QUANTIFICATION OF THE IMPACT OF PREDICTION MODELS:

MODEL FIRST, TRIAL LATER

—

Giske Lagerweij

Quantification of the impact of prediction models: model first trial later.

ISBN: 978-90-393-7090-2
Author: Giske Lagerweij
Layout: Jules Verkade, persoonlijkproefschrift.nl.
Printed by: Ipskamp Printing, proefschriften.net

Quantification of the impact of prediction models:

model first, trial later

Bepalen van de impact van predictiemodellen:
modelleer eerst, experimenteer later
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen
op dinsdag 12 maart 2019 des middags te 4.15 uur

door

Ghizelda Rebecca Lagerweij

geboren op 21 maart 1987
te Amersfoort

Promotor: Prof. dr. K.G.M. Moons

Copromotoren: Dr. H. Koffijberg
Dr. G.A. de Wit

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.

Contents

1.

General introduction

Since crystal balls do not exist, risk prediction models are often used to assist in medical decision making. Diagnostic models estimate the risk of *current* presence of a certain outcome (e.g. disease) whereas prognostic models estimate their *future* presence [1, 2]. The Pooled Cohort Equations, the Framingham risk score, and SCORE risk chart are examples of prognostic prediction models that may be used to estimate the risk of experiencing a cardiovascular event in the future [3-5]. In Dutch general practices for example, the SCORE risk chart is advocated and mostly used. A general practitioner determines the risk profile of an individual based on various characteristics (e.g. age, gender, smoking behaviour, blood pressure and cholesterol levels) to estimate the 10-year cardiovascular disease (CVD) risk. [6, 7]. According to CVD guidelines, individuals at elevated risk are recommended to use preventive medication in order to reduce their risk of CVD events [6, 7].

Over the last decades, in all medical domains, risk prediction models have been researched extensively, resulting in ten thousands of newly developed and updated prediction models. An underlying assumption of this research is that any new or updated model with a better predictive performance, indeed leads to improved health outcomes and lower health care costs since individuals with elevated risks are identified and treated more precisely and timely. However, this is not always the case. It is always important to investigate how the use of the model affects actual medical decision making and behaviour [8, 9]. After all, individuals' health outcomes, health care provision, and its costs are only impacted by the actions that are taken following the use of the prediction model to estimate risks. The use of a prediction model itself is often relatively inexpensive, although it sometimes requires input based on various laboratory, imaging or even genetic tests, which obviously have costs associated with them.

Impact assessment of prediction models, both on decision making behaviour and on eventual health outcomes and care, is complex and rarely performed [9, 10]. One complicating factor is that prediction models may predict multiple endpoints, using a so-called *composite endpoint*. The predicted risks are thus related to a combination of endpoints, which each in turn have different associated health outcomes and costs. For example, the consequences of a stroke versus myocardial infarction, both often combined in the prediction of CVD events, differ substantially.

Moreover, prediction models are often used to recommend treatment for individuals with a risk estimate above a certain risk threshold, and no treatment otherwise. However, when there are competing prediction models for the same targeted population, but predicting a different (composite) endpoint, the models' predicted risks have different interpretations and may require different decision making and management. Indeed, the same individuals can be classified differently according to their predicted risk by different models.

Another complicating factor is that a classical prospective study to assess the impact of the use of prognostic prediction models is often cumbersome. Ideally a comparative randomized design is used where healthcare providers are randomized to either use or not use of the prediction model under study. However, depending on the prediction model and time needed to follow-up, such a study is often time consuming, costly and actual health effects are hard to measure in case of long term outcomes.

The aim of this thesis is to explore the challenges in research on the impact assessment of prediction models, and to provide solutions for these challenges. The thesis is separated in two parts. First, the challenges of using different endpoints in competing prediction models for the same targeted population are described. Second, challenges in the design of impact assessments of prediction models are presented.

Chapter 2 describes the complexity of estimating the burden of cardiovascular disease in a Dutch population cohort based on four different prognostic prediction models with varying endpoints. **Chapter 3** describes how the use of different prediction models with varying endpoints may lead to different treatment recommendations. **Chapter 4** describes how the decision for preventive treatment for CVD based on predicted probabilities of prediction models may be optimized by explicitly considering the expected burden of the long term endpoints at different ages in the prediction calculations.

As the use of long term randomized studies for impact assessment of prognostic prediction models is often not feasible, other methods to assess the impact of a prediction model are required. One alternative is to perform a decision analytical modelling approach. This approach is less time consuming and costly. However, it requires the availability of sufficient high-quality data and evidence on various aspects. **Chapter 5** describes the development of a decision analytic model to explore key elements of evidence in the impact assessment

of a prediction model-based CVD prevention strategy in young individuals. In **Chapter 6**, this model is applied in a full cost-effectiveness analysis of CVD prevention strategies in young women with a history of preeclampsia. In **Chapter 7,** we describe the added value of performing a decision analytical model approach to optimize the design of a planned randomized trial investigating the use of CVD risk prediction in an emergency care setting.

Finally, in **Chapter 8**, we describe the advantages and disadvantages of performing a decision analytical model approach to assess the impact of prediction models, along with case studies and recommendations.

REFERENCES

1. Moons, K.G., et al., *Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker.* Heart, 2012. 98(9): p. 683-90.
2. Grobbee, D.E., Hoes, A.W., *Clinical Epidemiology - Principles Methods and Applications for Clinical Research*. 2009, Londen: Jones and Bartlett Publishers.
3. D'Agostino, R.B., Sr., et al., *General cardiovascular risk profile for use in primary care: the Framingham Heart Study.* Circulation, 2008. 117(6): p. 743-53.
4. Conroy, R.M., et al., *Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project.* Eur Heart J, 2003. 24(11): p. 987-1003.
5. Goff, D.C., Jr., 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, 2014. 63(25 Pt B): p. 2935-59.
6. Smulders, Y.M., et al., *Clinical practice guideline for cardiovascular risk management in the Netherlands.* Neth J Med, 2008. 66(4): p. 169-74.
7. Nederlands Huisartsen Genootschap. *Cardiovascular risk management*. 2012; Available from: https://www.nhg.org/standaarden/volledig/cardiovasculair-risicomanagement.
8. Moons, K.G., et al., *Risk prediction models: II. External validation, model updating, and impact assessment.* Heart, 2012. 98(9): p. 691-8.
9. Reilly, B.M. and A.T. Evans, *Translating clinical research into clinical practice: Impact of using prediction rules to make decisions.* Ann Intern Med, 2006. 144(3): p. 201-209.
10. van Giessen, A., et al., *Systematic Review of Health Economic Impact Evaluations of Risk Prediction Models: Stop Developing, Start Evaluating.* Value Health, 2017. 20(4): p. 718-726.

2.

From predicted risk to predicted burden of cardiovascular disease

G.R. Lagerweij
G.A. de Wit
K.G.M. Moons
W.M.M. Verschuren
J.M.A. Boer
H. Koffijberg

Journal of Clinical Epidemiology; 2018

ABSTRACT

Objectives: To explore the extent of the differences in definitions of composite endpoints and assess how these differences influence estimates of CVD burden.

Study design and settings: Data from a Dutch cohort study (n=19,484) were used to calculate 10-year risks according to four CVD risk prediction models: ATP-III, Framingham (FRS), Pooled Cohort Equations (PCE) and SCORE. Health loss was estimated based on the impact of event types included in the corresponding composite endpoints. Finally, each prediction model was used to estimate the expected CVD burden in high-risk individuals, expressed as Quality-Adjusted Life Years (QALYs) lost.

Results: The definition of the composite endpoints varied widely across the four models. FRS predicted the highest CVD risks and the composite endpoint used in SCORE was associated with the highest health burden. The predicted CVD burden in high-risk individuals was 0.23, 0.74, 0.43, and 0.39 QALYs lost per individual when using ATP, FRS, PCE and SCORE, respectively.

Conclusion: The investigated CVD risk prediction models showed huge variation in definition of composite endpoints and associated health burden. Therefore, health consequences related to predicted risks cannot be readily compared across prediction models, and estimates of burden of disease depend crucially on the prediction model used.

1. INTRODUCTION

Cardiovascular disease (CVD) is one of the leading causes of morbidity and mortality worldwide [1]. The annual number of CVD related deaths is expected to increase from 17.5 million in 2012 to 23.3 million by 2030 [2]. The burden of disease including all CVD related health loss gives an indication of the overall health loss due to CVD in the population. This CVD burden can also be interpreted as the maximum health gain achievable by any preventive CVD intervention, such as lifestyle improvements, and pharmacotherapy. To increase the effectiveness of prevention strategies, these are increasingly based on CVD risk stratification, i.e. CVD risk prediction models are used to allocate individuals to predefined risk categories to tailor preventive interventions. Numerous CVD risk prediction models have been developed for individualized CVD risk prediction and risk classification [3-5]. For example, the Framingham risk equation classifies individuals with a ≤ 20% 10-year CVD risk as low risk, and individuals with a > 20% 10-year CVD risk as high risk, whereas the Pooled Cohort Equation (PCE) uses a 7.5% 10-year CVD risk threshold instead of 20% [3, 6]. These prediction models can be used to estimate the risk of CVD for individuals, but can also be used to estimate the CVD burden in (sub)groups of individuals, for example, in individuals classified as high-risk [7, 8]. CVD burden estimates can be derived by simply aggregating all risk estimates of individuals in the (sub)group to get the expected total number of CVD events in that (sub)group. As long as the prediction model used is calibrated to (sub)group of individuals, the total number of CVD events can be validly estimated by the sum of the individual risk estimates. Estimating the expected CVD burden then requires deriving the expected health loss caused by these CVD events. For example, experiencing a stroke may persistently lower quality of life, or even lead to death. However, different CVD risk prediction models may predict different CVD events. In fact, these models are commonly developed based on composite endpoints, including multiple and different types of CVD events. For example, one CVD risk prediction model may predict only (fatal or non-fatal) stroke, whereas another may predict only (fatal and non-fatal) myocardial infarction (MI). Often, even more complex composite endpoints are used, in which >10 different types of CVD events are combined. The use of composite endpoints may be favourable from a clinical perspective, because it is more relevant to predict a range of CVD related events rather than a single event, and may increase statistical power [9]. However, the use of complex composite CVD endpoints makes it hard to estimate the health loss related to that endpoint, unless the included, separate

CVD events are considered. When, in addition, different CVD risk prediction models use different composite endpoints, this would further complicate the robust assessment of the expected CVD burden in (sub)groups of individuals.

To explore the extent of this problem, the expected CVD burden is estimated in a large cohort using four widely used CVD risk prediction models. Firstly, we investigate the definition and constitution of the composite endpoints used in these CVD risk prediction models. Secondly, we estimate the CVD risk for all individuals in the cohort, and the health loss of the CVD events included in the composite endpoint, for each prediction model. Finally, we assess how the identified differences in composite endpoints in the prediction models considered influence the estimated CVD burden in this cohort.

2. METHODS

2.1 Constitution of composite endpoints in MORGEN

Seven widely used CVD risk prediction models were initially selected for this study: Adult Treatment Panel III (ATP), Framingham Global Risk Score (FRS), Pooled Cohort Equations (PCE), SCORE-low (SCORE) model, PROCAM, QRISK and Reynolds risk score [3, 10-16]**.** The models were chosen based on their largely overlapping subsets of easy to measure and frequently available risk factors, e.g. gender, age and systolic blood pressure. Furthermore, all models were derived from general population cohorts. All prediction models except SCORE are Cox proportional-hazards (PH) regression models, i.e. semi parametric survival models where the form of the baseline hazard is not specified. The SCORE model is a Weibull model, i.e. a fully parametric survival model. More information on these CVD risk prediction models can be found in Appendix A). All models estimate the absolute risk of a composite endpoint, occurring within 10 years. The exact definition of the composite endpoint was identified from background articles for each prediction model [3, 10-16] and translated in terms of ICD-10 codes for each model (see Appendix B).

We compared the composite endpoints of the seven CVD risk prediction models using a large population cohort (MORGEN) in the Netherlands. The MORGEN cohort includes men and women aged 20 to 74 years at baseline, recruited from the general population between 1993 and 1997 [17]. After a follow-up time of 10 to 15 years (average 12.3 years), participant information on vital status, cause of death and comorbidity was obtained through municipal registries, Statistics Netherlands, and from the National Medical Registry (NMR), respectively. To apply the prediction models, information on both the recruitment and follow-up was required, leaving 19,484 individuals with adequate data from the original cohort for the analysis. Information on the composition of this cohort and exclusion criteria for the current use of cohort data can be found in Appendix B.

To investigate the constitution of the composite endpoints, the observed rates and distributions of the individual components were determined for each model separately, using the set of ICD-10 codes comprising the composite endpoint (Appendix B). As the different prediction models have different composite endpoints, whether individuals are registered as experiencing a CVD event thus depends on the applied prediction model. Furthermore, due to censoring mechanisms that vary per prediction model, the observed rate for a specific CVD event may also vary per prediction model. Interpretation of a first and secondary event within individuals depends on whether such event is included in the composite endpoint of each prediction model.

2.2 Consequences of dissimilarities in composite endpoints

Assessment of dissimilarities in the consequences of the composite endpoints requires estimations of the predicted risks and consequences of the included individual components. As evidence on certain risk factors, such as family history of CHD, C-reactive protein and social deprivation, was not available within the MORGEN cohort, the predicted risks according to prediction model QRISK, PROCAM and Reynolds could not be estimated. Hence, these three models were excluded from further analyses. To assure accurate predicted risks, we first validated and recalibrated the remaining four CVD risk prediction models ATP, FRS, PCE and SCORE to the cohort data. For the survival data (time-to-event data) considered in this study, recalibrating a prediction model typically involves updating the baseline hazard and adjusting the mean values of the predictors (the linear predictor of the “average” patient) [18]. Note that this was only to ensure that the model was well fitted, as we do not focus on statistical performance.

The selected prediction models all result in a predicted risk for a 10-year time horizon, therefore follow-up time was truncated at 10 years prior to validation, recalibration and subsequent analyses. The overall performance of the original and recalibrated models was expressed in the Brier Score [18]. Furthermore, the calibration of both the original and recalibrated models was assessed and expressed in terms of a calibration plot, including estimating the slope and intercept of each plot, and Hosmer-Lemeshow chi-square statistic [18]. The discrimination of the original and recalibrated models was also assessed, using Harrell's c-statistic [19]. The discrimination measure indicates the accuracy of the model by ordering individuals by their risk, i.e. a subgroup with high-risk individuals should exhibit higher event rates than a low-risk subgroup [20].

The original CVD risk prediction models were developed with other data than used for this study, hence only for the recalibrated models the 10-year CVD risks were predicted per individual in the MORGEN cohort and presented for six risk categories: 0-2%, 2-4%, 4-6%, 6-8%, 8-10%, >10%. Although age is included as a risk factor in all models, the actual effect of age differs per model. As age was skewed to the right, it was not possible to use age values expressed in whole years to create deciles. Therefore, the comparison of predicted risks according to the different models was also presented for deciles of age: 20.1-26.5, 26.6-32.1, 32.2-36.7, 36.8-40.4, 40.5-43.5, 43.6-47.0, 47.1-50.3, 50.4-53.5, 53.6-57.4, and 57.5-73.7 years. We defined *low-risk individuals* as those with the lowest 25% predicted risks and *high-risk individuals* as those with the highest 25% predicted risks, regardless of their absolute predicted risk, per prediction model. Reclassification tables were constructed to determine whether high-risk individuals correspond among the CVD risk prediction models.

