degree AV Block and the Need for Telemetry Monitoring. *JACC Cardiovasc Interv* **9,** 1269–76 (2016).

11. Tomlinson, D. R., Bashir, Y., Betts, T. R. & Rajappan, K. Accuracy of manual QRS duration assessment: its importance in patient selection for cardiac resynchronization and implantable cardioverter defibrillator therapy. *Europace* **11,** 638–42 (2009).

12. Brignole, M. *et al.* 2013 ESC Guidelines on cardiac pacing and cardiac resynchronization therapy: the Task Force on cardiac pacing and resynchronization therapy of the European Society of Cardiology (ESC). Developed in collaboration with the European Heart Rhythm Association (EHRA). *Eur Heart J* **34,** 2281–329 (2013).

13. Olshansky, B. *et al.* Is dual-chamber programming inferior to single-chamber programming in an implantable cardioverter-defibrillator? Results of the INTRINSIC RV (Inhibition of Unnecessary RV Pacing With AVSH in ICDs) study. *Circulation* **115,** 9–16 (2007).

14. Van Mieghem, N. M. *et al.* Persistent annual permanent pacemaker implantation rate after surgical aortic valve replacement in patients with severe aortic stenosis. *Ann Thorac Surg* **94,** 1143–9 (2012).

15. Abdel-Wahab, M. *et al.* Comparison of balloon-expandable vs self-expandable valves in patients undergoing transcatheter aortic valve replacement: the CHOICE randomized clinical trial. *JAMA* **311,** 1503–14 (2014).

16. Gilard, M. *et al.* Registry of transcatheter aortic-valve implantation in high-risk patients. *N Engl J Med* **366,** 1705–15 (2012).

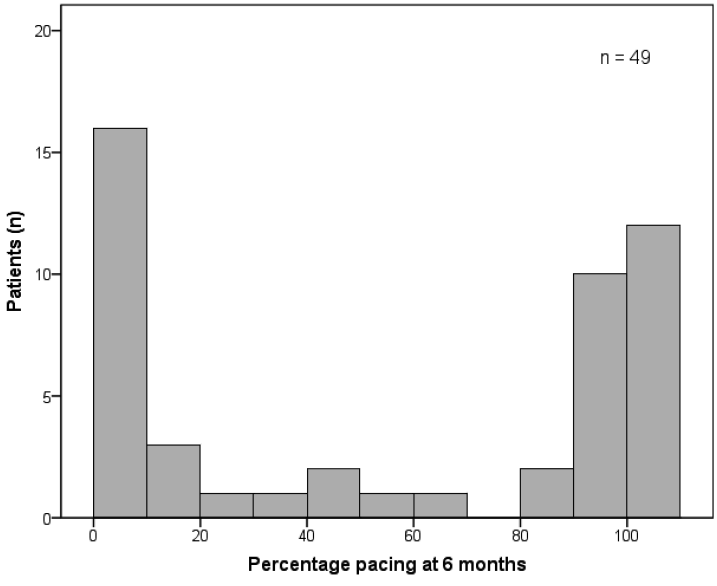
17. Kodali, S. *et al.* Early clinical and echocardiographic outcomes after SAPIEN 3 transcatheter aortic valve replacement in inoperable, high-risk and intermediate-risk patients with aortic stenosis. *Eur Heart J* **37,** 2252–62 (2016).

18. Strik, M. *et al.* Performance of a specific algorithm to minimize right ventricular pacing: A multicenter study. *Heart Rhythm* **13,** 1266–73 (2016).

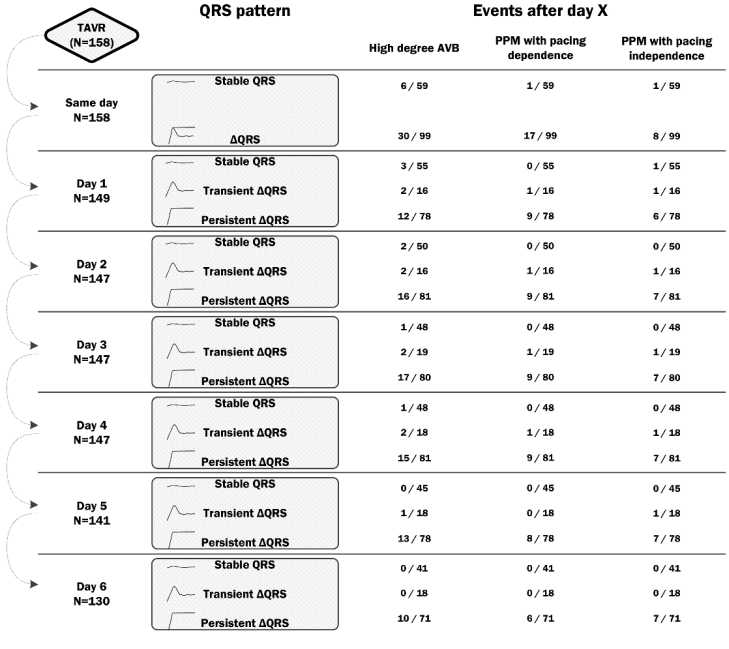
19. Shrivastav, M., Padte, S. & Sinha, N. Patient experience with a novel patch-like external loop recorder for cardiac arrhythmia detection in India. *Expert Rev Med Devices* **11,** 259–64. ISSN: 1745-2422 (Electronic) 1743-4440 (Linking) (2014).

20. Armaganijan, L. V. *et al.* Are elderly patients at increased risk of complications following pacemaker implantation? A meta-analysis of randomized trials. *Pacing Clin Electrophysiol* **35,** 131–4 (2012).

Supplemental Material



Supplemental Figure 1



Supplemental Figure 2

Supplemental Table 1. QRS dynamics from baseline to 30 days follow-up in Cohort A.

	Stable (n=47)	Dynamic transient (n=26)	Dynamic persistent (n=85)	P-value
Baseline QRS duration	95 ± 11	100 ± 9	97 ± 11	0.129
Δ QRS baseline – maximum	+8 ± 6*	+48 ± 16	+57 ± 15	<0.001
Δ QRS maximum – discharge	-3 ± 6	-45 ± 17*	-6 ± 12	<0.001
Δ QRS discharge – follow-up 30 days	-5 ± 6	-1 ± 10	-15 ± 20*	<0.001
Δ QRS baseline – follow-up 30 days	0 ± 5	+3 ± 9	+39 ± 23*	<0.001
Follow-up 30 days QRS duration	98 ± 10*	102 ± 11*	137 ± 23*	<0.001

*denotes significant differences compared to both other subgroups at a p<0.05 confidence level.

Supplemental Table 2. Overview of death causes during follow-up for Cohort A and B.

	Cohort A		Cohort B	
	PPM +	PPM -	PPM +	PPM -
Mortality at 30 days	1/27 (4%)	7/130 (5%)	1/39 (3%)	5/94 (5%)
Mortality at 1 year*	1/26 (4%)	22/125 (18%)	7/36 (18%)	12/85 (13%)
Death cause	- Stroke	- Sepsis (2x) - Multi-organ failure - Pneumonia (2x) - Stroke (3x) - Pulmonary embolism - Liver cirrhosis - Heart failure - Cancer (2x) - Myocardial infarction - Unknown (8x)	- Sepsis - Subdural hematoma - Pneumonia - Unknown (4x)	- Sepsis - Stroke (2x) - Multi-organ failure - Hemothorax - Valve embolization - Colitis - Liver cirrhosis - Unknown (4x)

Incidence is shown as counts (n/N) and percentages.
*19 patients were lost to follow-up.
PPM; permanent pacemaker implantation.

Discussion

In this thesis we developed and applied statistical methods to improve prediction of clinical outcomes in CVD. We will discuss the main findings below.

In **Part I** we investigated the consequences of applying a case-cohort design when using the joint modelling framework. A joint model was estimated in a study with patients admitted for acute coronary syndrome that underwent frequent, repeated biomarker measurements. For financial reasons, only a selection of the biomarker measurements were ascertained by a case-cohort design. Standard statistical modelling on the case-cohort design would have resulted in bias due to oversampling of the cases. This situation could also be regarded as a missing data problem, where the patients left out of the case-cohort design contain missing data. Based on the idea by Dong et al. [1] we included the survival information of the complete cohort in the analysis, which is available due to the nature of the case-cohort design: namely, if these patients had experienced the study end point, they would have been included in the case-cohort subset. Therefore we can conclude that the excluded patients were censored. If the survival information of these patients is included in the analysis, only their biomarker measurements remain missing. However, these are now missing randomly conditional on observed data (survival status) and are therefore missing at random (MAR). Joint models fitted with MAR yield unbiased estimates and therefore the results will be valid. Simulation studies showed that indeed case-cohort studies with complete survival information perform similarly to studies that assess the biomarkers in the full cohort, both in terms of unbiased parameter estimates and predictive accuracy of the models.

Part II deals with dynamic modelling in a different manner. Based on methods which are popular in oncology research, we investigated the impact of different measures of survival on estimates of prognosis of patients. In Chapter 2 we calculated relative conditional survival in a large cohort of patients undergoing percutaneous coronary intervention (PCI) with long term follow-up. Chapter 3 applies the same techniques to a cohort of patients diagnosed with heart failure (HF). Patients undergoing PCI showed one-year survival rates between 83%-98%, depending on the age of the patient, estimated using the Kaplan-Meier method. At ten years their survival was between 36%-91%. The patients from the HF cohort showed lower survival rates (between

56%-64% after one year), that persisted throughout follow-up (9%-35% survival at 10-years). When time already survived was taken into account in the analysis in a dynamic manner and prognosis was compared to the general population, both cohorts showed different prognosis patterns. For the patients undergoing PCI, it was clear that the first month after the intervention was the most crucial, and if patients survived this period, survival for the remainder of the follow-up was much higher as demonstrated by conditional survival. Additionally, survival approached that of persons from the general population with the same age and gender as shown by relative conditional survival. For heart failure patients, on the other hand, although improvements in survival were present after the first year since diagnosis, estimated survival probabilities remained below the survival of the general population throughout follow-up. This means that for these patients prognosis never becomes 'normal' once they are diagnosed with HF. In general, the information from these additional survival methods could be useful in interpreting prognosis and discussing this with patients experiencing different types of CVD.

In **Part III** we focused on predicting outcomes in women. Recently, differences between men and women in terms of CVD have received increasing attention. In Chapter 4 we aimed to investigate how this has translated into developing distinct prediction models for men and women, and whether there are differences in risk factors used in the models. We performed a systematic review of all prediction models developed for women in the general population. We distinguished between models that were developed solely on women (female-specific models) and models that included gender as a covariate (sex-predictor models). Through our systematic search we identified 285 distinct prediction models, although a large majority of these have not been externally validated. We found that the female-specific models did not perform substantially better than the sex-predictor models. However we also found that only two female-specific models (1.3%) included predictors that are actually female-specific. Based on the currently available studies, we were unable to conclude whether the lack of difference in performance between the two types of models was due to not using informative female-specific predictors or due to the fact that static prediction

models may have reached their maximum performance capability and more advanced dynamic models are needed.

In Chapter 5 we modelled outcomes in pregnant women with structural heart disease. Methodological difficulties occurred due to the three-level structure of the data. Women were included in different hospitals in different countries. By adding random effects for hospitals and countries we were able to model the hierarchical structure in the data. Additionally, we investigated the degree of variability in the random effects with two versions of the R^2 adapted to be suitable for generalized linear mixed models [2], the $R^2_{GLMM(c)}$ (which denotes the proportion of variability explained by both the random and fixed effects) and the $R^2_{GLMM(m)}$ (which denotes the proportion of variability explained only by the random effects). When these measures are compared they indicate the amount of variability that is found in the random effects. We found that, when known important covariates on patient and country level were entered into the model as fixed effects, the difference between these measures was small. This means that in this case the random effects only account for a small portion of the total variability.