For measuring the consequences, i.e. the individualized (weighted) impact, of a "composite endpoint", Quality Adjusted Life Years (QALYs) were used. The QALY is a measure combining the length of life and quality of life (QoL) of individuals [21]. As morbidity and mortality due to disease decrease the number of QALYs experienced by individuals, burden of disease can be expressed in terms of QALY loss. To correct a year of life lived in a sub-optimal health status, i.e. following a CVD event, life years were weighted by a utility (value) for the QoL during that year. Evidence on QoL following different CVD event types was collected from a clinical guideline defined in 2014 by the National Institute for Health and Care excellence (NICE) [22]. This guideline presents utilities for different health states after a CVD event and a baseline utility for normal health by age (see Appendix C). The ICD-10 codes used to define all

CVD events were linked to corresponding utilities. Furthermore, information from Statistics Netherlands was used to determine the survival rates per gender and age category, for the years 2007-2012, after excluding mortality due to CVD events. These survival probabilities were applied to establish the average life expectancy per gender and for each age category, in absence of CVD.

Furthermore, for simplification, a persistent, lifetime impact of events was estimated based on the observed QoL following (partial) recovery of a CVD event (see Appendix C). The occurrence of multiple (recurrent) CVD events or other diseases was not taken into account. In addition, it was assumed that the CVD events (according to the predicted risks) occurred, on average, after five years (for details see Appendix C).

The overall estimated CVD burden of disease was assessed by combining predicted absolute (individualized) risks of an event with the consequences of the composite endpoint. The estimated overall CVD burden from each prediction model gives an indication of the expected health loss due to CVD events per individual, and can also be interpreted as the maximum health gain achievable by any preventive CVD intervention, according to the corresponding prediction model. In addition to assessing the CVD burden based on the consequences of the composite endpoint as defined per model, this burden was also assessed using the most comprehensive endpoint used in the four models.

3. RESULTS

3.1 Constitution of composite endpoints in MORGEN

Table 1 (column 1-2) shows that composite endpoints of the investigated prediction models are very different with varying types of individual components included. The definition and ICD-10 code per component is shown in column 1-2. Per prediction model, the type of individual components and observed number of individuals experiencing this component (event) is shown in Table 1 (column 3-16).FRS and QRISK had the highest observed numbers and largest variety in individual components as compared to the other CVD risk prediction models. All models include MI, either alone (ATP) or in combination with different sets of other manifestations of cardiovascular diseases (Figure 1). There was also a clear difference in the severity of the different components included, most notably mortality and morbidity. Furthermore, absolute numbers for SCORE were about eight times smaller than FRS, as SCORE only predicts fatal CVD events.

3.2 Consequences of dissimilarities in composite endpoints

Calibration and discrimination results for the original ATP, FRS, PCE and SCORE models and the recalibrated models, based on the endpoints as defined in Table 1, can be found in Appendix D. The performance of the four models is good and very similar; c-statistic of 0.81, 0.78, 0.78, and 0.81 for ATP, FRS, PCE and SCORE respectively. Moreover, the predicted number of events now closely matches the observed number of events, for each of the four models (Appendix D – Table 2). However, the observed differences in the definition of the composite endpoints, and type and number of individual components, directly led to large differences in predicted risks, as shown in Figure 2. Incorporation of more individual components into the composite endpoint automatically lead to higher predicted risks and prediction models focusing only on more severe events, e.g. SCORE provided lower predicted risks due to a lower incidence of such events. For the SCORE model, 90% of the individuals had a predicted risk lower than 2%, while according to FRS only 33% of the individuals were classified into this lowest risk category. The average predicted risks for the four prediction models are 1.4%, 5.9%, 2.2%, and 0.7% for APT, FRS, PCE and SCORE, respectively.

Figure 3A shows that differences in mean values of the predicted risks were already present at a young age, and became more pronounced at older age. Furthermore, the predicted risk increased supra-linear for all models, except ATP. Reclassification tables showed, however, that individuals identified as low and high risk still mostly correspond among the prediction models (Appendix E). The consequences of the composite endpoint (in terms of QALYs lost) according to prediction model SCORE was expected to be highest due to the severity of the incorporated individual components, i.e. only fatal CVD events. For the other models, the consequences of the composite endpoints were much lower and in the same order of magnitude. For all models, the risk and consequences of the composite endpoint was assessed per individual, based on age and gender dependent CVD patterns. For example CVD burden decreased with age, even though the risk of fatal versus non-fatal events increases with age, due to decreasing life expectancy (see Appendix F). SCORE showed the most rapid decrease in consequences of the composite endpoint (Figure 3B). Figure 3C illustrates the results for the predicted individualized CVD burden per individual, i.e. the maximum potentially preventable health loss per individual from CVD, as function of age. The predicted CVD burden is highest for FRS, at all ages, and is relatively stable with age for ATP and PCE. The predicted CVD burden for SCORE was highly age dependent, resulting in a very low predicted burden at young age, which was even lower than ATP. At older age, the predicted burden for SCORE was substantial, much higher than ATP and PCE.

The expected CVD burden in the high-risk individuals is 0.23, 0.74, 0.43, and 0.39 QALYs lost per individual for ATP, FRS, PCE and SCORE, respectively (Appendix F). Hence, FRS predicts a CVD burden 1.9 times as high as SCORE. This large variation in burden is caused by the differences in composite endpoints.

Figure 4A illustrates that a predicted risk according to ATP results in a lower CVD burden per individual than a similar predicted risk according to PCE due to the different composite endpoints. Of the four models considered, the Framingham model used the most comprehensive endpoint (Table 1). Using the Framingham composite endpoint to predict the CVD burden in the high-risk individuals, resulted in 0.74, 0.74, 0.72, and 0.65 QALYs lost per individual for ATP, FRS, PCE and SCORE, respectively (Figure 4B).

4. DISCUSSION

In this study, the definitions and constitution of composite endpoints for four widely used CVD risk prediction models, ATP, FRS, PCE and SCORE, have been investigated regarding both the number and type of CVD events included. Results indicate that these CVD risk prediction models vary substantially regarding the definition of their composite endpoint, that is, they include different sets of CVD event types (individual components). This variation in individual components induces large differences in predicted risk, i.e. individuals in our cohort have different predicted CVD risks according to these four prediction models. However, the group of individuals classified as high-risk is very similar when different prediction models are used. The variation in included individual components also induces a large variation in the expected health loss associated with the occurrence of a composite endpoint across prediction models. In addition, the estimated CVD burden is highly age dependent when applying SCORE [11, 13]. Consequently, the estimated CVD burden in individuals classified as high-risk in our cohort varies widely, with FRS predicting a 1.9 times higher burden than SCORE.

Previous (clinical) research has shown that the use of composite endpoints in studies may be more relevant to patients and clinicians as they cover more aspects and outcomes of the disease [23]. The usefulness of composite endpoints in the context of randomized trials, however, is still debated, due to the ensuing difficulty of interpreting differences in 'sets of outcomes' [24-30]. Moreover, even commonly used prediction models, such as the four models considered here, often have hard to find, or unclear, definitions of the composite endpoint in terms of ICD codes included. This affects a direct comparison of CVD risk prediction models, as each different composite endpoint has to be unravelled into its individual components, and each component has to be linked to a unique disease code. This process complicates the statistical analysis, e.g. evaluation, comparison and external validation of prediction models. Still, a transparent description of the composite endpoint and incorporated components is unavoidable to 1) translate changes in statistical prediction performance to expected health benefits for individuals, and 2) estimate the expected health benefits from new risk-based preventive interventions [31-33]. For example, assuming that preventive statin treatment reduces the risk of a *composite* CVD endpoint by a certain percentage will result in estimated health benefits which are highly dependent on the prediction model used [34]. Appropriate impact analysis of risk-based preventive

interventions require evidence of a) the initial risk of different types of CVD events, b) their consequences, and c) how the intervention reduces these risks. Finally, standardization of impact analysis in a single disease area also requires including the exact same (broad set of) event types in all such analyses, to make impact aspects comparable.

Strengths

Four widely used CVD risk prediction models are compared regarding their composite endpoints, their risk estimates and the associated burden of disease. Furthermore, this study unambiguously links all CVD endpoints of interest to ICD-10 codes, thereby improving clarity and ensuring replicability of the analyses in other cohorts. In addition, the size of the dataset used allowed for stratified analyses per risk- and age category. Finally, following from the recalibration, the prediction models considered have similar statistical performance, and the group of individuals categorized as high-risk is very similar across the prediction models. The large differences regarding predicted CVD risks and CVD burden can therefore reliably be attributed to differences in the constitution of their composite endpoints.

Limitations

The actual results from this study are dependent on the dataset used, i.e. the observed differences between CVD risk prediction models may be different in other datasets and populations. The cohort used consists of relatively young and healthy individuals, so even high-risk individuals have few CVD events. Thus, all predicted absolute risks are low compared with typical categories for high-risk individuals. However, the presented analyses can easily be generalized to other populations. Moreover, the methodology can also be applied to other disease areas in which composite endpoints are common such as for example the C-WATCH risk score for upper gastrointestinal bleeding [35]. Further analyses in other disease areas require large individual patient datasets with long follow up and accurate registration of all event types included in the prediction models, as well as registration of sequences of events in individuals.

For the translation of composite endpoints into ICD-10 codes, certain assumptions are required due to unclear definitions of the composite endpoints in the original publications. For PCE and ATP the defined endpoints "non-fatal MI and CHD death" are translated in "non-fatal and fatal MI" for consistency reasons. CVD risk prediction models PCE and ATP are both based on a formal Framingham prediction model, with ATP defining composite

endpoints “hard CHD” as developing a MI or MI death event, whereas PCE does not clearly specify the definition of “hard CHD”. These assumptions may have led to slight underestimations of predicted risks and consequences and therefore the overall predicted CVD burden. They are, however, unavoidable when unclear definitions of events needs to be linked to unique disease codes. In this study, we only accounted for the first CVD event in individuals even though in practice individuals may experience multiple CVD events. This limitation will lead to underestimation of the CVD burden but was necessary because the CVD risk prediction models considered are only validated for predicting first CVD events and are not appropriate for estimating the risk of recurrent CVD events [36].

Recommendations

Firstly, it is recommended that developers of CVD risk prediction models with a composite endpoint clearly describe the definition of that composite endpoint, as well as all its individual components, and their incidence in the development cohort. Secondly, studies comparing (the performance of) different prediction models should clearly describe the dataset(s) used and the link defined between the composite endpoints and the disease codes, preferably using the most recent ICD codes. Finally, impact assessments of preventive interventions should separate the individual components, and include their respective health consequences and costs, rather than focus on the composite endpoint.

Conclusions

Our results suggest that the number of different composite endpoints and included individual components used in CVD risk prediction models may almost be as large as the actual number of models itself. Furthermore, many CVD risk prediction models have unclear or hard to establish definitions of the composite endpoint in terms of ICD codes included. Hence, estimating the CVD burden using risk prediction models is not straightforward, and results should be interpreted with caution as they are highly dependent on the prediction model used. When using prediction models that include only a very limited set of CVD events, such as SCORE (fatal events only) and ATP (only MI), both the estimated CVD burden and the health benefits from preventive intervention will be underestimated. Moreover, the estimated health impact of preventive interventions may be biased if too narrow composite outcomes are used to estimate health benefits, or too narrow endpoints are used to reflect risks and side effects from such treatments. Whereas a broad common set of endpoints may be defined to reflect health benefits of preventive strategies in CVD, this may not be

feasible or useful for the negative consequences of treatment, as different treatments may have widely different negative side effects. More comprehensive prediction models, such as for example FRS and QRISK, cover more manifestations of CVD and might therefore yield more meaningful estimates regarding the (preventable) burden of CVD.

Table 1: Individual components and structure of composite endpoints in cohort

Individual components	ICD-10 code	ATP		FRS		PCE		SCORE		QRISK		PROCAM		Reynolds	
			#		#		#		#		#		#		#
Morbidity															
Myocardial infarction (MI)	I21,I22	X	232	X	208	X	223			X	208	X	217	X	223
Other Coronary heart disease (OCHD)	I20,I23,I24,I25			X	435					X	435				
Cardiac arrest, sudden death	I46,R96			X	4					X	4	X	4		
Haemorrhagic stroke (CVAH)	I60,I61,I62			X	41	X	41			X	41			X	41
Ischemic stroke (CVAI)	I63,I65			X	72	X	76			X	72			X	76
Other stroke (OCVA)	I64,I66			X	33	X	34			X	33			X	34
Other Cardiovascular diseases (OCVD)	G45,I67,I69,I70-I74,I50			X	267					X	267				
Total observed events			**232**		**1060**		**374**		**0**		**1060**		**221**		**374**
Mortality															
Myocardial infarction (MI)	I21,I22	X	50	X	43	X	50	X	61	X	43	X	50	X	50
Other Coronary heart disease (OCHD)	I20,I23,I24			X	6			X	17	X	6				
Cardiac arrest, sudden death	I46,R96			X	10			X	13	X	10	X	91		
Haemorrhagic stroke (CVAH)	I60,I61,I62			X	6	X	6	X	16	X	6			X	6
Ischemic stroke (CVAI)	I63,I65			X	3	X	3	X	6	X	3			X	3
Other stroke (OCVA)	I64,I66			X	2	X	3	X	3	X	2			X	3
Other Cardiovascular diseases (OCVD)	G45,I67,I69,I70-I74,I50			X	18			X	25	X	18				
Total observed events			**50**		**88**		**62**		**141**		**88**		**141**		**62**
Composite endpoints (morbidity + mortality)															
Ischemic Heart disease (IHD)	I20-I25														
Coronary heart disease (CHD)	I20-I25,I46,R96														
Cerebrovascular accident (CVA)	I60-I66					X								X	
Cardiovascular disease (CVD)	I20-I26,I46,R96,G45,I60-I67, I69,I70-I74,I50			X				X (only fatal events)		X					
Overall observed events			**282**		**1148**		**436**		**141**		**1148**		**362**		**436**

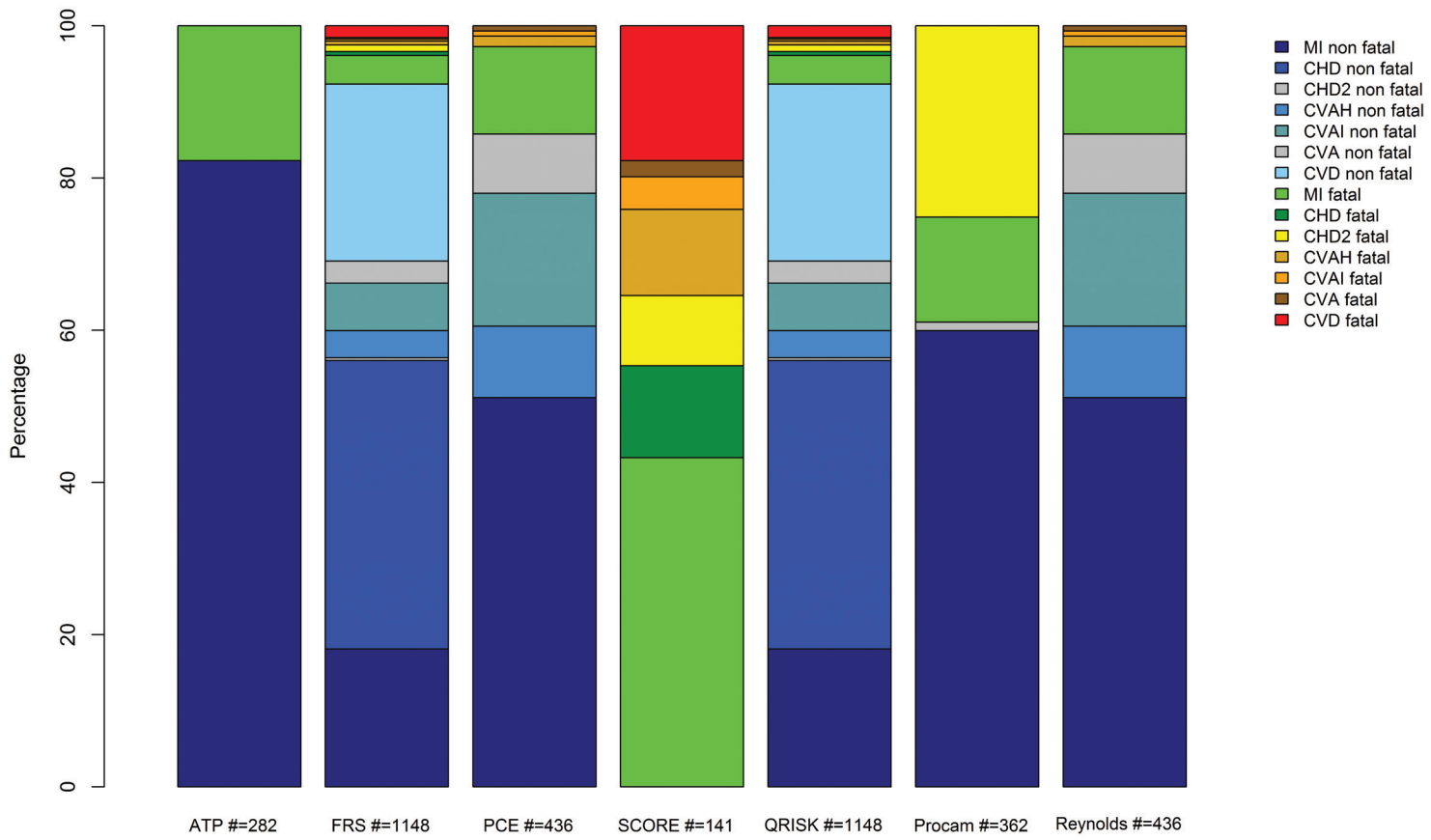


Figure 1: Overall distribution of included individual components per CVD risk prediction model