In **Part IV** we applied hierarchical modelling techniques to clinical data sets with longitudinal data. In Chapters 6 and 7 we used the joint modelling framework to investigate the relationship between repeated values of Galectin-3 (Chapter 6) and ST2 (Chapter 7) biomarkers with the risk of an event in patients with acute heart failure. Both biomarkers showed stronger associations with the event when the whole trajectory of the biomarkers was used compared to only the baseline measurement. In Chapter 8 we assessed repeated measurements of quality of life questionnaires to predict outcomes in children with dilated cardiomyopathy. Due to the nature of the data we were unable to obtain a parametric fit of the longitudinal trajectories of the questionnaires and used a time-dependent Cox model to predict the outcomes. Chapter 9 deals with daily electrocardiogram measurements after transcatheter aortic valve implantation (TAVI). The longitudinal profiles were used to identify groups with three different patterns. The predictive ability of the profiles depended strongly on the baseline conduction status. In the group of patients with normal conduction

before TAVI, pacemaker implantations were only needed when there was persistent QRS prolongation.

Future Research

The results presented in this thesis provide room for improvements and extensions. Currently we are investigating optimization for two-phase sample designs in the joint modeling framework. Based on the results found in Chapter 1, we want to investigate whether the case-cohort design is the most efficient way to select a subset of longitudinal measurements for analysis or if we can find a more optimal subset. Instead of the case-cohort design, other two-phase sample designs could be used in studies to assess only a subset of the longitudinal measurements, such as the nested case-control study.

First we need to determine what defines an efficient design. A design is considered more efficient than another, if it uses the same number of longitudinal measurements, yet provides more information about the research question. In the framework of experimental design, a most informative or optimal design is usually found by defining a certain utility function (such as a function over the precision of the parameter estimates or prediction of events) and identifying the design that maximizes this function. [3] A popular choice for the utility function in Bayesian experimental designs is the Kullback-Leibler divergence (KLD), which measures how different one distribution is from another, and can be used to calculate the amount of information lost or gained by using one distribution over the other to describe the data. [4] The methodology for computing the KLD in the joint modeling framework needs to be developed further so that it can be used to compare the efficiency of different sample designs.

Additionally, we are not only interested in comparing efficiency of certain sampling designs, but we want to find the optimal design. For this we will need to develop tools to evaluate the optimality of numerous possible designs. The optimal design will depend on several features, which will need to be incorporated into the methodology.

First of all, the shape of the trajectories will have an impact. Non-linear evolutions will require more samples to recover the shape than linear evolutions. Secondly, the aim of the study also plays a role. Different subsets might be optimal when the goal is to model the longitudinal trajectory and predict new longitudinal measurements than when the goal is to predict the probability of experiencing the study end point.

Another challenge we faced in using the joint modelling framework on the real-life case-cohort data occurred, when we assessed the predictive performance of the estimated model in Chapter 1. For the predictive accuracy, we evaluated how often the model, based on the biomarker values of a patient, correctly identifies a patient as a case if they had the event and as a non-case if the patient was event-free. Although the survival information of the full cohort was used to estimate the models, the patients without biomarker measurements could not be used when calculating the predictive performance. Additionally, the event rate should be comparable to the event rate in the population. Therefore, the predictive performance measures were calculated on a data set consisting of only the randomly drawn subcohort. In the case of our real-life cohort, this meant that the predictive accuracy of the model was evaluated on only eight patients with the study end point (the event patients in the randomly drawn subcohort). To calculate predictive accuracy measures in case-cohort studies in a more reliable manner, new methods are needed.

In the chapters on relative conditional survival, we have seen that a new light can be shed on prognosis by presenting survival estimates in different ways. By calculating relative survival, the prognosis of a patient can be placed in context of the general population. A downside to relative survival, which is also the case for most measures of survival, is that it is presented as a percentage. Percentages are often difficult for people to interpret. What exactly does a 95% relative survival measure mean? To help patients understand their prognosis even better we could consider representing relative survival in terms of 'age' instead of percentages. That way, a treating physician can, for example, demonstrate to a 60-year old male patient, that his estimated survival probability corresponds to the survival probability of a 70-year old man from the general population. This may help patients to grasp the severity of their situation.

We are currently developing this concept further.

General conclusion

In this thesis we explored and applied different modelling techniques to predict clinical outcomes in cardiovascular diseases. We have focused on dynamic models, which incorporate the repeatedly updated characteristics of the patient to obtain better predictions. This repeated update occurs either through repeated biomarker measurements or merely by accounting for the fact that the patient is still alive to return to the hospital. These models require complex data structures with multiple dependencies. However, if the correct methodological techniques are used, the models form powerful tools to improve predictions of patient outcomes and additionally help patients better understand their prognosis.

References

1. Dong, X., Kong, L. & Wahed, A. S. Accelerated failure time model for case-cohort design with longitudinal covariates subject to measurement error and detection limits. *Stat Med* **35,** 1327–39 (2016).
2. Nakagawa, S. & Schielzeth, H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* **4,** 133–142 (2013).
3. Chaloner, K. & Verdinelli, I. Bayesian experimental design: A review. *Statistical Science* **10,** 273–304 (1995).
4. Kullback, S. & Leibler, R. A. On information and sufficiency. *The annals of Mathematical Statistics* **22,** 79–86 (1951).

Appendix

Summary

In this thesis we developed and applied statistical methods to improve prediction of clinical outcomes in cardiovascular diseases (CVD). Prediction models are an important tool in (cardiovascular) research, because they can identify individuals at high risk of developing CVD and indicate which factors are important contributors to the development of CVD. Additionally, prediction models can be used for patients who have already been diagnosed with CVD to estimate their risk of a recurrent event. Most prediction models are static, in the sense that they use one measurement of the health status of a patient to estimate their risk for a long period of time. Improvements can be made when patients are measured repeatedly over time and their updated status is incorporated in risk prediction. An often used framework for this is the joint modeling framework for longitudinal and time-to-event data.

In **Part I**, Chapter 1 we investigated the consequences of applying a case-cohort design when using the joint modelling framework. A joint model was estimated in a study with patients admitted for acute coronary syndrome that underwent frequent, repeated biomarker measurements. For financial reasons, only a selection of the biomarker measurements were ascertained by a case-cohort design. Standard statistical models on the case-cohort design would have resulted in bias due to oversampling of the cases. We included the survival information of the complete cohort in the analysis. Using simulations we were able to show that case-cohort studies where the complete survival information is used for analysis, indeed perform similarly to studies that assess the full cohort of biomarkers, both in terms of unbiased parameter estimates and predictive accuracy of the models.