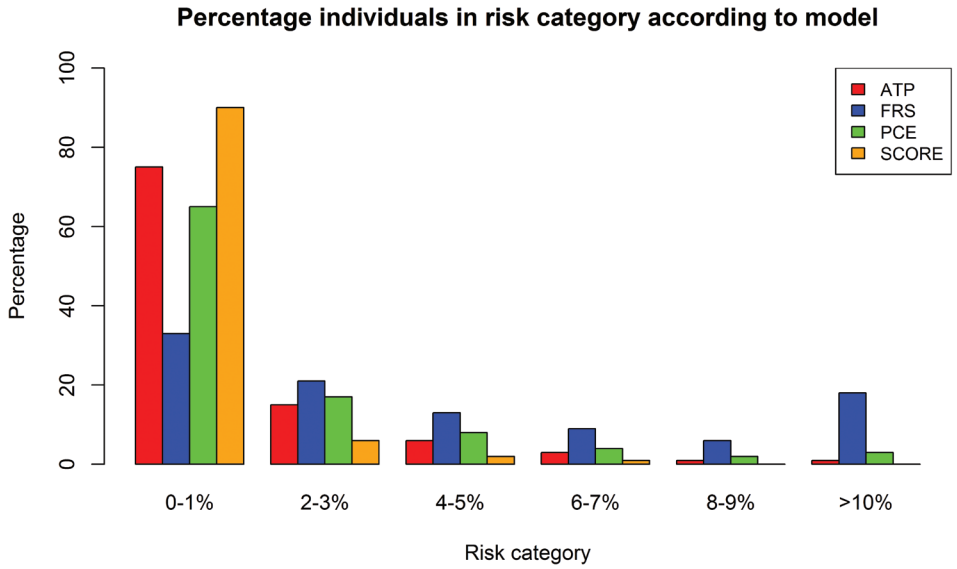


Figure 2: Distribution of individuals per risk category and CVD risk prediction model

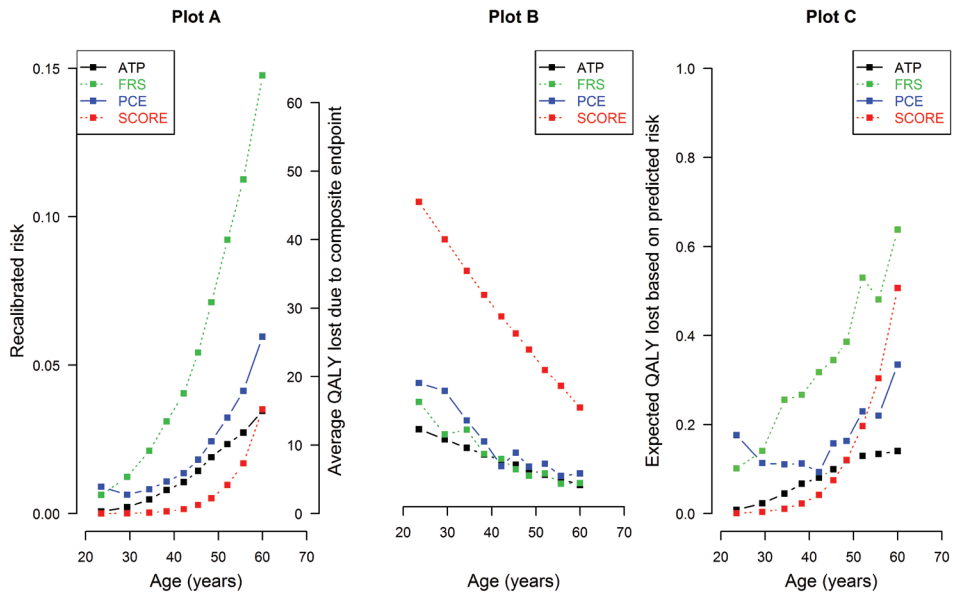


Figure 3: Three figures as function of age, with Plot A) the 10-year CVD predicted risks, Plot B) the expected (lifetime) consequence of a composite endpoints per individual, and Plot C) the expected (potentially preventable) CVD burden per individual[1].

1 Distribution of individual components was evaluated per age category, except for ATP where this distribution was assessed in the entire population due to limited number of included endpoints

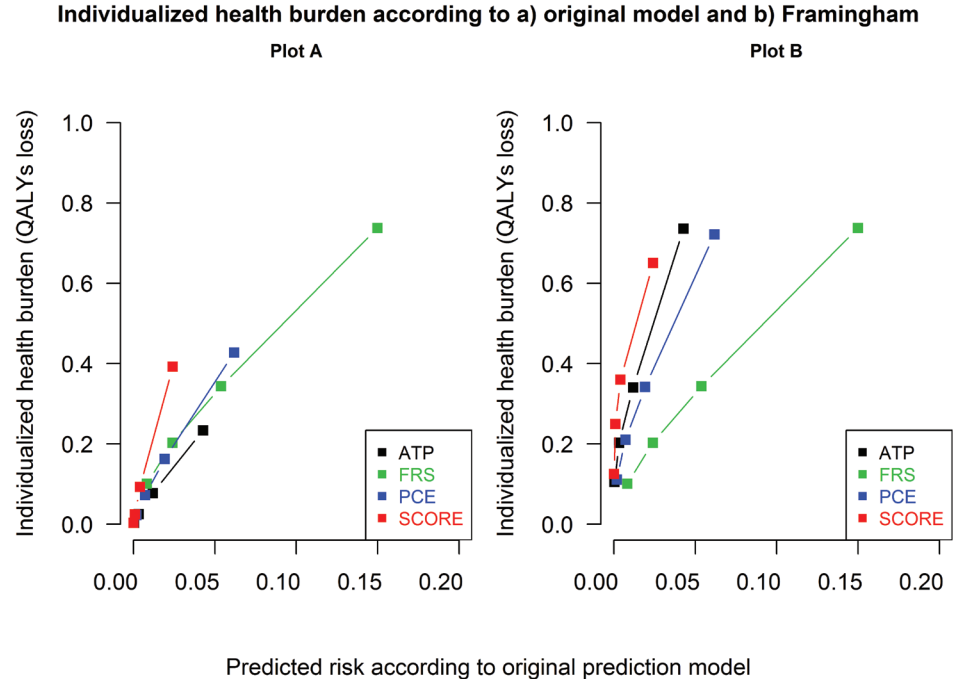


Figure 4: Individualized CVD burden[2] (QALYs loss) according to Plot A) the original endpoint and Plot B) comprehensive endpoint (Framingham model).

2 The CVD burden was estimated for categories based on risk quartiles.

REFERENCES

1. Go, A.S., D. Mozaffarian, V.L. Roger, E.J. Benjamin, J.D. Berry, W.B. Borden, et al., *Executive summary: heart disease and stroke statistics--2013 update: a report from the American Heart Association.* Circulation, 2013. 127(1): p. 143-52.
2. *Cardiovascular disease*. [cited 2017 10-19]; Available from: http://www.who.int/cardiovascular_diseases/about_cvd/en/.
3. Goff, D.C., Jr., D.M. Lloyd-Jones, G. Bennett, S. Coady, R.B. D'Agostino, R. Gibbons, 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.* Circulation, 2014. 129(25 Suppl 2): p. S49-73.
4. Hippisley-Cox, J., C. Coupland, J. Robson, and P. Brindle, *Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database.* BMJ, 2010. 341: p. c6624.
5. Pencina, M.J., R.B. D'Agostino, Sr., M.G. Larson, J.M. Massaro, and R.S. Vasan, *Predicting the 30-year risk of cardiovascular disease: the framingham heart study.* Circulation, 2009. 119(24): p. 3078-84.
6. Grundy, S.M., J.I. Cleeman, C.N. Merz, H.B. Brewer, Jr., L.T. Clark, D.B. Hunninghake, et al., *Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III Guidelines.* J Am Coll Cardiol, 2004. 44(3): p. 720-32.
7. Hennekens, C.H., Buring, J.E., *Epidemiology in Medicine*. First ed. 1987: Lippincott Williams & Wilkins.
8. Bailey, L., Vardulaki, K., Langham, J., Chandramohan D., *Introduction to Epidemiology*. First ed. 2005: Open University Press.
9. Lim, E., A. Brown, A. Helmy, S. Mussa, and D.G. Altman, *Composite outcomes in cardiovascular research: a survey of randomized trials.* Ann Intern Med, 2008. 149(9): p. 612-7.
10. *Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III).* JAMA, 2001. 285(19): p. 2486-97.
11. D'Agostino, R.B., Sr., R.S. Vasan, M.J. Pencina, P.A. Wolf, M. Cobain, J.M. Massaro, et al., *General cardiovascular risk profile for use in primary care: the Framingham Heart Study.* Circulation, 2008. 117(6): p. 743-53.
12. Ridker, P.M., J.E. Buring, N. Rifai, and N.R. Cook, *Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score.* JAMA, 2007. 297(6): p. 611-9.
13. Conroy, R.M., K. Pyorala, A.P. Fitzgerald, S. Sans, A. Menotti, G. De Backer, et al., *Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project.* Eur Heart J, 2003. 24(11): p. 987-1003.

14. Hippisley-Cox, J., C. Coupland, Y. Vinogradova, J. Robson, M. May, and P. Brindle, *Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study.* BMJ, 2007. 335(7611): p. 136.
15. Assmann, G., P. Cullen, and H. Schulte, *Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Munster (PROCAM) study.* Circulation, 2002. 105(3): p. 310-5.
16. Ridker, P.M., N.P. Paynter, N. Rifai, J.M. Gaziano, and N.R. Cook, *C-reactive protein and parental history improve global cardiovascular risk prediction: the Reynolds Risk Score for men.* Circulation, 2008. 118(22): p. 2243-51, 4p following 2251.
17. Beulens, J.W., E.M. Monninkhof, W.M. Verschuren, Y.T. van der Schouw, J. Smit, M.C. Ocke, et al., *Cohort profile: the EPIC-NL study.* Int J Epidemiol, 2010. 39(5): p. 1170-8.
18. Steyerberg, E.W., A.J. Vickers, N.R. Cook, T. Gerds, M. Gonen, N. Obuchowski, et al., *Assessing the performance of prediction models: a framework for traditional and novel measures.* Epidemiology, 2010. 21(1): p. 128-38.
19. FE, H. Regression Modeling Strategies. 2001, New York: Springer.
20. Royston, P. and D.G. Altman, *External validation of a Cox prognostic model: principles and methods.* BMC Med Res Methodol, 2013. 13: p. 33.
21. Philipp, C.S. *Health economics* [cited 2017 10-19]; 2nd [Available from: http://www.bandolier.org.uk/painres/download/What%20is%202009/What_is_a_QALY.pdf.
22. *Lipid Modification: Cardiovascular Risk Assessment and the Modification of Blood Lipids for the Primary and Secondary Prevention of Cardiovascular Disease*, 2014, NICE: London.
23. Ferreira-Gonzalez, I., P. Alonso-Coello, I. Sola, V. Pacheco-Huergo, A. Domingo-Salvany, J. Alonso, et al., *[Composite endpoints in clinical trials].* Rev Esp Cardiol, 2008. 61(3): p. 283-90.
24. Montori, V.M., G. Permanyer-Miralda, I. Ferreira-Gonzalez, J.W. Busse, V. Pacheco-Huergo, D. Bryant, et al., *Validity of composite end points in clinical trials.* BMJ, 2005. 330(7491): p. 594-6.
25. Ross, S., *Composite outcomes in randomized clinical trials: arguments for and against.* Am J Obstet Gynecol, 2007. 196(2): p. 119 e1-6.
26. Chi, G.Y., *Some issues with composite endpoints in clinical trials.* Fundam Clin Pharmacol, 2005. 19(6): p. 609-19.
27. Kip, K.E., K. Hollabaugh, O.C. Marroquin, and D.O. Williams, *The problem with composite end points in cardiovascular studies: the story of major adverse cardiac events and percutaneous coronary intervention.* J Am Coll Cardiol, 2008. 51(7): p. 701-7.
28. Carneiro, A.V., *Composite outcomes in clinical trials: uses and problems.* Rev Port Cardiol, 2003. 22(10): p. 1253-63.
29. Huque, M.F., M. Alosh, and R. Bhore, *Addressing multiplicity issues of a composite endpoint and its components in clinical trials.* J Biopharm Stat, 2011. 21(4): p. 610-34.
30. Cordoba, G., L. Schwartz, S. Woloshin, H. Bae, and P.C. Gotzsche, *Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review.* BMJ, 2010. 341: p. c3920.

31. Gondrie, M.J., K.J. Janssen, K.G. Moons, and Y. van der Graaf, *A simple adaptation method improved the interpretability of prediction models for composite end points.* J Clin Epidemiol, 2012. 65(9): p. 946-53.

32. Glynn, R.J. and B. Rosner, *Methods to evaluate risks for composite end points and their individual components.* J Clin Epidemiol, 2004. 57(2): p. 113-22.

33. Ferreira-Gonzalez, I., G. Permanyer-Miralda, J.W. Busse, D.M. Bryant, V.M. Montori, P. Alonso-Coello, et al., *Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns.* J Clin Epidemiol, 2007. 60(7): p. 651-7; discussion 658-62.

34. Heeg, B.M. and B.A. van Hout, *Assessing uncertainties surrounding combined endpoints for use in economic models.* Med Decis Making, 2014. 34(3): p. 300-10.

35. Hoffmann, V., H. Neubauer, J. Heinzler, A. Smarczyk, M. Hellmich, A. Bowe, et al., *A Novel Easy-to-Use Prediction Scheme for Upper Gastrointestinal Bleeding: Cologne-WATCH (C-WATCH) Risk Score.* Medicine (Baltimore), 2015. 94(38): p. e1614.

36. Wattanakit, K., A.R. Folsom, L.E. Chambless, and F.J. Nieto, *Risk factors for cardiovascular event recurrence in the Atherosclerosis Risk in Communities (ARIC) study.* Am Heart J, 2005. 149(4): p. 606-12.

37. Struijk, E.A., A.M. May, N.L. Wezenbeek, H.P. Fransen, S.S. Soedamah-Muthu, A. Geelen, et al., *Adherence to dietary guidelines and cardiovascular disease risk in the EPIC-NL cohort.* Int J Cardiol, 2014. 176(2): p. 354-9.

38. InterAct, C., A.M. Spijkerman, A.D. van der, P.M. Nilsson, E. Ardanaz, D. Gavrila, et al., *Smoking and long-term risk of type 2 diabetes: the EPIC-InterAct study in European populations.* Diabetes Care, 2014. 37(12): p. 3164-71.

APPENDICES

A. PREDICTION MODELS

A.1 - Risk factors included in CVD risk prediction models

Appendix A - Table 1: The risk factors used in this study, with their description and unit, by risk prediction model

Risk factor	Description	Unit	ATP	FRS	PCE	SCORE	Qrisk	Procam	Reynolds
Gender	Gender of respondent	0=male, 1=female	X	X	X	X	X*		X*
Age	Age at baseline (rounded)	Years	X	X	X	X	X	X	X
Total cholesterol (TC)	Baseline level, single measurement	mg/dL	X	X	X	X	X		X
High-density lipoprotein cholesterol (HDL-C)	One measurement	mg/dL	X	X	X		X	X	X
Low-density lipoprotein cholesterol (LDL-C)	One measurement	mg/dL						X	
Systolic blood pressure (SBP)	Average SBP of two measurements at baseline	mmHg	X	X	X	X	X	X	X
Treatment (trt)	Use of medication for high blood pressure	0 = no, 1 = yes	X	X	X		X		
Smoking (smok)	Current smoking status, former smoker is non-smoker	0 = no, 1 = yes	X	X	X	X	X	X	X
Diabetes (diab)	Diabetes mellitus	0 = no, 1 = yes	X	X	X			X	
Risk region (region)	Low-risk or high-risk region in Europe	0 = low, 1 = high				X	X		
Body mass index (BMI)	Body mass index	kg/m2					X		
Family history of CHD	Family history of CHD in first degree under the age of 60 years	0 = no, 1 = yes					X	X	X
Deprivation score (Townsend)	Townsend deprivation score (2001 census data)	-					X		
C-reactive protein (CRP)	C-reactive protein	mg/L							X
Haemoglobin A1c (HbA1c)	Haemoglobin A1c	-							X (only for female)

* Gender is not incorporated in the model, but there are two separate models for men and women.

A.2 - Cox proportional hazard model

A general CVD risk function, from a Cox-PH model, looks like

$$\hat{p} = 1 - S_0\text{-}(t)^{e^{\left(\sum_{i=1}^{p} \beta_i X_i - \sum_{i=1}^{p} \beta_i \bar{X}_i\right)}}$$

where $S_0(t)$ is the baseline survival at follow-up time (t = 10 years in this case), β_i are the Cox regression coefficients and X_i are the (log-transformed) predictors or risk factors. The value of $S_0(t)$ and the (values of the) regression coefficients β_i differ for each prediction model.

A.3 - Weibull proportional hazard model

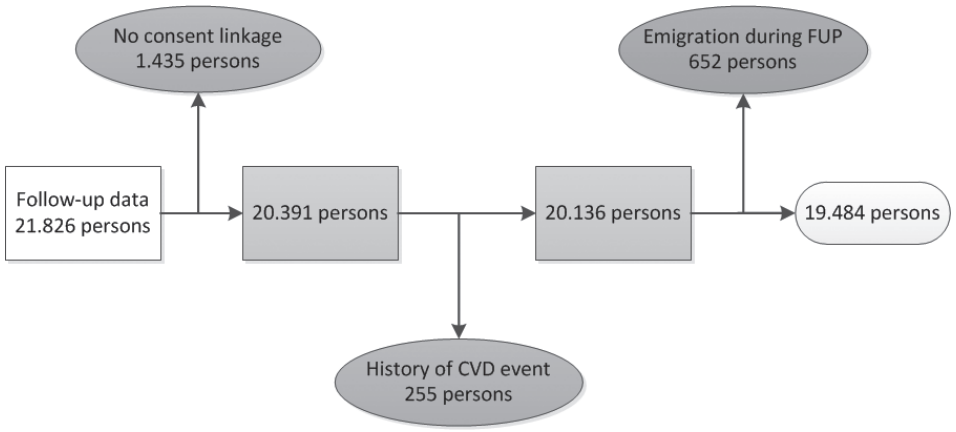
First, the shape of the baseline survival is modelled and then the relative risks regarding the risk factors are calculated. The SCORE model exists of a CHD part and a non CHD part, i.e. different sub models are used to estimate the probability of different endpoints within the composite endpoint.

Appendix A - Table 2 shows the risk formulas of the SCORE model. Here α and g differ for CHD, non-CHD, gender and low/high risk region. The regression coefficients β differ for CHD and non-CHD.

Appendix A - Table 2: Risk formulas SCORE

Shape baseline hazard	$S_0(age) = e^{-e^{\alpha}(age-20)^g}$
	$S_0(age+10) = e^{-e^{\alpha}(age-10)^g}$
Relative risk	$w = \beta_{TCL} \cdot (TCL-6) + \beta_{SBP} \cdot (SBP-120) + \beta_{smok} \cdot (smok)$
CHD Risk	$CHD\_Risk_{10}(age) = 1 - \frac{S_0(age+10)^{e^w}}{S_0(age)^{e^w}}$
Total CVD risk	$CVD\_Risk_{10}(age) = CHD\_Risk_{10}(age) + non\_CHD\_Risk_{10}(age)$

B. STUDY POPULATION



Appendix B - Figure 1: Flowchart of the inclusion and exclusion of individuals from the MORGEN cohort; the white rectangle presents the inclusions, the grey eclipses present the exclusions, the grey rectangles show the sub totals and in the light grey oval the final result is shown.

Appendix B - Table 1: Baseline characteristics of the MORGEN cohort, N = 19.484

	Male	Female
Gender (m / f)	8.855 / -	- / 10.629
Age (years)	43.4 ± 11.02	42.5 ± 11.29
Total cholesterol (mg/dL)	205.8 ± 41.30	204.2 ± 40.82
High-density lipoprotein	45.9 ± 11.62	58.4 ± 14.51
Systolic blood pressure (mmHg)	125.2 ± 15.24	117.9 ± 16.28
Treatment for hypertension (%)	4.2	5.1
Current smoking (%)	35.6	35.7
Diabetes (%)	13.1	10.8
Follow-up time (years)	11.82 ± 2.71	12.20 ± 2.16

All information was (re)coded according to ICD-10 [17, 37], as information recruited before 1997 was coded using ICD-9 codes. Therefore, in this study only ICD-10 codes were used.

Appendix B - Table 2: Translation ICD codes

	ICD 9	ICD 10
Coronary heart disease (CHD)	410-414	
Ischematic heart disease (IHD)	427.5	I20-I25
Cardiac arrest		I46
Sudden death		R96
Stroke	430-434, 436	I60-I66
Cardiovascular disease (CVD)	410-414, 427.5, 428, 415.1, 443.9, 430–438, 440–442, 444, 798.1, 798.2, 798.9	I20–I25, I46, R96, G45, I60–I67, I69, I70–I74, I50

C. BURDEN OF DISEASE

The impact ($I_{i,j}$), in terms of QALY loss, when an event predicted by model *j* occurs in individual i is given by

$$I_{i,j} = \left(\sum_{n=age_i}^{age_i+LE_i} QoL_n \right) \cdot \sum_{k=1}^{\Theta} \left(\pi_{i,k,j} \cdot (1 - u_k) \right)$$

with the number of individual components Θ under consideration, average life expectancy LE_i after surviving an event with remaining quality of life QoL_n, utility u_k and probability $\pi_{i,k,j}$ that an event predicted by prediction model j is of a specific type *k*.
Here, the left component represents the remaining life years – adjusted for their quality – in the absence of CVD events. The right component represents the total expected loss in quality of life due to all predicted CVD events in model *j*. Note that $\pi_{i,k,j}$ is zero for any type of event *k* not included in the composite endpoint of model *j*.
The quality of life of individual i with age n is given by

$$QoL_n = -0.00425 \cdot n + 1.06$$

The formula for the expected CVD burden of disease ($BD_{i,j}$) is given by

$$BD_{i,j} = I_{i,j} \cdot r_{i,j}$$

where $r_{i,j}$ is the recalibrated predicted risk for individual i and prediction model j.

Appendix C - Table 1 - Utility values [38]

	Utility
Morbidity	
Myocardial infarction (MI)	0.88
Other Coronary heart disease (OCHD)	0.88
Cardiac arrest	0.81
Ischemic stroke (CVAI)	0.63
Haemorrhagic stroke (CVAH)	0.63
Other stroke (OCVA)	0.63
Other Cardiovascular diseases (OCVD)	0.68
Mortality	
All seven cardiovascular events	0

D. CALIBRATION AND DISCRIMINATION

Appendix D - Table 1: Discrimination and calibration of original and recalibrated models[3]

Model		Intercept	Slope	Chi-square	p-value	C-statistic	95% low limit	95% high limit
ATP	*Original*	0.002	0.331	326.3	0.00	0.804	0.759	0.850
	Recalibrated	0.001	0.956	22.01	0.01	0.807	0.762	0.852
FRS	*Original*	-0.001	0.855	45.9	0.00	0.781	0.756	0.806
	Recalibrated	-0.002	1.054	10.98	0.28	0.782	0.757	0.807
PCE	*Original*	0.002	0.57	130.65	0.00	0.779	0.738	0.820
	Recalibrated	0.002	0.943	28.68	0.00	0.779	0.738	0.820
SCORE	*Original*	0.001	0.819	79.46	0.00	0.813	0.747	0.879
	Recalibrated	0.002	0.719	75.42	0.00	0.805	0.740	0.871

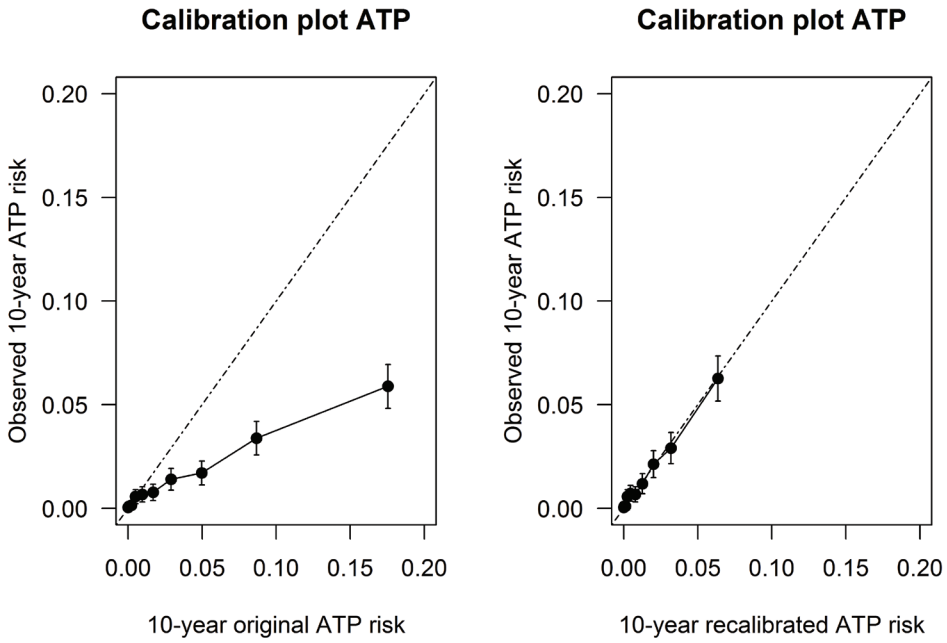
3 Appendix D - Table 1 shows the statistical performance of the original and recalibrated models with the estimated average value for slope and intercept (column 1-2) corresponding with Appendix D - Figure 1 - 4. Calibration plot ATP1–4. A good calibration fit has an intercept and slope estimate of 0 and 1, respectively.

Appendix D - Table 2: Overall observed and expected number of CVD events

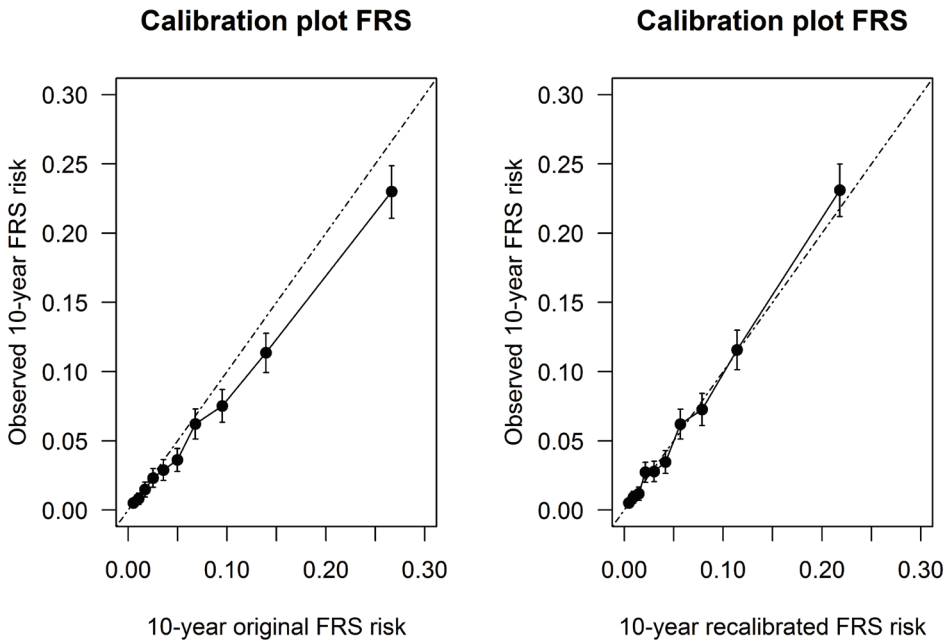
	# Observed CVD events	# Expected CVD events	
		Original model	*Recalibrated model*
ATP	282	732	282
FRS	1148	1,388	1,148
PCE	436	711	436
SCORE	141	144	141

Appendix D - Table 3: Overall performance measured by the Brier Score for the original and recalibrated models

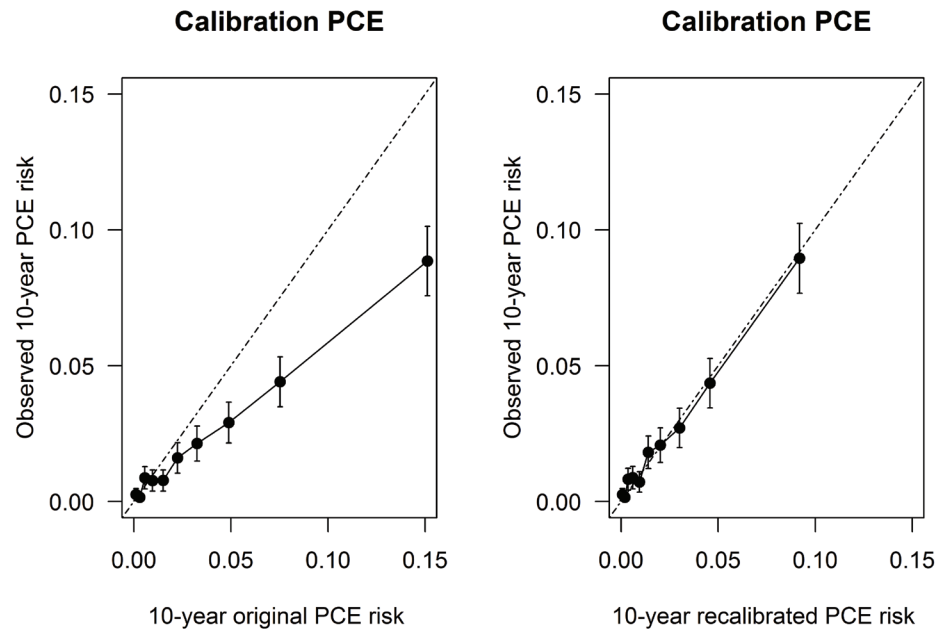
	Original model	Recalibrated model
ATP	0.016	0.014
FRS	0.051	0.051
PCE	0.022	0.021
SCORE	0.007	0.007