Part II deals with dynamic modelling in a different manner. Based on methods which are popular in oncology research, we investigated the impact of different measures of survival on the estimates of prognosis of patients. In Chapter 2 we calculated relative conditional survival in a large cohort of patients undergoing Percutaneous Coronary Intervention (PCI) with a long term follow-up. Chapter 3 applies the same techniques to a cohort of patients diagnosed with heart failure (HF). In conditional

survival, the prognosis of a patient is updated repeatedly by taking the time already survived into account. Relative survival relates the survival probabilities of a patient to that of someone from the general population with the same age and gender, and the combination of the two (relative conditional survival) tells us at what point in time the prognosis of the patient becomes similar to that of someone without the disease. In the PCI patients this method of calculating survival demonstrated that the first month after the procedure was crucial and after a year the prognosis of the patients was on the same level as the general population. For the HF patients, on the other hand, it showed that prognosis always stays below population level, indicating that these patients carry the burden of their disease for the rest of their lives.

Recently, differences between men and women in terms of CVD have received increasing attention. In **Part III**, Chapter 4 we aimed to investigate how this has translated into developing distinct prediction models for men and women, and whether there are differences in risk factors used in the models. We identified 285 distinct prediction models for women in the general population. We found that models that were developed solely on women (*female-specific* models) did not perform substantially better than the models that only included gender as a covariate. However, only two female-specific models (1.3%) included predictors that are actually female-specific (such as menopause, age at menarche and pregnancy status). In Chapter 5 we modelled outcomes in pregnant women with structural heart disease. Women were included in this study from different hospitals in different countries around the world. By incorporating the hierarchical structure of the data in the model, we were able to deal with clustering found in this data set.

In **Part IV** we applied hierarchical modelling techniques to clinical data sets with longitudinal data. In Chapters 6 and 7 we used the joint modelling framework to investigate repeated values of *Galectin-3* and *ST2* biomarkers and the risk of an event in patients with acute heart failure. In Chapter 8 we assessed repeated measurements of quality of life questionnaires to predict outcomes in children with dilated cardiomyopathy. The psychical functioning of children that is measured multiple times was an independent predictor of death of risk of transplantation.

Chapter 9 deals with daily electrocardiogram measurements after transcatheter aortic valve implantation (TAVI). The longitudinal profiles were used to identify three groups with different patterns.

General conclusion

In this thesis we explored and applied different modelling techniques to predict clinical outcomes in cardiovascular diseases. We have focused on dynamic models, which incorporate the repeatedly updated characteristics of the patient to obtain better predictions. This repeated update occurs either through repeated biomarker measurements or merely by accounting for the fact that the patient is still alive to return to the hospital. These models require complex data structures with multiple dependencies. However, if the correct methodological techniques are used, the models form powerful tools to improve predictions of patient outcomes and additionally help patients understand their prognosis better.

Nederlandse samenvatting

In dit proefschrift hebben we statistische methoden ontwikkeld en toegepast om klinische uitkomsten beter te kunnen voorspellen in hart- en vaatziekten (HVZ). Voorspelmodellen spelen een belangrijke rol in het (cardiovasculaire) onderzoek, omdat zij patiënten kunnen identificeren die een hoog risico hebben om HVZ te ontwikkelen en aantonen welke risico factoren belangrijk zijn voor het ontwikkelen van HVZ. Tevens kunnen voorspelmodellen worden gebruikt bij patiënten die al gediagnosticeerd zijn met HVZ, om voor hen het risico op een tweede incident te voorspellen. Veel voorspelmodellen zijn statische modellen, wat betekent dat zij een meting van de gezondheidsstatus op één tijdsmoment gebruiken om het risico van een patiënt over een lange tijd in te schatten. De modellen kunnen worden verbeterd door patiënten vaker te meten en bijgewerkte informatie mee te nemen in het voorspellen van het risico. Een methode die hier vaak voor gebruikt wordt zijn de gecombineerde modellen voor longitudinale en overlevingsdata (*joint models*).

In **Deel I**, Hoofdstuk 1 hebben we onderzocht wat de gevolgen zijn als *joint models* worden toegepast in een *case-cohort* studie. Een joint model is gemaakt op data uit een studie met patiënten die in het ziekenhuis waren opgenomen met acuut coronair syndroom (ACS), waarvan veelvuldig herhaalde biomarkers werden gemeten. Wegens financiële redenen werd bij slechts een deel van de metingen de biomarker waarde daadwerkelijk bepaald, waarbij het case-cohort studie ontwerp is gebruikt. Gebruikelijke statistische modellen op deze case-cohort studie zouden geresulteerd hebben in onzuivere resultaten, omdat de patiënten met het eindpunt ('cases') oververtegenwoordigd zijn in de analyse data. Voor de analyses hebben wij de overlevingsdata van het volledige cohort meegenomen. Door middel van simulaties hebben we aangetoond dat een joint model geschat op een case-cohort studie, maar met complete overlevingsdata, vergelijkbare resultaten oplevert als wanneer de volledige data beschikbaar zou zijn. We hebben hier zowel gekeken naar de zuiverheid van de parameter schattingen, als de nauwkeurigheid waarmee overleving voorspeld kan worden.

Deel II behandelt dynamische modellen op een andere manier. Gebaseerd op methodes die populair zijn in het oncologische onderzoek, hebben wij onderzocht wat het effect op de prognose van patiënten is, wanneer overlevingskansen op verschillende manieren worden uitgedrukt. In Hoofdstuk 2 hebben wij *relatieve conditionele overleving* berekend in patiënten met Percuntane Coronaire Interventie (PCI), die een lange vervolg termijn hadden. Hoofdstuk 3 past dezelfde technieken toe op een patiëntencohort dat gediagnosticeerd is met hartfalen. Met conditionele overleving wordt de prognose van de patiënt steeds bijgewerkt door de tijd die de patiënt al overleefd heeft mee te nemen in de overlevingsberekening. Relatieve overleving vergelijkt de overlevingskansen van de patiënt met iemand uit de algemene populatie van dezelfde leeftijd en hetzelfde geslacht, en de combinatie van beiden vertelt ons op welk moment in de tijd de prognose van een patiënt overeenkomt met de prognose van een soortgelijk iemand zonder de ziekte. In de PCI patiënten laat deze methode zien dat de eerste maand na de ingreep cruciaal is en dat patiënten na een jaar gelijke prognoses hebben als mensen uit de algemene populatie. Voor de hartfalen patiënten geldt echter dat hun overleving altijd lager blijft dan het populatie niveau. Dit betekent dat deze patiënten de last van de diagnose hun hele leven bij zich dragen.