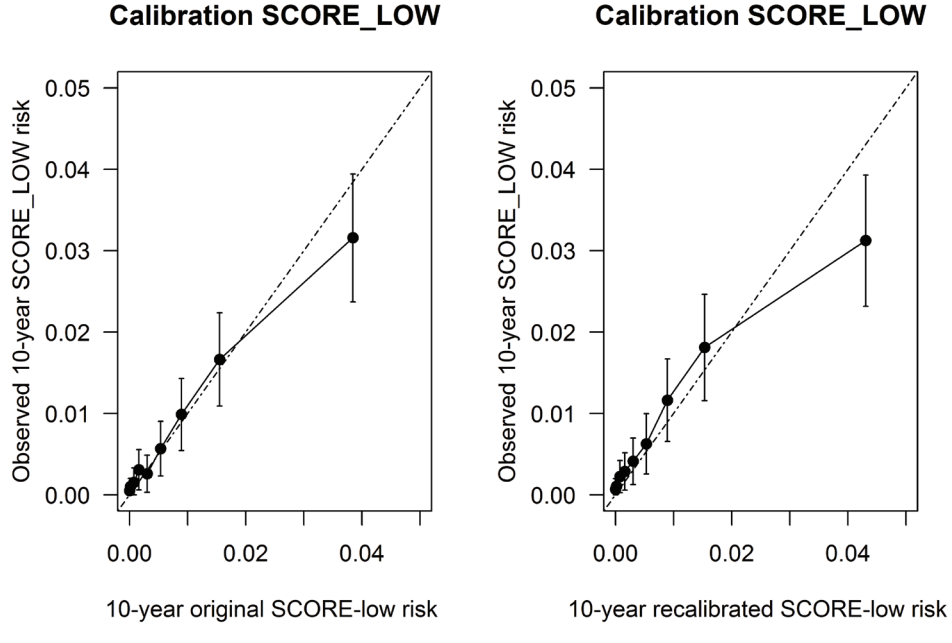
Appendix D - Figure 1: Calibration plot ATP



Appendix D - Figure 2: Calibration plot FRS



Appendix D - Figure 3: Calibration plot PCE



Appendix D - Figure 4: Calibration plot SCORE

E. RECLASSIFICATION TABLES

According to risk prediction model FRS, all individuals were classified into four quartiles based on their relative predicted CVD risk, and reclassified based on ATP, PCE, and SCORE respectively.

Appendix E - Table 1: Reclassification table FRS to ATP in percentage

FRS \ ATP	Low risk (r =0.000)	Moderate low risk (r =0.003)	Moderate high risk (r =0.012)	High risk (r =0.043)
Low risk (r =0.008)	85.22	14.76	0.02	0.00
Moderate low risk (r =0.024)	14.19	70.05	15.64	0.12
Moderate high risk (r =0.054)	0.58	14.86	72.55	12.01
High risk (r =0.150)	0.02	0.33	11.78	87.87

Appendix E - Table 2: Reclassification table FRS to PCE in percentage

FRS \ PCE	Low risk (r =0.002)	Moderate low risk (r =0.007)	Moderate high risk (r =0.019)	High risk (r =0.062)
Low risk (r =0.008)	75.98	24.00	0.02	0.00
Moderate low risk (r =0.024)	15.52	63.13	21.35	0.00
Moderate high risk (r =0.054)	5.48	11.09	71.38	12.05
High risk (r =0.150)	3.02	1.79	7.25	87.95

Appendix E - Table 3: Reclassification table FRS to SCORE in percentage

FRS \ SCORE	Low risk (r =0.000)	Moderate low risk (r =0.001)	Moderate high risk (r =0.004)	High risk (r =0.024)
Low risk (r = 0.008)	82.39	17.16	0.45	0.00
Moderate low risk (r =0.024)	17.29	59.56	21.82	1.33
Moderate high risk (r =0.054)	0.33	22.75	56.99	19.93
High risk (r =0.150)	0.00	0.53	20.74	78.73

F. OVERALL AND INDIVIDUALIZED CVD BURDEN

Appendix F - Table 1: Number of observed events for deciles of age for the MORGEN cohort (n=19,484)

	Age groups	20.1; 26.5	26.6; 32.1	32.2; 36.7	36.8; 40.4	40.5; 43.5	43.6; 47.0	47.1; 50.3	50.4; 53.5	53.6; 57.4	57.5; 73.7
ATP*	CHD Non-Fatal#	1	1	9	11	18	18	32	42	52	48
	CVA Non-Fatal^	-	-	-	-	-	-	-	-	-	-
	CVD Non-Fatal^	-	-	-	-	-	-	-	-	-	-
	CHD Fatal#	0	0	4	2	0	4	6	11	8	15
	CVA Fatal^	-	-	-	-	-	-	-	-	-	-
	CVD Fatal^	-	-	-	-	-	-	-	-	-	-
FRS	CHD Non-Fatal	2	7	17	27	51	60	90	104	157	132
	CVA Non-Fatal	6	3	8	7	10	12	11	25	29	35
	CVD Non-Fatal	6	8	10	12	14	33	25	33	56	70
	CHD Fatal	0	0	6	3	2	5	7	10	8	18
	CVA Fatal	1	1	0	1	1	0	0	2	1	4
	CVD Fatal	0	0	0	0	6	0	1	6	3	2
PCE	CHD Non-Fatal#	1	1	9	11	17	17	32	40	50	45
	CVA Non-Fatal	6	3	8	7	10	12	12	25	30	38
	CVD Non-Fatal^										
	CHD Fatal#	0	0	4	2	0	5	6	10	8	15
	CVA Fatal	1	1	0	1	1	0	0	3	1	4
	CVD Fatal^										
SCORE*	CHD Non-Fatal^										
	CVA Non-Fatal^										
	CVD Non-Fatal^										
	CHD Fatal	0	0	6	3	2	8	13	16	15	28
	CVA Fatal	1	1	0	1	1	0	0	7	3	11
	CVD Fatal	0	0	0	0	6	0	1	8	5	5

** Due to limited number of observed endpoints in some age categories for ATP and SCORE, the overall distribution of CVD events in the entire population was used for the calculation of the consequences and CVD burden.*
^ not included in the composite endpoint
 # only MI included in the composite endpoint

Appendix F - Table 3: Overall and individualized CVD burden according to four CVD risk prediction models, classified to four risk quartiles

		ATP	FRS	PCE	SCORE
Low risk	Overall CVD burden (QALYs lost)	16	490	110	15
	Individualized CVD burden (QALYs lost)	0.00	0.10	0.02	0.00
Moderate low risk	Overall CVD burden (QALYs lost)	120	988	352	119
	Individualized CVD burden (QALYs lost)	0.02	0.20	0.07	0.02
Moderate high risk	Overall CVD burden (QALYs lost)	377	1672	791	453
	Individualized CVD burden (QALYs lost)	0.08	0.34	0.16	0.09
High risk	Overall CVD burden (QALYs lost)	1,137	3,594	2,080	1,911
	Individualized CVD burden (QALYs lost)	0.23	0.74	0.43	0.39
Cohort	Overall CVD burden (QALYs lost)	1,651	6,744	3,333	2,498
	Individualized CVD burden (QALYs lost)	0.08	0.35	0.17	0.13

3.

A new selection method to increase the health benefits of CVD prevention strategies

G.R. Lagerweij
G.A. de Wit
K.G.M. Moons
Y.T. van der Schouw
W.M.M. Verschuren
J.A.N. Dorresteijn
H. Koffijberg
on behalf of the CREW consortium

European Journal of Preventive Cardiology; 2018

ABSTRACT

Background: Cardiovascular disease (CVD) prevention is commonly focused on providing individuals at high predicted CVD risk with preventive medication. Whereas CVD risk increases rapidly with age, current risk-based selection of individuals mainly targets the elderly. However, the lifelong (preventable) consequences of CVD events may be larger in younger individuals. The purpose of this paper is to investigate if health benefits from preventive treatment may increase when the selection strategy is further optimized.

Methods: Data from three Dutch cohorts was combined (n=47469, men:women=1:1.92) and classified into subgroups based on age and gender. Framingham Global Risk Score (FRS) was used to estimate 10-year CVD risk. The associated lifelong burden of CVD events according to this 10-year CVD risk was expressed as Quality-Adjusted Life Years (QALYs) lost. Based on this approach, the additional health benefits from preventive treatment, reducing this 10-year CVD risk, from selecting individuals based on their expected CVD burden rather than their expected CVD risk was estimated. These benefits were expressed as QALYs gained over lifetime.

Results: When using current selection strategy (10% risk threshold), 32% of the individuals was selected for preventive treatment. When the same proportion was selected based on burden, more younger and less older individuals would receive treatment. Across all individuals, the gain in QALYs was 217 between the two strategies, over a 10 year time horizon. Additionally, when combining the strategies 5% extra eligible individuals were selected resulting in a gain of 628 QALYs.

Conclusion: Improvement of the selection approach of individuals can help to further reduce the CVD burden. Selecting individuals for preventive treatment based on their expected CVD burden will provide more younger and less older individuals with treatment, and will reduce the overall CVD burden.

1. INTRODUCTION

One of the leading causes of mortality and morbidity worldwide is cardiovascular disease (CVD) with an expected burden of disease of 143 million disability adjusted-life years in 2020 [1, 2]. An important target of the World Health Organization (WHO) is to reduce the CVD burden with 30% before 2030 [3]. A substantial part of this CVD burden can be prevented by positively influencing behavioural risk factors, e.g. blood pressure, smoking, diabetes and cholesterol, through preventive strategies [4].

In the last decade, CVD risk prediction models have increasingly been used to predict individualized CVD risks. Based on their predicted 10-year CVD risk, several guidelines recommend the use of such prediction models to stratify individuals into risk categories, with a corresponding, recommended preventive treatment strategy. For example, cholesterol lowering drug prescription is advocated in case of a risk above the threshold of 7.5% risk of CVD events in 10 year according to prediction model Pooled Cohort Equations (PCE) or above 10% risk according to Framingham Global Risk Score [4, 5]. Over the years, many CVD risk prediction models have been developed, each with a specific risk threshold for 'high risk', implying that the classification of individuals who qualify for (preventive) treatment is, to say the least, not uniform. As different CVD risk prediction models may use different predictors and different coefficients, different models may classify different individuals as having 'high risk' [6]. However, individuals have different health-related consequences of CVD, that is, loss in terms of life years due to (earlier) death, and health-related quality of life due to non-fatal CVD events. Consequently, the health-related consequences of being classified as 'high risk' may differ per individual, depending on the current quality of life and life expectancy. Additionally, through combining the individual predicted risk of a CVD event with the expected consequences of this CVD event, it is possible to estimate the expected CVD burden of an individual. Aggregating these individual burden estimates can then provide an estimate of the total expected CVD burden in a specific population or group of individuals.

Inferences of the value and usefulness of current risk-based prevention strategies for CVD, however, are solely based on their ability to provide accurate risk estimations and do not include any consideration of the expected health loss due to occurrence of the predicted CVD event(s). As a consequence, the selection of high risk individuals, in whom preventive treatment is initiated, may include individuals at high risk but with a low expected CVD burden,

for example due to high age. Similarly, young individuals with relatively low CVD risk would not receive preventive treatment, even though their expected CVD burden (because they will lose life years due to the consequences of a non-fatal or fatal CVD event) could be substantial. For example, the overall expected burden due to fatal strokes may be higher in young individuals than in older individuals even though the risk of a stroke being fatal increases with age. Indeed, it has been shown that the estimated consequences in terms of health loss of having a CVD event vary widely with age [7, 8]. This implies that selecting individuals for preventive treatment based on predicted risks only, may not necessarily result in the most effective nor the most efficient strategy to reduce CVD burden on population or group level.

In this paper, we investigate if the selection strategy for preventive CVD treatment can be improved by considering a threshold based on expected CVD burden rather than on predicted CVD risk. We illustrate how the selected individuals differ, and how this influences the effectiveness of a hypothetical preventive treatment strategy, in a combination of Dutch cohorts.

2. METHODS

To illustrate if CVD preventive strategies can be improved by considering a burden threshold rather than a risk threshold, we combined different cohort datasets from the Netherlands. This resulted in a heterogeneous large dataset with different age groups and risks.

First, the MORGEN cohort was used [9, 10]. The MORGEN cohort is a subset of the general population from Maastricht, Amsterdam and Doetinchem, including 20.423 males and females with baseline and follow up data [11, 12]. The second cohort (PROSPECT) is a cohort 16.401 females of whom baseline and follow up data are available after linkage [13]. Finally, we used data from 10.645 patients with a history or recent diagnosis of manifest atherosclerotic disease enrolled In the Secondary Manifestations of ARTerial disease (SMART) Study between January 1996 and February 2014 [15]. A reason to include individuals with a CVD history in the analysis was that these individuals were older, had more risk factors and higher occurrence of CVD events. Combined across these three cohorts 47.469 individuals were eligible for the analysis. The men to women ratio was 1:1.92. Baseline information on the individuals per cohort is shown in Appendix A - Table 1.

2.1 Estimating CVD risk and CVD burden

Framingham Global Risk Score (FRS) is a widely used CVD risk prediction model containing easy to measure predictors s, e.g. age, gender and systolic blood pressure (see Appendix A - Table 2) [5, 14]. Prediction model FRS was originally not developed to estimate the risk of a CVD event for individuals with a CVD history. Currently, there are multiple prediction models available for secondary risk estimations, for example the SMART risk score [15, 16]. Therefore, FRS was used to estimate CVD risks of MORGEN and Prospect individuals and SMART risk score was used to estimate CVD risks of SMART individuals. As FRS predicts 10-year risk of CVD events (Appendix A - Table 3, column 1-3), only events occurring within 10 years from baseline (start of cohort) were included in the event distribution and further analysis. Furthermore, for each cohort, the event distribution according to FRS was determined and presented separately for men and women. The observed event distribution was determined per ICD-10 code. The event distribution for men and women is shown per cohort in Appendix A - Table 3.

Conform methodological guidelines in prediction modelling, the FRS prediction model was first recalibrated to each cohort, by updating the baseline hazard and linear predictor to better match each of the three separate cohorts. [17, 18]. Measures of the statistical performance after this recalibration, i.e. discrimination and calibration, were determined per cohort and separately for men and women. The statistical performance of FRS after recalibration is presented in Appendix B. Note that the SMART risk score was originally developed on the SMART cohort, therefore recalibrating was not necessary and only the statistical performance of FRS is presented.

To estimate each individual's expected burden of CVD, the individual's predicted risk was multiplied by the consequences of the events, i.e. multiplying the probability of getting CVD events with the consequences of experiencing CVD events. The consequences of the occurrence of CVD events were estimated and expressed in Quality-Adjusted Life Years (QALYs) lost and determined per individual. The consequences were determined as a product of the observed event distribution, the impact on quality of life (QoL) following different CVD event types (utilities), and average life expectancy, representing years of life lost for fatal CVD events. The observed event distribution was determined separately for cohort, gender, and age groups. The life expectancy of a (healthy) individual only depended on age and

gender and the impact of CVD events on QoL was assumed to be similar for all individuals, i.e. no separate values were used for cohort, gender, or age groups. Appendix C presents more information on the impact of CVD events on QoL and the formula used to estimate CVD burden. Additionally, Appendix C also provides an example on the calculation of CVD burden, per individual.

2.2 Description of the selection process

In this study, we compared four different scenarios of selection of high risk individuals for preventive strategies.

For scenario 1, risk-based selection, we investigated individuals at high absolute risk, according to FRS, with a 10% risk threshold as recommended in the US guideline for CVD preventive strategies [4]. For consistency, we used one single threshold instead of two thresholds for each prediction model.

For scenario 2, burden-based selection, individuals were ranked according to the individual expected CVD burden. Individuals with the highest burden were then selected, until exactly the same number of individuals was selected as when applying scenario 1. The CVD burden of the selected individual with lowest CVD burden was defined as the burden threshold (i.e. all individuals with burden exceeding this threshold were selected). Selecting exactly the same number of individuals in scenarios 1 and 2 allowed comparison of the expected benefits from preventive treatment across these two scenarios.

Scenario 3 combined the selection procedures of scenario 1 and 2. Here, individuals were selected if they had a high predicted risk (scenario 1), a high expected burden (scenario 2) or both. It was expected that the two groups of selected individuals would largely overlap between scenario 1 and 2 because 'high predicted risk' would often lead to 'high expected burden'. However, applying a burden-based selection as in scenario 2 might result in not selecting, and thus withholding treatment from, a small subgroup of individuals with a high CVD risk currently considered for preventive treatment (scenario 1). Scenario 3 thus reflects the notion that withholding relatively cheap and effective preventive medication from individuals with a high CVD risk but with low expected burden may not be desirable.

Given selection on both risk and burden scenario 3 will select a larger number of individuals for preventive treatment than scenario 1 and 2. Therefore, comparison of outcomes between these scenarios is not possible. To assess the impact of combined selection, scenario 4 was defined as an extension of scenario 2, again selecting individuals on burden, but now selecting exactly the same number of individuals as in scenario 3.

In Appendix F, two figures are presented to show more details on the four investigated scenarios, where the marks represent the selected individuals according to the four scenarios.

2.3 Description of hypothetical treatment

For those individuals selected for preventive strategies in any of the 4 scenarios, hypothetical treatment was considered, for example poly pill, blood pressure lowering medication or aspirin. We assumed that all individuals adhered to this medication and that medication would lower the risk of CVD events with 35%, similarly across all types of CVD events included in the composite endpoint [19]. As preventive CVD medication often has side effects, these were included in the analysis (for details see Appendix C - Table 2).

After preventive treatment, the risk reduction on CVD events was applied to each individual. For all four scenarios of selecting individuals, the number of selected individuals, average values of risk and burden, and average values of reduction in CVD burden after treatment were determined. The expected number of events was calculated by summing the estimated CVD risk of the selected individuals. After preventive treatment, the individual risk estimates were multiplied with 35% which resulted in a decrease of the average CVD risk and number of expected CVD events. Furthermore, the gain in QALYs was determined for scenarios 2-4 compared with scenario 1 (reference scenario). Scenarios were also compared among each other.

3. RESULTS

3.1 Estimating CVD risk and CVD burden

Figure 1 (upper part) shows the barplot for the predicted CVD risks according to FRS per age group and gender, with vertical lines representing the 5th and 95th percentile values. The risk estimates vary widely between men and women. As a consequence, men have on average a higher predicted risk compared to women. As expected, there was a trend towards higher CVD risk with increasing age, both for men and women.

The expected CVD burden increased up till the age of 75 years for men and women and decreased thereafter (Figure 1, lower part). Additionally, the burden estimates were higher for men than for women due to two reasons. First, CVD risk also partially determines CVD burden and CVD risks were higher for men. Second, there were differences in the observed event rates, for example, men experienced more often a fatal CVD event than women (Appendix D - Table 2). This was due to the high proportion of men in the SMART cohort. As this subgroup has the highest risk of fatal CVD, in the combined cohort the CVD risk for men exceeds that for women. This effect is therefore not apparent in the other cohorts. The relation between CVD risk and CVD burden is shown in Appendix E - Figure 1.

3.2 Description of the selection process

In the combined cohort, applying the risk based strategy to select individuals resulted in a selection of 32.1% of all individuals, i.e. 15,263 individuals had a predicted risk above the threshold of 10% (scenario 1). However, the percentage of selected men and women was not similar due to different risk estimates, 50.4% and 22.7% of all men and women were considered for the hypothetical treatment, respectively (Table 1, selection part). Additionally, men were selected for treatment at a much earlier age than women, with about 50% of men over 45 years of age qualifying for treatment, while a similar percentage was reached only in women between 65 and 75 years of age.

At the same proportion of selected individuals for scenario 2 (the burden based strategy), the burden threshold was 0.59 QALYs. This indicates that, on individual level, an expected lifelong health loss due to CVD (i.e. CVD risk multiplied with CVD event consequences)

exceeding 0.59 QALYs would make individuals eligible for preventive treatment. Overall, 32.1% of all individuals had an expected CVD burden exceeding 0.59 QALYs and therefore were assigned to preventive treatment in scenario 2.

The total number of the selected individuals was, by definition, similar, the percentage of men and women changed across the selection strategies. In scenario 2, more men, and therefore fewer women, were considered compared to scenario 1 (Table 1, column 2). Although in scenario 2 the total number of selected individuals did not change, individuals were selected based on their estimated CVD burden rather than on their CVD risk. Consequently, certain individuals were selected in scenario 2 who were not selected in scenario 1 since their CVD risk was below the risk threshold of 10%, and vice versa. Appendix G provides more details on the percentage of selected individuals per age group, separately for men and women, in each scenario. Comparing scenario 2 with scenario 1, on average the risk estimates were similar and burden estimates were higher. The difference in burden estimates was mainly caused by the fact that on average younger individuals qualify for treatment in scenario 2, compared to scenario 1. The lower part of Figure 1 shows that, for example, for the age group 35-45 years, the part of the percentile interval line above the threshold (grey dotted line) is largerthan in the upper part of Figure 1. As an example of different implications of scenario 1 and 2, 15.2% and 4.4% of the men and women between the age of 35-45 years were selected for preventive treatment in scenario 1, whereas much more individuals (28.6% and 13.9% of the men and women, respectively) were selected in scenario 2. Additional details on selected individuals and average values of risk and burden per age group and gender are provided in Appendix F.

Following scenario 3 and 4, an additional 2,351 individuals were selected which resulted in a group that included 37.1% instead of 32.1% of all individuals.

3.3 Description of the impact of hypothetical treatment

Treatment following a risk based selection strategy (scenario 1) is estimated to yield 6,474 QALYs, compared to no treatment. Treatment following the burden based selection strategy (scenario 2) is estimated to yield 6,691 QALYs (Table 1). Hence, without treating more individuals, 217 QALYs can be gained from switching to a burden rather than a risk based selection strategy.

When comparing scenario 3 to scenario 1, the expected gain was 628 QALYs. This difference in QALYs was due to the effect of treating the additionally selected individuals hence comparison of the current and new risk based strategy is not directly informative. For a more informative comparison the burden based strategy was extended (scenario 4) by also selecting 37.1% of the individuals as in scenario 3, this was achieved by decreasing the burden threshold with 0.08 QALYs to a threshold of 0.51 QALYs. After hypothetical treatment, the gain from scenario 4 compared to scenario 3 was 41 QALYs (Table 1, column 8)

Overall, the gain was 217 QALYs without treating additional individuals (scenario 2 versus 1) and 669 QALYs when additional individuals were selected based on their CVD burden (scenario 4 versus 1). Hence, scenario 2 is has potential for greater health gain than scenario 1, and, likewise, scenario 4 has potential for greater health gain than scenario 3. As it may not be desirable to withhold preventive treatment from individuals currently eligible for preventive treatment (according to scenario 3), the opportunity loss of not implementing scenario 4, compared to scenario 3, is 41 QALYs.

The analysis on the combined dataset is also performed on the different cohorts separately; the results are presented in appendix H.

4. DISCUSSION

This study illustrates how health benefits from preventive hypothetical treatment increase when the selection of individuals qualifying for preventive intervention changes rather than the intervention itself. Our study results illustrates that the current risk-based selection mainly targets at older individuals, since CVD risk rapidly increases with age. Furthermore, when exactly the same number of individuals was selected based on their CVD burden, both old and young individuals are selected, with the selected young individuals having a low absolute risk but potentially a high health loss with corresponding high expected burden, when a CVD event would occur. As individuals selected on their expected CVD burden have, on average, a higher expected burden than individuals selected on their predicted risk, burden-based selection increases the health benefits of preventive treatment. When both selection strategies are combined, the yield of preventive treatment further increases without the need to withhold preventive treatment in older individuals at (relatively) high risk.

Furthermore, aggregating the individual estimates of CVD burden provides an estimated total CVD burden in a specific population which can also be interpreted as the maximum theoretical health gain achievable by any preventive strategy in this population. When assessing and comparing CVD prevention strategies, the extent to which they would be able to reduce CVD burden on population or group level should be the primary 'effectiveness' outcome of an impact assessment, as it better matches the targets set by the WHO and other organizations [20].

Clinical impact

Many western countries have implemented a risk-based selection strategy to select individuals that should use medication to prevent cardiovascular disease. Following this strategy, many older persons use such medication. This study illustrates how a new selection method may increase the health benefits from CVD preventive treatment. However, in clinical practice the results may vary. A switch from a risk-based to a burden-based strategy implicates earlier intervention with preventive strategies in younger individuals. This raises the question whether prescription of medication is desirable for these individuals since they have to take medication lifelong which may complicate and lower adherence. A lifestyle intervention may be more appropriate for this group [21]. Furthermore, communication is key here since young individuals have a low absolute risk but their relative risk may be high. For example, a 2.6% risk for a 30 year old woman may not seem very threatening, but as this risk falls in the highest risk quintile, it is very high relative to the risk of peers.

Another implication of the switch from risk to burden based strategies is withdrawal of preventive medication for some using or currently starting with preventive medication. This is especially the case in older people with low expected burden. The exact benefits are still under debate, however, as multiple studies investigated the added value of medication, for example statins, in older individuals. Han et al. concluded that there was no benefit in giving pravastatin in primary prevention for all-cause mortality or CHD event among adults 65 years and older [22]. Additionally, Thompson et al. showed that the payoff time of using statins in primary CVD prevention lengthened when the direct treatment disutility of medication increased [23]. In 2012, the American Geriatrics Association recommended that clinicians should balance the benefits and harms of interventions in older individuals [24]. For example, the benefits of most medication are long-term, i.e. decrease CVD risk, while the harms are short-term, e.g. muscle weakness, in elderly people. Balder et al.

showed a large discrepancy between CVD risk guidelines and the current practice of statin prescription in the Netherlands, i.e. a large group of individuals had no discernible cause for statin treatment [25].

Strengths

This study provides a transparent and detailed illustration of different strategies to select individuals for a preventive CVD intervention rather than improving the intervention itself. Risk-based selection compresses the individualized characteristic of the patient into one single number. Moreover, individuals with a similar high predicted CVD risk can be significantly different. Since there is no direct relation between risk and burden estimates, similar risk estimates may result in varying health consequences of the disease and health benefits of preventive treatment. However, burden-based selection accounts for patient age and gender, in addition to predicted risk. Individuals with a similar burden may also be different but the impact of preventive treatment will not vary significantly when expressed as a reduction of this burden. Furthermore, this way of considering individuals for preventive strategies is a move into the direction of more individualized care and increasing the effectiveness of a preventive strategy without (possibly) increasing costs, as similar numbers in a population use preventive medication [26]. The used dataset consists of individuals from a broad age range with relatively young individuals whereas studies on CVD prevention often only included older individuals. However, the power in this young age group is low due to a low number of observed CVD events.

Limitations

As the main focus of this study is to illustrate a proof of concept, our data may not accurately represent the population of the Netherlands. Preferably, we would have used one large cohort consisting of individuals from 20-90 years with a follow-up time of 15 years or longer. Unfortunately, such a cohort is not available in the Netherlands, and probably not even world-wide, therefore we used and combined three existing cohorts. Although, our approach may yield slightly different results in population cohorts from other countries, the analysis itself is generalizable across other populations or can be repeated in a large population cohort whenever data are available.

In this study, only first CVD events are accounted, although in practice individuals may experience more CVD events in a 10 year period. This simplification may lead to an underestimation

of the burden estimates but was necessary since the registration of sequential events in individuals was not always very accurate in the included cohorts. Should we have had more data on follow up and/or the sequence of CVD events within individuals, the expected burden would probably have been higher. Consequently, the absolute differences in QALYs gained between scenarios would have been different but the relative differences would still be marginal. Since this study is an illustration of a proof of concept, we had to make a choice on the treatment effect and possible side effects of the preventive treatment. In our analysis only a single overall effectiveness estimate of event reduction was applied, however the treatment effect of the medication may be different per age group since the observed CVD event rates differ for sex and gender. Furthermore, in practice different treatments (e.g. statins, antihypertensive drugs, or both) will likely be provided to different individuals depending on their risk profile. Along with large variation in individuals' baseline risks, the expected risk reduction of preventive treatment in practice will vary substantially across individuals. The impact of assuming no variation could be assessed, for example, in a patient-level model with separate risk reduction estimates for each individual, or in deterministic sensitivity analyses based on a range of plausible risk reductions. However, this would further complicate the analyses and is therefore mainly of interest when either the average risk reduction in groups of individuals would deviate substantially from 35%, or when absolute outcomes of selection scenarios are appraised rather than differences between scenarios as in our analysis. As our purpose was merely to demonstrate a proof-of-concept and not to evaluate absolute outcomes or real-world implementation of burden-based selection strategies, we chose to not account for individual variation of treatment effectiveness. A concrete application of this burden based selection strategy requires detailed information on effectiveness and consequences of the medication, the risk of the associated side effects and should account for, sequences of CVD events within individuals.

The results show that after hypothetical treatment the burden based selection strategy provides more QALYs prevented compared to the risk based strategy. However, the gain in QALYs was not equally distributed for men and women. The reduction in CVD burden according to the burden strategy was mainly caused by selecting more men and consequently fewer women. In other words, women had to sacrifice health benefits such that men had more health benefits from preventive treatment. Although this is a disadvantage of the burden based strategy, the analyses can easily be performed for men and women separate, for example with a gender specific burden threshold. To investigate this matter

more, we applied both selection strategies on the cohorts separately (Appendix H). Results showed that more women were selected according to the burden strategy compared to the risk strategy. In other words, the switch in selected individuals was in the opposite direction. For the combined dataset, this means that the reduction in CVD burden may be biased due to a cohort effect because Prospect only consists of women. Additionally, the imbalance in gender was already present in the risk based strategy where a larger proportion of men was selected compared to women. The effect of risk selection was enhanced since the health consequences of CVD events, e.g. more severe CVD events, were larger in men. The imbalance in health benefits from preventive treatment for men and women is undesirable, however it is logical since our present goal was to maximize the number of QALYs to be gained from preventive treatment. Furthermore, the unfavourable effects are resolved in the extended selection strategy where currently selected individuals retained their preventive treatment. The additionally selected individuals were mostly men, hence the QALY gain is caused by giving men preventive medication only now without withdrawing women from preventive medication.

Although in our analysis we fixed the burden threshold value to select the exact same number of individuals for preventive treatment as in traditional risk based selection, in practice different burden thresholds can be set, and thresholds could also vary across subgroups of individuals. For example, in a formal health economic analysis the optimal value of the burden threshold for preventive treatment may be determined, even separately for men and women, or depending on co-morbidities of individuals, if such distinctions would be deemed socially and ethically acceptable. Similarly, for consistency we used one single risk threshold since different thresholds may have complicated the analysis and interpretation of the results.