Recentelijk is er meer aandacht gekomen voor de verschillen tussen mannen en vrouwen op het gebied van HVZ. In **Deel III**, Hoofdstuk 4 hebben wij onderzocht in hoeverre dit zich heeft vertaald in aparte voorspelmodellen voor mannen en vrouwen en of er in deze modellen andere risico factoren zijn gebruikt. We hebben 285 verschillende voorspelmodellen geïdentificeerd voor vrouwen uit de algemene populatie. De modellen die puur op vrouwen zijn ontwikkeld (de *vrouwspecifieke* modellen) presteerden niet substantieel beter dan de modellen die geslacht als risicofactor meenamen. Slechts twee vrouwspecifieke modellen (1.3%) hadden ook daadwerkelijk vrouwspecifieke risicofactoren geïncludeerd (zoals menopauze, leeftijd van menarche en zwangerschap). In Hoofdstuk 5 hebben we uitkomsten gemodelleerd in zwangere vrouwen met structurele hartziektes. In deze studie waren vrouwen geïncludeerd vanuit verschillende ziekenhuizen in verschillende landen. Door de hiërarchische structuur mee te nemen in de modellen, konden we rekening houden met de clustering in de

data.

In **Deel IV** hebben we hiërarchische modeltechnieken toegepast op klinische studies met longitudinale data. In Hoofdstuk 6 en 7 hebben we joint models gebruikt om de relatie aan te tonen tussen herhaald gemeten *Galectine-3* en *ST2* biomarkers en het risico op een eindpunt in patiënten met acuut hartfalen. In Hoofdstuk 8 hebben onderzocht of we met herhaald afgenomen vragenlijsten over de kwaliteit van leven uitkomsten kunnen voorspellen in kinderen met gedilateerde cardiomyopathie (DCM). De fysieke functionering van de kinderen die herhaaldelijk werd bepaald was een onafhankelijke voorspeller voor dood of transplantatie. Hoofdstuk 9 beschrijft dagelijkse metingen van het elektro-cardiogram van patiënten na een *transcatheter aortic valve implantation (TAVI)*. De patronen van de metingen zijn gebruikt om drie groepen patiënten te identificeren.

Algemene conclusie

In dit proefschrift hebben we verschillende modelleertechnieken onderzocht en toegepast om klinische uitkomsten te voorspellen in HVZ. We hebben ons gefocust op dynamische modellen, die door het herhaald bijwerken van de gezondheidsstatus van de patiënt, beter in staat zijn om uitkomsten te kunnen voorspellen. Deze update kan komen door herhaaldelijke (biomarker)metingen of door simpelweg de informatie dat de patiënt nog steeds in leven is, mee te nemen in het model. Deze modellen vereisen complexe data structuren met meerdere afhankelijkheden. Als, echter, de juiste methodologische technieken worden gebruikt, kunnen deze modellen krachtige hulpmiddelen vormen om klinische uitkomsten beter te voorspellen of patiënten te helpen met het begrijpen en interpreteren van hun prognose.

List of Publications

1. **Baart, S. J.**, Berge, J. C., Akkerhuis, K. M., Boersma, H., Kardys, I., "Relative conditional survival analysis provides additional insights into the prognosis of heart failure patients". *Submitted*.

2. **Baart, S. J.**, Boersma, E., Rizopoulos, D., "Joint models for longitudinal and time-to-event data in a case-cohort design". In: *Stat Med* (2019).

3. **Baart, S. J.**, Dam, V., Scheres, L. J. J., Damen, J., Spijker, R., Schuit, E., Debray, T. P. A., Fauser, B., Boersma, E., Moons, K. G. M., Schouw, Y. T., on behalf of the CREW consortium, "Cardiovascular risk prediction models for women in the general population: A systematic review". In: *PLoS One* 14.1 (2019).

4. Berg, V. J., Strachinaru, M., Akkerhuis, K. M., **Baart, S.**, Brankovic, M., Constantinescu, A. A., Cornel, J. H., Manintveld, O. C., Umans, V., Rizopoulos, D., Geleijnse, M. L., Boersma, E., Dalen, B. M., Kardys, I., "Repeated Echocardiograms Do Not Provide Incremental Prognostic Value to Single Echocardiographic Assessment in Minimally Symptomatic Patients with Chronic Heart Failure: Results of the Bio-SHiFT Study". In: *J Am Soc Echocardiogr* (2019). [Epub ahead of print].

5. Bouwens, E., Brankovic, M., Mouthaan, H., **Baart, S.**, Rizopoulos, D., Boven, N., Caliskan, K., Manintveld, O., Germans, T., Ramshorst, J., Umans, V., Akkerhuis, K. M., Kardys, I., "Temporal Patterns of 14 Blood Biomarker candidates of Cardiac Remodeling in Relation to Prognosis of Patients With Chronic Heart Failure-The Bio- SH i FT Study". In: *J Am Heart Assoc* 8.4 (2019).

6. Grootel, R. W. J., Bosch, A. E., Baggen, V. J. M., Menting, M. E., **Baart, S. J.**, Cuypers, J., Witsenburg, M., Roos-Hesselink, J. W., "The Prognostic Value of Myocardial Deformation in Adult Patients With Corrected Tetralogy of Fallot". In: *J Am Soc Echocardiogr* 32.7 (2019), 866–875 e2.

7. Baggen, V. J. M., **Baart, S. J.**, Bosch, A. E., Eindhoven, J. A., Witsenburg, M., Cuypers, J., Roos-Hesselink, J. W., Boersma, E., "Prognostic Value of Serial N-Terminal Pro-B-Type Natriuretic Peptide Measurements in Adults With Congenital Heart Disease". In: *J Am Heart Assoc* 7.7 (2018).