Conclusion

For decades, risk based prevention has been applied to optimize the selection of individuals eligible for preventive interventions, from a perspective in which risk reduction is seen as the ultimate goal. With the increasing emphasis on actual health outcomes of patients, and on the improvements in these health outcomes provided by (preventive) interventions, it is now time to add a burden component to selection strategies. This is straightforward and easily implementable in clinical practice, and can efficiently improve the health benefits from preventive interventions.

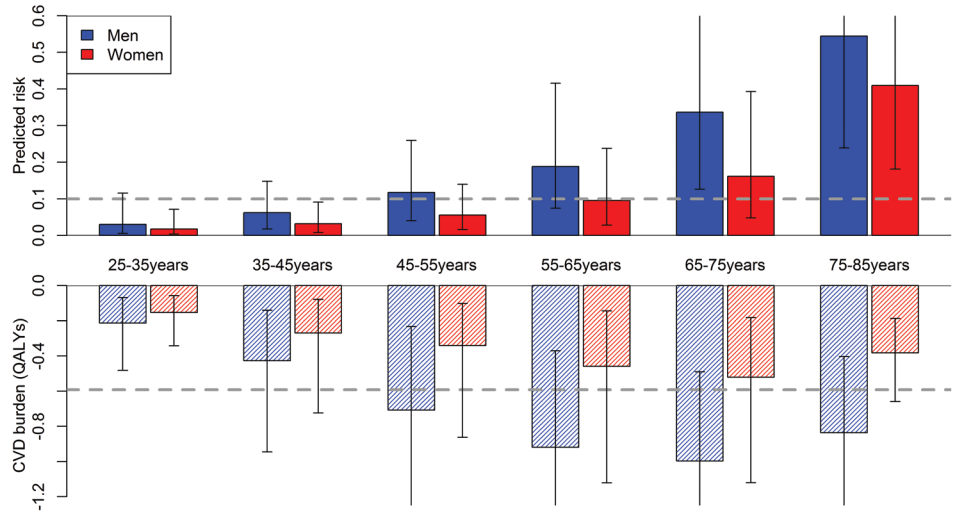


Figure 1: Bar plot of the average values for the predicted CVD risk (upper part) and for the expected CVD burden (lower part), per age group. The vertical lines represent the 5th- and 95th-percentile values of the predicted risks in each group and not the confidence intervals for the expected mean CVD risk estimates. Furthermore, the grey dotted lines represent the threshold with a risk threshold of 10% (upper part) and a burden threshold of 0.59 QALYs (lower part), with the lower part indicating that individuals with an expected lifelong health loss due to CVD (i.e. CVD risk multiplied with CVD event consequences) exceeding 0.59 QALYs would be eligible for preventive treatment. For visual clearance, this figure is presented with some limits value for the estimates of CVD risk and burden.

Table 1: Overall impact of hypothetical preventive treatment when the selected individuals are based on estimates of A) CVD risk (threshold of 10%) or B) CVD burden (threshold of 0.59 QALYs) according to FRS.

A: Scenario 1 - risk based strategy

	Selection				**Impact**		
	Total selected individuals (%)	Average CVD risk	Expected number of events	Estimated CVD burden (QALYs lost)	With preventive treatment (QALYs lost)	Expected number of events	Gain in QALYs
Men	8,182 (50.4%)	0.24	1961	11,320 (1.38)	7,362 (0.90)	1,275	3,958 (0.48)
Women	7,081 (22.7%)	0.19	1334	7,191 (1.02)	4,675 (0.66)	867	2,516 (0.36)
Men and Women	15,263 (32.1%)	0.21	3295	18,511 (1.21)	12,037 (0.79)	2,142	6,474 (0.42)

B: Scenario 2 - burden based selection

	Selection				**Impact**		
	Total selected individuals (%)	Average CVD risk	Expected number of events	Estimated CVD burden (QALYs lost)	With preventive treatment (QALYs lost)	Expected number of events	Gain in QALYs
Men	8,887 (54.7%)	0.23	2003	11,937 (1.34)	7,764 (0.87)	1302	4,174 (0.47)
Women	6,376 (20.4%)	0.19	1202	7,195 (1.13)	4,678 (0.73)	781	2,517 (0.39)
Men and Women	15,263 (32.1%)	0.21	3205	19,133 (1.25)	12,441 (0.82)	2083	6,691 (0.44)

Table 2: Scenario analyses for four different selection strategies

	Selection				**Impact**		**Gain in QALYs compared to scenario 1**
	Total selected individuals		Estimated CVD burden (QALYs lost)		Gain in QALYs		
	N	%	Total	Average	Total	Average	Total
Scenario 1 – Risk based strategy risk ≥ 0.10	15,263	32.1	18,511	1.20	6,474	0.42	-
Scenario 2 – Burden based strategy burden ≥ 0.59 QALYs	15,263	32.1	19,133	1.24	6,691	0.44	217
Scenario 3 – Extended risk based strategy risk ≥ 0.10 or burden ≥ 0.59 QALYs	17,614	37.1	20,310	1.15	7,103	0.40	628
Scenario 4 – Extended burden based strategy burden ≥ 0.51 QALYs	17,614	37.1	20,426	1.16	7,143	0.41	669

Funding

This research forms part of the CREW NHS project, grand number 2013T083, co-funded by the Dutch Heart Foundation. The Dutch Heart Foundation had no role in the collection, analysis, and interpretation of data, nor in the decision to submit the article for publication.

Acknowledgement

The authors gratefully acknowledge the contribution of the SMART research nurses, R. van Petersen (data manager), B. Dinther (vascular manager) and the members of the SMART Study Group: P.A. Doevendans, Department of Cardiology; A. Algra; Y. van der Graaf; D.E. Grobbee; G.E.H.M. Rutten, Julius Center for Health Sciences and Primary Care; L.J. Kappelle, Department of Neurology; T. Leiner, Department of Radiology; G.J. de Borst, Department of Vascular Surgery; F.L.J. Visseren, Department of Vascular Medicine.

Author contributions

AdW, GRL, HK and KGMM contributed to the conception and design of the work. All contributed to the acquisition and interpretation of data for the work, and GRL and HK contributed to the analysis of the data. AdW, GRL, HK, and KGMM drafted the manuscript. All critically revised the manuscript and gave final approval and agreed to be accountable for all aspects of work ensuring integrity and accuracy.

REFERENCES

1. Goff, D.C., Jr., 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, 2014. 63(25 Pt B): p. 2935-59.
2. *CVD atlas Stroke.* Available from: http://www.who.int/cardiovascular_diseases/en/cvd_atlas_15_burden_stroke.pdf?ua=1.
3. Smith, S.C., Jr., et al., *AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update: a guideline from the American Heart Association and American College of Cardiology Foundation endorsed by the World Heart Federation and the Preventive Cardiovascular Nurses Association.* J Am Coll Cardiol, 2011. 58(23): p. 2432-46.
4. 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, 2011. 57(12): p. 1404-23.
5. Goff, D.C., Jr., 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.* Circulation, 2014. 129(25 Suppl 2): p. S49-73.
6. van Giessen, A., et al., *Patient selection for cardiac surgery: Time to consider subgroups within risk categories?* Int J Cardiol, 2016. 203: p. 1103-8.
7. Bots, S.H., S.A.E. Peters, and M. Woodward, *Sex differences in coronary heart disease and stroke mortality: a global assessment of the effect of ageing between 1980 and 2010.* BMJ Glob Health, 2017. 2(2): p. e000298.
8. Leening, M.J., et al., *Sex differences in lifetime risk and first manifestation of cardiovascular disease: prospective population based cohort study.* BMJ, 2014. 349: p. g5992.
9. Beulens, J.W., et al., *Cohort profile: the EPIC-NL study.* Int J Epidemiol, 2010. 39(5): p. 1170-8.
10. Riboli, E., et al., *European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection.* Public Health Nutr, 2002. 5(6B): p. 1113-24.
11. Verschuren, W.M., et al., *Cohort profile: the Doetinchem Cohort Study.* Int J Epidemiol, 2008. 37(6): p. 1236-41.
12. Blokstra, A., Smit, H.A., Bueno de Mesquita H.B., Seidell J.C., Verschuren W.M.M. , *Monitoring of risk factors and health in the Netherlands (MORGEN-cohort), 1993–1997*, 2005, RIVM: Bilthoven, The Netherlands.
13. Boker, L.K., et al., *Prospect-EPIC Utrecht: study design and characteristics of the cohort population. European Prospective Investigation into Cancer and Nutrition.* Eur J Epidemiol, 2001. 17(11): p. 1047-53.
14. D'Agostino, R.B., Sr., et al., *General cardiovascular risk profile for use in primary care: the Framingham Heart Study.* Circulation, 2008. 117(6): p. 743-53.

15. Dorresteijn, J.A., et al., *Development and validation of a prediction rule for recurrent vascular events based on a cohort study of patients with arterial disease: the SMART risk score.* Heart, 2013. 99(12): p. 866-72.

16. Kaasenbrood, L., et al., *Distribution of Estimated 10-Year Risk of Recurrent Vascular Events and Residual Risk in a Secondary Prevention Population.* Circulation, 2016. 134(19): p. 1419-1429.

17. Steyerberg, E.W., et al., *Assessing the performance of prediction models: a framework for traditional and novel measures.* Epidemiology, 2010. 21(1): p. 128-38.

18. FE, H. Regression Modeling Strategies. 2001, New York: Springer.

19. Taylor, F., et al., *Statins for the primary prevention of cardiovascular disease.* Cochrane Database Syst Rev, 2013. 1: p. CD004816.

20. Dehmer, S.P., et al., *Aspirin for the Primary Prevention of Cardiovascular Disease and Colorectal Cancer: A Decision Analysis for the U.S. Preventive Services Task Force.* Ann Intern Med, 2016. 164(12): p. 777-86.

21. Aeschbacher, S., et al., *Healthy lifestyle and heart rate variability in young adults.* Eur J Prev Cardiol, 2016. 23(10): p. 1037-44.

22. Han, B.H., et al., *Effect of Statin Treatment vs Usual Care on Primary Cardiovascular Prevention Among Older Adults: The ALLHAT-LLT Randomized Clinical Trial.* JAMA Intern Med, 2017.

23. Thompson, A., B. Guthrie, and K. Payne, *Using the Payoff Time in Decision-Analytic Models: A Case Study for Using Statins in Primary Prevention.* Med Decis Making, 2017: p. 272989X17700846.

24. Guiding principles for the care of older adults with multimorbidity: an approach for, c., *Guiding principles for the care of older adults with multimorbidity: an approach for clinicians: American Geriatrics Society Expert Panel on the Care of Older Adults with Multimorbidity.* J Am Geriatr Soc, 2012. 60(10): p. E1-E25.

25. Balder, J.W., et al., *Time to improve statin prescription guidelines in low-risk patients?* Eur J Prev Cardiol, 2017. 24(10): p. 1064-1070.

26. Authors/Task Force, M., 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 & Rehabilitation (EACPR).* Eur J Prev Cardiol, 2016. 23(11): p. NP1-NP96.

27. Nederlands Huisartsen Genootschap, *Multidisciplinaire Richtlijn Cardiovasculair Risicomanagement 2011.* Utrecht: Bohn Stafleu van Loghum.

28. *Technology assessment report commissioned by the HTA Programme on behalf of The National Institute for Clinical Excellence*. 2005 [cited 2017; Available from: https://www.nice.org.uk/guidance/ta94/documents/assessment-report-coronary-heart-disease-statins2.

29. Kind, P., et al., *Variations in population health status: results from a United Kingdom national questionnaire survey.* BMJ, 1998. 316(7133): p. 736-41.

30. *Lipid Modification: Cardiovascular Risk Assessment and the Modification of Blood Lipids for the Primary and Secondary Prevention of Cardiovascular Disease*, 2014, NICE: London.

31. Lee, K.K., et al., *Cost-effectiveness of using high-sensitivity C-reactive protein to identify intermediate- and low-cardiovascular-risk individuals for statin therapy.* Circulation, 2010. 122(15): p. 1478-87.

APPENDICES

A. DETAILED INFORMATION ON THE DIFFERENT COHORTS

A.1 – Baseline information

Appendix A - Table 1 shows that individuals from SMART were less healthy and older compared to Morgen and Prospect participants. Moreover, they had significantly higher blood pressure, had diabetes more often, smoked more often and used more medication. On average, the values of the Morgen women regarding their CVD risk factors were slightly lower than the values for the Prospect women, except for smoking. The follow up time across cohorts varied widely because the SMART cohort is an ongoing cohort, whereas Morgen and Prospect are closed cohorts.

Appendix A - Table 1: Baseline table for the different cohorts

	SMART men	SMART women	Morgen men	Morgen women	Prospect women
Number individuals	7,088	3,557	9,153	11,270	16,401
Age (SD)	58.22 (11.61)	55.31 (13.49)	43.32 (11.01)	42.49 (11.25)	57.62 (6.01)
Total cholesterol (SD)	192.65 (51.32)	210.62 (55.82)	205.64 (41.35)	203.94 (40.78)	236.97 (41.38)
High density lipoprotein (SD)	44.88 (12.69)	56.24 (16.59)	45.88 (11.69)	58.32 (14.50)	57.58 (15.62)
Using preventive treatment	44	52%	4%	5%	16%
Systolic blood pressure (SD)	141.26 (20.57)	142.16 (23.49)	125.02 (15.28)	117.89 (16.27)	133.15 (20.02)
Smoking	30%	29%	36%	36%	22%
Diabetes	20%	18%	5%	4%	8%
Follow up time (min-max)	6.99 (0.00-17.49)	7.22 (0.00-17.49)	14.38 (0.04-17.97)	14.93 (0.02-17.97)	14.04 (0.01-17.51)

A.2 - Background information on Framingham Global Risk Score

Appendix A - Table 2: Risk predictors

Risk factor	Description	Unit	FRS
Gender	Gender of respondent	0=male, 1=female	X
Age	Age at baseline (rounded)	Years	X
Total cholesterol (TC)	Baseline level, single measurement	mg/dL	X
High-density lipoprotein cholesterol (HDL-C)	One measurement	mg/dL	X
Low-density lipoprotein cholesterol (LDL-C)	One measurement	mg/dL	
Systolic blood pressure (SBP)	Average SBP of two measurements at baseline	mmHg	X
Treatment (trt)	Use of medication for high blood pressure	0 = no, 1 = yes	X
Smoking (smok)	Current smoking status, former smoker is non-smoker	0 = no, 1 = yes	X
Diabetes (diab)	Diabetes mellitus	0 = no, 1 = yes	X

A.3 - Event distribution regarding different cohorts

Only events occurring within 10 years after recruitment were included in the analysis, because the timeline of the CVD risk prediction models used was 10 years. The event distribution varied hugely between the cohorts, for example, SMART individuals suffered more events resulting into a higher observed incidence rate per individual (Appendix A - Table 3). Moreover, the percentage fatal CVD events was higher for SMART individuals compared to Morgen and Prospect. There were also differences between men and women. Overall and per cohort, it can be seen that men suffered more non-fatal myocardial infarctions whereas women suffered more from strokes.