8. Brankovic, M., Akkerhuis, K. M., Boven, N., Anroedh, S., Constantinescu, A., Caliskan, K., Manintveld, O., Cornel, J. H., **Baart, S. J.**, Rizopoulos, D., Hillege, H., Boersma, E., Umans, V., Kardys, I., "Patient-specific evolution of renal function in chronic heart failure patients dynamically predicts clinical outcome in the Bio-SHiFT study". In: *Kidney Int* 93.4 (2018), pp. 952–960.

9. Brankovic, M., Kardys, I., Hoorn, E. J., **Baart, S. J.**, Boersma, E., Rizopoulos, D., "Personalized dynamic risk assessment in nephrology is a next step in prognostic research". In: *Kidney Int* 94.1 (2018), pp. 214–217.

10. Gils, L., **Baart, S. J.**, Kroon, H., Rahhab, Z., El Faquir, N., Rodriguez Olivares, R., Aga, Y., Maugenest, A. M., Theuns, D. A., Boersma, E., Szili Torok, T., De Jaegere, P. P., Van Mieghem, N. M., "Conduction dynamics after transcatheter aortic valve implantation and implications for permanent pacemaker implantation and early discharge: the CONDUCT-study". In: *Europace* 20.12 (2018), pp. 1981–1988.

11. Hagen, I. M., **Baart, S. J.**, Fong Soe Khioe, R., Sliwa-Hahnle, K., Taha, N., Lelonek, M., Tavazzi, L., Maggioni, A. P., Johnson, M. R., Maniadakis, N., Fordham, R., Hall, R., Roos-Hesselink, J. W., on behalf of the ROPAC investigators, "Influence of socioeconomic factors on pregnancy outcome in women with structural heart disease". In: *Heart* 104.9 (2018), pp. 745–752.

12. Hoven, A. T., Bons, L. R., **Baart, S. J.**, Moelker, A., Laar, I., Bosch, A. E., Bekkers, J. A., Verhagen, H. J. M., Linde, D., Roos-Hesselink, J. W., "Aortic Dimensions and Clinical Outcome in Patients With SMAD3 Mutations". In: *Circ Genom Precis Med* 11.11 (2018), e002329.

13. Velzen, H. G., Schinkel, A. F. L., **Baart, S. J.**, Huurman, R., Slegtenhorst, M. A., Kardys, I., Michels, M., "Effect of Gender and Genetic Mutations on Outcomes in Patients With Hypertrophic Cardiomyopathy". In: *Am J Cardiol* 122.11 (2018), pp. 1947–1954.

14. Velzen, H. G., Schinkel, A. F. L., **Baart, S. J.**, Oldenburg, R. A., Frohn-Mulder, I. M. E., Slegtenhorst, M. A., Michels, M., "Outcomes of Contemporary Family Screening in Hypertrophic Cardiomyopathy". In: *Circ Genom Precis Med* 11.4 (2018).

15. **Baart, S. J.**, Domburg, R. T., Janssen-Heijnen, M. L. G., Deckers, J. W., Akkerhuis, K. M., Daemen, J., Geuns, R. J., Boersma, E., Kardys, I., "Impact of Relative Conditional Survival Estimates on Patient Prognosis After Percutaneous Coronary Intervention". In: *Circ Cardiovasc Qual Outcomes* 10.6 (2017).

16. Boer, S. L., **Baart, S. J.**, Meulen, M. H., Iperen, G. G., Backx, A. P., Ten Harkel, A. D., Rammeloo, L. A., Marchie Sarvaas, G. J., Tanke, R. B., Helbing, W. A., Utens, E. M., Dalinghaus, M., "Parent reports of health-related quality of life and heart failure severity score independently predict outcome in children with dilated cardiomyopathy". In: *Cardiol Young* 27.6 (2017), pp. 1194–1202.

17. Cuypers, J. A., Menting, M. E., Opic, P., Utens, E. M., Helbing, W. A., Witsenburg, M., Bosch, A. E., Domburg, R. T., **Baart, S. J.**, Boersma, E., Meijboom, F. J., Bogers, A. J., Roos-Hesselink, J. W., "The unnatural history of pulmonary stenosis up to 40 years after surgical repair". In: *Heart* 103.4 (2017), pp. 273–279.

18. Vark, L. C., Lesman-Leegte, I., **Baart, S. J.**, Postmus, D., Pinto, Y. M., Boer, R. A., Asselbergs, F. W., Wajon, E., Orsel, J. G., Boersma, E., Hillege, H. L., Akkerhuis, K. M., for the TRIUMPH investigators, "Prognostic Value of Serial Galectin-3 Measurements in Patients With Acute Heart Failure". In: *J Am Heart Assoc* 6.12 (2017).

19. Vark, L. C., Lesman-Leegte, I., **Baart, S. J.**, Postmus, D., Pinto, Y. M., Orsel, J. G., Westenbrink, B. D., Brunner-la Rocca, H. P., Miltenburg, A. J. M., Boersma, E., Hillege, H. L., Akkerhuis, K. M., for the TRIUMPH investigators, "Prognostic Value of Serial ST2 Measurements in Patients With Acute Heart Failure". In: *J Am Coll Cardiol* 70.19 (2017), pp. 2378–2388.

20. Domburg, R. T., Kardys, I., Lenzen, M., **Baart, S. J.**, Boersma, E., Hoeks, S. E., "Tools and techniques - statistical: it's statistically significant, but is it clinically relevant?" In: *EuroIntervention* 10.11 (2015), pp. 1368–9.

21. Kardys, I., **Baart, S. J.**, Domburg, R., Lenzen, M., Hoeks, S., Boersma, E., "Tools and Techniques - Statistical: A brief non-statistician's guide for choosing the appropriate regression analysis, with special attention to correlated data and repeated measurements". In: *EuroIntervention* 11.8 (2015), pp. 957–62.

Phd Portfolio

Name PhD Student:	Sara Johanna Baart
Research School:	COEUR
Erasmus MC Department:	Cardiology
PhD Period:	2014 - 2019
Promotors:	Prof. dr. ir. H. Boersma
	Prof. dr. D. Rizopoulos
Co-promotor:	Dr. I. Kardys

	Year	Workload
1. Phd Training		
Courses		
Cardiovascular Epidemiology (COEUR)	2014	1.5
Applied Multiple Imputation in R (ISCB)	2015	0.3
Advanced R programming (KULeuven)	2016	0.5
Workshop Joint Modeling and beyond (UHasselt)	2016	0.5
Multivariate dimension reduction for biological data integration (ISCB)	2018	0.3
Research Integrity Course	2019	0.3
Conferences		
36th International Society for Clinical Biostatistics Annual Conference, Utrecht, NL	2015	1.2
9th Eastern Mediterranean Region of the International Biometric Society, Thessaloniki, Greece (*oral presentation*)	2017	1.7
38th International Society for Clinical Biostatistics Annual Conference, Vigo, Spain (*oral presentation*)	2017	1.7
29th International Biometric Conference, Barcelona, Spain (*poster presentation*)	2018	1.5
Seminars and workshops		
COEUR Symposium Personalised Medicine	2015	0.2
COEUR PhD day	2016	0.3
Presentation COEUR PhD day	2017	0.8
Presentation CQM-seminar, Erasmus MC	2016-2018	1.5

2. Teaching activities		
Teacher		
Basic Introduction Course on SPSS - MolMed	2015 - 2018	12
Introductie statistische methoden - Minor onderwijs	2016	0.5
Assistent teacher		
Biostatistical Methods I (CC02) - NIHES	2017	0.5
Supervision		
Supervising 2nd year medical students in a systematic review	2014, 2015, 2017	1.8
3. Other		
Organization PhD Day BMS-ANed	2017	3
Member EuroIntervention Statistical Board	2014-2018	7.5
	Total	**37.6**

About the author

Sara Johanna Baart was born on August 24 1988 in Amsterdam, the Netherlands. After graduating high school in 2006 (Vossius Gymnasium, Amsterdam), she started the bachelor program of Econometrics and Operational Research at the University of Amsterdam in 2007. In 2011 she commenced the Masters program Statistics at the Katholieke Universiteit Leuven in Belgium. She chose the specialization track Biometrics and graduated with honours in February 2014. That year she started her PhD project within the CREW consortium at the Erasmus MC, in Rotterdam under the supervision of prof. dr. Eric Boersma. Currently she works as a post doc at the department of Biostatistics and as a statistician for the department of Pulmonology and the Pain center (department of Anesthesiology) in the Erasmus MC.

Dankwoord

Nu mijn promotietraject aan zijn einde is gekomen, wil ik graag degenen bedanken die ervoor gezorgd hebben dat dit boekje ook daadwerkelijk tot stand is gekomen.

Allereerst mijn promotor, professor Boersma. Beste Eric, ik weet nog heel goed hoe ik, inmiddels ruim vijf jaar geleden, vertwijfeld de trein naar Rotterdam pakte voor een sollicitatiegesprek bij jou voor het CREW consortium. Als geboren en getogen Amsterdammer leek het me niks om in Rotterdam te gaan werken. Ongeveer één minuut nadat ons gesprek was begonnen, waren mijn twijfels als sneeuw voor de zon verdwenen en had ik het gevoel dat ik me bij jouw groep goed op mijn plek zou voelen. De afgelopen jaren hebben dat vermoeden alleen maar bevestigd. Tijdens het grootste deel van mijn promotie zat jij in de kamer naast mij en Cordula, en stond jouw deur niet alleen figuurlijk maar ook letterlijk bijna altijd open. Hier heb ik dankbaar en veelvuldig gebruik van gemaakt om even snel iets te vragen of te bespreken als ik ergens tegen aan liep. Hartelijk bedankt voor jouw adviezen en begeleiding de afgelopen jaren. Hoewel wij nu in principe niet meer samenwerken, hoop ik dat jouw deur altijd figuurlijk open blijft staan.

Ook mijn tweede promotor, professor Rizopoulos. Beste Dimitris, de afgelopen jaren ben jij langzaamaan steeds meer bij mijn onderzoek betrokken geraakt. In het begin van mijn promotietraject hadden we feitelijk niet iets met elkaar te maken, maar toen kon ik al met mijn joint model vragen bij jou terecht. Toen Eric en ik tegen het case-cohort probleem waren aangelopen, was je bereid om mij hierin te begeleiden en later kreeg je ook een officiële rol in de begeleiding van mijn promotie. Inmiddels werk ik nu drie dagen per week op de afdeling biostatistiek. Ik zal de data niet verder extrapoleren over toekomstige samenwerkingen, maar ik hoop deze in elk geval altijd plezierig zullen blijven. Bedankt voor je begeleiding de afgelopen jaren. Ik heb veel van je kunnen leren over joint models, maar hoop ook iets over te nemen van jouw enthousiasme voor programmeren en je geduld wanneer wekenlang durende simulaties weer eens niet convergeerden.

Mijn co-promotor, dr. Kardys. Beste Isabella, ik wil jou voornamelijk bedanken voor jouw begeleiding in ons relatieve overleving avontuur. Dankzij jou heb ik mijn eerste paper tot een succesvol eindresultaat weten te brengen. Jij hebt mij geleerd om een goed paper te schrijven en ik vind het knap hoe jij passages - die ik tot tien keer toe onsuccesvol herschreven had - eenvoudig tot goedlopende zinnen om kan toveren.

Ik wil graag ook de leden van mijn commissie bedanken, zonder wie ik mijn proefschrift niet kan verdedigen. Bedankt Prof. dr. Steyerberg, prof. dr. ir. van der Schouw en prof. dr. Takkenberg voor het lezen van mijn proefschrift. Ook wil ik prof. dr. Roos-Hesselink, prof. dr. Eijkemans en dr. Lingsma bedanken dat jullie in de oppositie willen plaatsnemen.

Alle leden van het CREW-consortium, bedankt voor de samenwerking de afgelopen jaren. Vooral bedankt aan de CREW-promovendi voor de steun die we elkaar hebben kunnen geven tijdens dit grote project. Veerle en Luuk, mijn review buddies; wat ben ik blij dat ik kan schrijven dat de review af EN gepubliceerd is. Ik heb me goed vermaakt tijdens de dagen die we samen eraan hebben gewerkt, zowel in Amsterdam, Rotterdam als Utrecht. Als de technologie ooit zo ver is dat we niet meer zelf door alle abstracts heen hoeven te kammen, kunnen we ons misschien toch nog een keer aan een update wagen!

Ik wil graag alle co-auteurs bedanken voor de kans om mee te werken aan alle interessante onderzoeksprojecten. In het bijzonder Iris van Hagen, Susanna den Boer, Lennart van Gils en Laura van Vark; jullie onderzoeken vormen een belangrijk deel van mijn proefschrift. Laura, tijdens het analyseren van de TRIUMPH data had jij je eigen stoel bij ons op de kamer en zat je af en toe vaker achter mijn pc dan je eigen. Ondanks de analyse-frustraties was het erg gezellig en zijn er een paar mooie papers uitgekomen!

Mijn kamergenoot Cordula, ik kijk met veel plezier terug op de jaren tegen over elkaar in Ba-044. Met name jouw lunchboterhammen die om 9 uur 's ochtends al op waren, onze ladeblokken (eigenlijk bedoeld voor office supplies) die altijd gevuld waren met snacks en thee om de lange uren tussen 9 en 12 door te komen en natuurlijk de niet aflatende strijd tegen de monitor. Ik heb me dankzij jou goed vermaakt in ons

hok! De andere Cardio Phd's, bedankt voor de borrels en de kerstdiners. En de Klep promovendi; Sharda, Nermina, Milos, Maxime, Victor, Anne-Sophie, Elke, Michelle, Jan-Kees, Marie, Nathalie: Bedankt voor alle gezelligheid, biertjes, bitterballen, escape room avonturen en de (jaarlijkse?) Jacc party!

Mijn collega's in de Room Of Excellence. Sanne, ik ben heel blij dat ik met jou de SPSS cursus mag geven. Het is volgens mij niet vanzelfsprekend dat zo'n samenwerking zo soepel verloopt. We zijn nu ook (deeltijd)kamergenoten, en hoewel we elkaar eigenlijk alleen maandag zien, zijn mijn maandagen een stuk gezelliger daardoor. Mattie, bedankt voor de prettige samenwerking de afgelopen jaren, vooral tijdens het doorkammen van de statistiek van de statistical board papers. Paul, vrolijke noot in onze hardwerkende kamer. De kennis die ik tijdens mijn promotie heb op gedaan in het voorspellen van uitkomsten zal vast nog vaak gebruikt worden om uitslagen van bepaalde voetbalwedstrijden te voorspellen.
Ik wil ook graag Ron van Domburg bedanken. Beste Ron, wij hebben uiteindelijk alleen in het begin van mijn promotietraject samen gewerkt. Maar ik wil je heel erg bedanken voor je hulp de eerste paar maanden dat ik naast je op de kamer zat. Het was een hele fijne manier om mijn promotie te starten.
Yvonne, bedankt voor alle administratieve zaken en de onmogelijke taak om de CREW vergaderingen in te plannen.

I would also like to thank all my colleagues from the department of Biostatistics; Ani, Eline, Elrozy, Floor, Greg, Hongchao, Katya, Joost, Meng, Nicole, Sten, and honorary colleague Daan. Especially the ones who answered my statistical and equally important non-statistical questions during our Wednesday consulting/lunch meetings and helped me in the task of dealing with medical doctors. Additional thanks to Nicole, Greg, and my office mate Elrozy for the good times during the conferences and on the Katsuiberg!

Dobri Pet; Fren, Son, Claire, Maniet, met aanhang en inmiddels met een nieuwe generatie aanhang. Jullie hebben niets hoeven doen voor mijn proefschrift, maar toch horen jullie hier thuis. Bedankt voor de jarenlange vriendschap, steun, vele vakanties, etentjes, lunches, koffietjes, en steeds braver wordende (café) avondjes. Benne, je

woont inmiddels al jaren in Engeland, maar dat is nog niets in vergelijking met hoe lang onze vriendschap al duurt. Bedankt voor alle goeie herinneringen!

Mijn paranimfen; Frances en Thomas. Fren, wij kennen elkaar al sinds de peuterspeelzaal, zijn klasgenoten geweest op de basisschool en middelbare school, en huisgenoten in onze eerste studiejaren. Tijdens al die jaren zijn we altijd bevriend geweest en ik weet zeker dat dit altijd zal blijven. Thomas, je familie kies je niet zelf uit, maar ik heb je wel zelf uitgekozen om mijn paranimf te zijn. Hoewel wij in veel dingen erg van elkaar verschillen, hebben wij toch een band die je alleen met dichte familieleden kan hebben. Allebei bedankt dat jullie mijn paranimf willen zijn en naast mij willen staan!

Ik wil ook Sifu, Ying en iedereen van de LHM-familie bedanken, mijn tweede familie. Alles wat ik bij jullie heb geleerd, neem ik voor de rest van mijn leven mee. Salome, I'm glad you are my sister-in-law. Thank you for enlightening our lives with your Georgian charm and enthusiasm! Mijn schoonfamilie, bedankt voor jullie hartelijkheid de afgelopen jaren.

Natuurlijk wil ik ook mijn ouders bedanken voor alles. Lieve mam en pap, zonder jullie was het nooit gelukt om hier te staan en deze promotie tot een succesvol einde te brengen. Bedankt voor jullie onvoorwaardelijke steun in alles wat ik doe, het bieden van een luisterend oor en het geven van goede adviezen, die ik (natuurlijk) niet altijd opvolg.

Richard, lief, mijn laatste woorden zijn voor jou. Hoewel het er van buitenaf uitzag dat mijn promotietraject behoorlijk soepel verliep, ben jij degene die thuis mijn frustraties aan heb moeten horen. Bedankt voor het geduld waarmee je dit hebt gedaan. Daarnaast wil ik je ook bedanken dat je mijn cover heb willen designen. Op basis van mijn tekening heb je er -zoals verwacht- iets spectaculair moois van gemaakt. Ik ben blij dat mijn boekje nu ook iets van ons allebei is. Ik houd van je en kijk uit naar al onze avonturen die we verder gaan beleven in de Baarsjes en daarbuiten!