Appendix A - **Table 3:** Observed event distribution for the different cohorts

	ICD-10 code	FRS	SMART men	SMART women	Morgen men	Morgen women	Prospect women
			N	N	N	N	N
Morbidity							
Myocardial infarction (MI)	I21,I22	X	307 (32.8%)	88 (31.5%)	144 (19.6%)	69 (15.2%)	173 (13.4%)
Other Coronary heart disease (OCHD)	I20,I23,I24,I25	X	0 (0%)	0 (0%)	290 (39.5%)	158 (34.73%)	455 (35.4%)
Cardiac arrest	I46,R96	X	8 (0.9%)	1 (0.4%)	2 (0.3%)	2 (0.4%)	6 (0.5%)
Ischemic stroke (CVAI)	I63,I65	X	240 (25.7%)	97 (34.8%)	64 (8.7%)	60 (13.2%)	195 (15.2%)
Haemorrhagic stroke (CVAH)	I60,I61,I62	X	0 (0%)	0 (0%)	13 (1.8%)	29 (6.4%)	55 (4.3%)
Other stroke (OCVA)	I64,I66	X	0 (0%)	0 (0%)	29 (4.0%)	19 (4.2%)	32 (2.5%)
Other Cardiovascular diseases (OCVD)	G45,I67,I69,I70-I74,I50	X	5 (0.5%)	8 (2.9%)	131 (17.8%)	89 (19.6%)	250 (19.4%)
Mortality							
Myocardial infarction (MI)	I21,I22	X	19 (2.0%)	6 (2.2%)	34 (4.6%)	11 (2.4%)	42 (3.3%)
Other Coronary heart disease (OCHD)	I20,I23,I24	X	0 (0%)	0 (0%)	5 (0.7%)	1 (0.2%)	5 (0.4%)
Cardiac arrest, sudden death	I46,R96	X	138 (14.8%)	34 (12.2%)	10 (1.4%)	1 (0.2%)	19 (1.5%)
Ischemic stroke (CVAI)	I63,I65	X	18 (1.9%)	5 (1.8%)	2 (0.3%)	1 (0.2%)	5 (0.4%)
Haemorrhagic stroke (CVAH)	I60,I61,I62	X	0 (0%)	0 (0%)	2 (0.3%)	4 (0.9%)	9 (0.7%)
Other stroke (OCVA)	I64,I66	X	0 (0%)	0 (0%)	1 (0.1%)	2 (0.4%)	21 (1.6%)
Other Cardiovascular diseases (OCVD)	G45,I67,I69,I70-I74,I50	X	200 (21.3%)	40 (14.3%)	8 (1.1%)	9 (2.0%)	20 (1.6%)
Total number of events			935	279	735	455	1287
Total individuals			7088	3557	9153	11270	16401
Prevalence events (%)			13.19	7.84	8.03	4.04	7.85

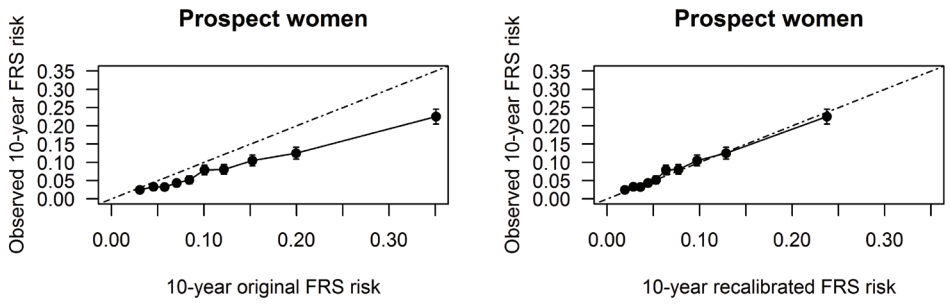
B. STATISTICAL PERFORMANCE

Implementation of a prediction model typically follows updating or recalibration of the model in the target setting, as the target cohort may differ from the original development cohort [28]. Therefore, we recalibrated FRS to the MORGEN and PROSPECT cohorts to ensure that the models provide accurate risk estimates in these cohorts. For the survival data (time-to-event data) considered in this study, recalibrating a prediction model typically involves updating the baseline hazard and centering each predictor around the mean value of all patient characteristics (i.e. linear predictor) in our cohorts, for men and women separately [18]. Furthermore, we incorporated an additional correction factor to ensure that the updated baseline hazards actually reflect the observed probability of survival after 10 years. The regression coefficients of the risk factors of the original FRS model were not changed [15].

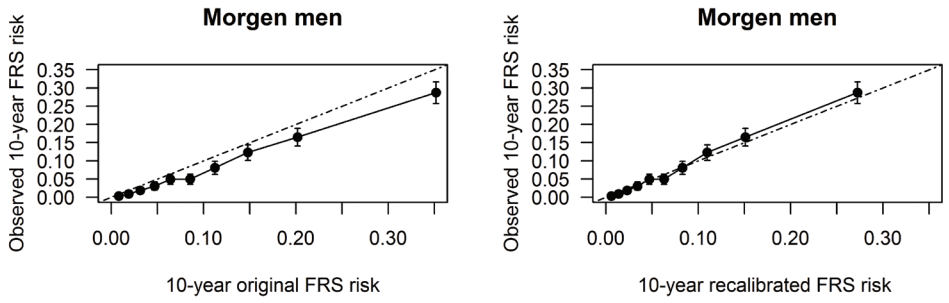
Results of the updated values for the linear predictor and baseline hazard can be seen in Appendix B - Table 1 (column 2-5). Calibration and discrimination results according to the original and recalibrated model can be found in Appendix B - Table 1 (column 6-10). The performance of the subgroups of individuals is good and very similar, see the column "c-statistic". Moreover, the predicted number of events now closely matches the observed number of events. Furthermore, the calibration plots according to FRS for Morgen and Prospect cohort are shown below, see Appendix B - Figure 1 and Appendix B - Figure 2.

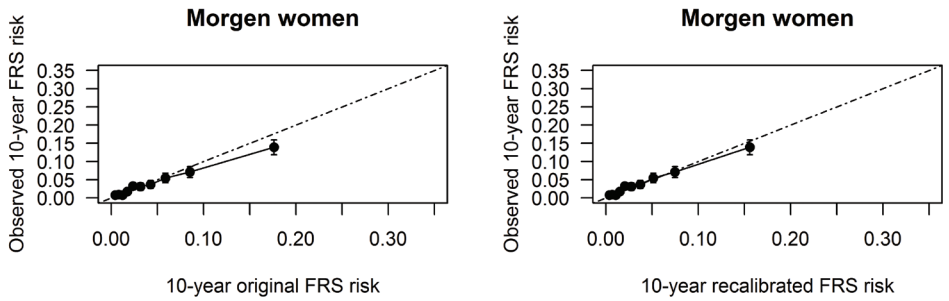
Appendix B - Table 1: Statistical performance of FRS for cohort and gender

	Linear predictor		Baseline hazard		Observed events	Expected events		C-statistic	
	Original model	Recalibrated model	Original model	Recalibrated model		Original model	Recalibrated model	Original model	Recalibrated model
Morgen men	24.35	23.44	0.89	0.92	735	978	735	0.78 (0.75; 0.81)	0.78 (0.75; 0.81)
Morgen women	26.97	25.57	0.95	0.96	455	518	455	0.75 (0.70; 0.79)	0.745 (0.70; 0.79)
Prospect women	26.97	26.87	0.95	0.92	1,286	1,989	1,286	0.70 (0.67; 0.73)	0.70 (0.67; 0.73)



Appendix B - Figure 1: Calibration plot for Prospect individuals





Appendix B - Figure 2: Calibration plot for Morgen individuals

C. INFORMATION ON THE ESTIMATION OF CVD BURDEN

The impact ($I_{i,j}$), in terms of QALY loss, when an event predicted by model *j* occurs in individual i is given by

$$I_{i,j} = \left(\sum_{n=age_i}^{age_i+LE_i} QoL_n \right) \cdot \sum_{k=1}^{\Theta} \left(\pi_{i,k,j} \cdot (1 - u_k) \right)$$

with the number of individual components Θ under consideration, average life expectancy LE_i after surviving an event with remaining quality of life QoL_n, utility u_k and probability $\pi_{i,k,j}$ that an event predicted by prediction model j is of a specific type *k*.

Here, the left component represents the remaining life years – adjusted for their quality – in the absence of CVD events. The right component represents the total expected loss in quality of life due to all predicted CVD events in model *j*. Note that $\pi_{i,k,j}$ is zero for any type of event *k* not included in the composite endpoint of model *j*.

The baseline quality of life of individual i with age n is given by [29, 30]

$$QoL_n = -0.00425 \cdot n + 1.06$$

The formula for the expected CVD burden of disease ($BD_{i,j}$) is given by

$$BD_{i,j} = I_{i,j} \cdot r_{i,j}$$

where $r_{i,j}$ is the recalibrated predicted risk for individual i and prediction model j.

For example, the CVD burden estimate of a 59 year old man was 0.9 QALYs, resulting from a 20.8% 10-year CVD risk and health related consequences equal to losing 4.2 QALYs when CVD events occurred.

Here, the loss of 4.2 QALYs was calculated from a life expectancy of 20.4 years, adjusted for decrease in quality of life with age (first part of the impact equation), multiplied with the expected impact when a CVD event occurred (second part of the equation). This expected impact is the summation of the expected impact of all 14 CVD event types we observed in the data. For each event type, there was a disutility (Appendix C - Table 1) and marginal probability, i.e. the probability that a specific event type occurred given occurrence of CVD. All these 14 marginal probabilities summed up to 1 and were determined by dividing the number of observed events (for that event type) with the total number of observed event (all types). The marginal probabilities were determined per cohort and age group, and separate for gender.

Appendix C - Table 1: Utility values [31]

Morbidity	Utility	Disutility
Myocardial infarction (MI)	0.88	0.12
Other Coronary heart disease (OCHD)	0.88	0.12
Cardiac arrest	0.81	0.19
Ischemic stroke (CVAI)	0.63	0.37
Haemorrhagic stroke (CVAH)	0.63	0.37
Other stroke (OCVA)	0.63	0.37
Other Cardiovascular diseases (OCVD)	0.68	0.32
Mortality		
All CVD event types	0	1

Side effect of the preventive treatment were chosen as known as the side effects of statins, see Appendix C - Table 2 for details.

Appendix C - Table 2: Side effects of preventive treatment [32]

Side effect	Probability	Health loss (utility)
Minor	0.18	2 days of lost life
Major	1/18000	14 days of lost life
Death, given major side effects	0.09	

D. INFORMATION OF THE COMBINED COHORT

Appendix D - Table 1 shows baseline information of the combined cohort, separately for men and women. Although women were older and had higher total cholesterol, men were less healthy regarding the other risk factors. Men more often had diabetes, smoked more often, used more preventive treatment medication and had a slightly higher blood pressure. The event distribution for men and women varied substantially between men and women, as shown in Appendix D - Table 2. The percentage of non-fatal strokes was higher for women whereas men had more myocardial infarctions. Additionally, the percentage of observed fatal events in men was higher than in women.

Appendix D - Table 1: Baseline characteristics of the combined cohort

	Men	Women
Number individuals	16,241	31,228
Age (SD)	49.82 (13.48)	51.9 (11.65)
Total cholesterol (SD)	199.97 (46.42)	222.05 (45.87)
High density lipoprotein (SD)	45.44 (12.14)	57.69 (15.35)
Using preventive treatment	22%	16%
Systolic blood pressure (SD)	132.11 (19.52)	128.67 (21.03)
Smoking	33%	28%
Diabetes	11%	7%
Follow up Time (min-max)	12.07 (0.00-17.97)	13.58 (0.00-17.97)

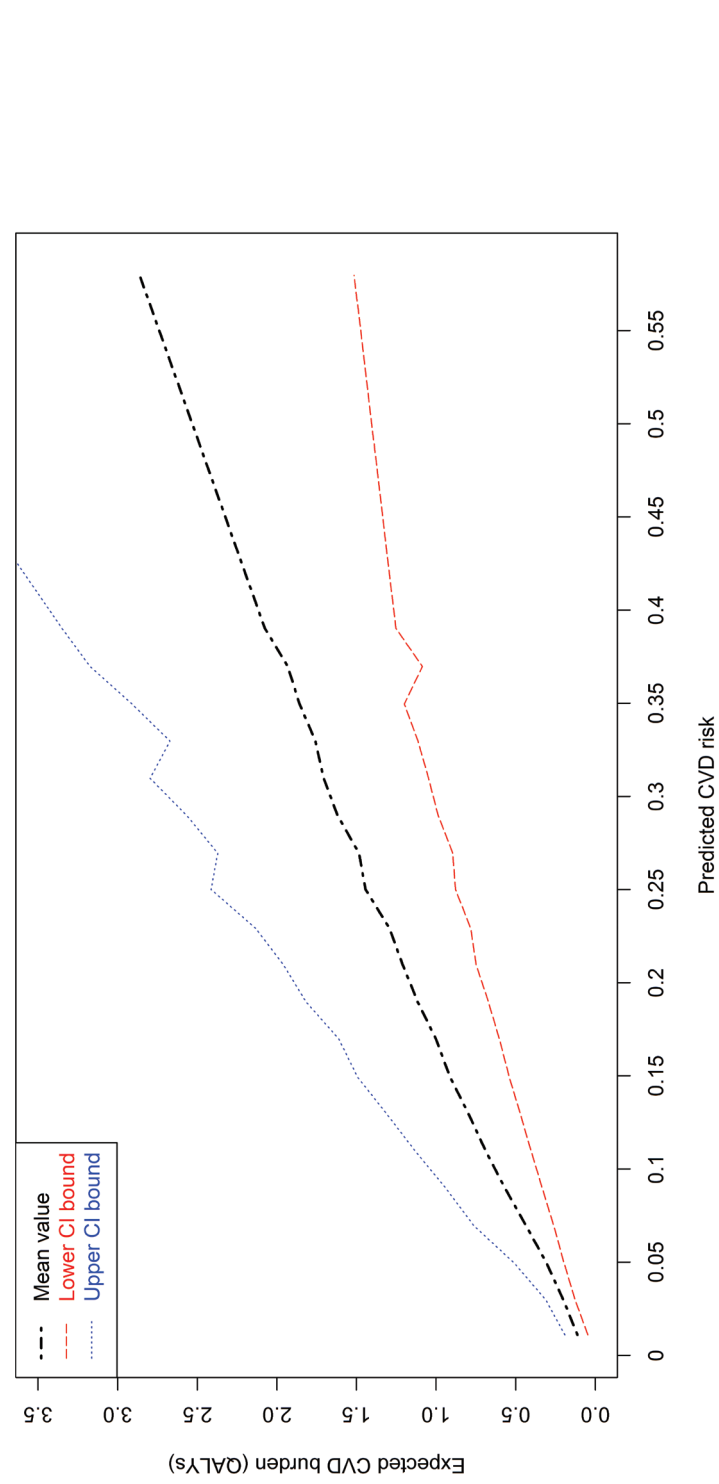
Appendix D - Table 2: Observed event distribution for the combined cohort

	Men (n=16,241)		Women (n=31,228)	
	N	%	N	%
Morbidity				
Myocardial infarction (MI)	451	27	330	16.3
Other Coronary heart disease (OCHD)	290	17.4	613	30.3
Cardiac arrest	10	0.6	9	0.4
Ischemic stroke (CVAI)	304	18.2	352	17.4
Haemorrhagic stroke (CVAH)	13	0.8	84	4.2
Other stroke (OCVA)	29	1.7	51	2.5
Other Cardiovascular diseases (OCVD)	136	8.1	347	17.2
Mortality				
Myocardial infarction (MI)	53	3.2	59	2.9
Other Coronary heart disease (OCHD)	5	0.3	6	0.3
Cardiac arrest, sudden death	148	8.9	54	2.7
Ischemic stroke (CVAI)	20	1.2	11	0.5
Haemorrhagic stroke (CVAH)	2	0.1	13	0.6
Other stroke (OCVA)	1	0.1	23	1.1
Other Cardiovascular diseases (OCVD)	208	12.5	69	3.4
Total number of events‡		1,670		2,021
Percentage fatal events		26.2%		11.6%
Prevalence events		10.3		6.5

‡ Up to 10 years

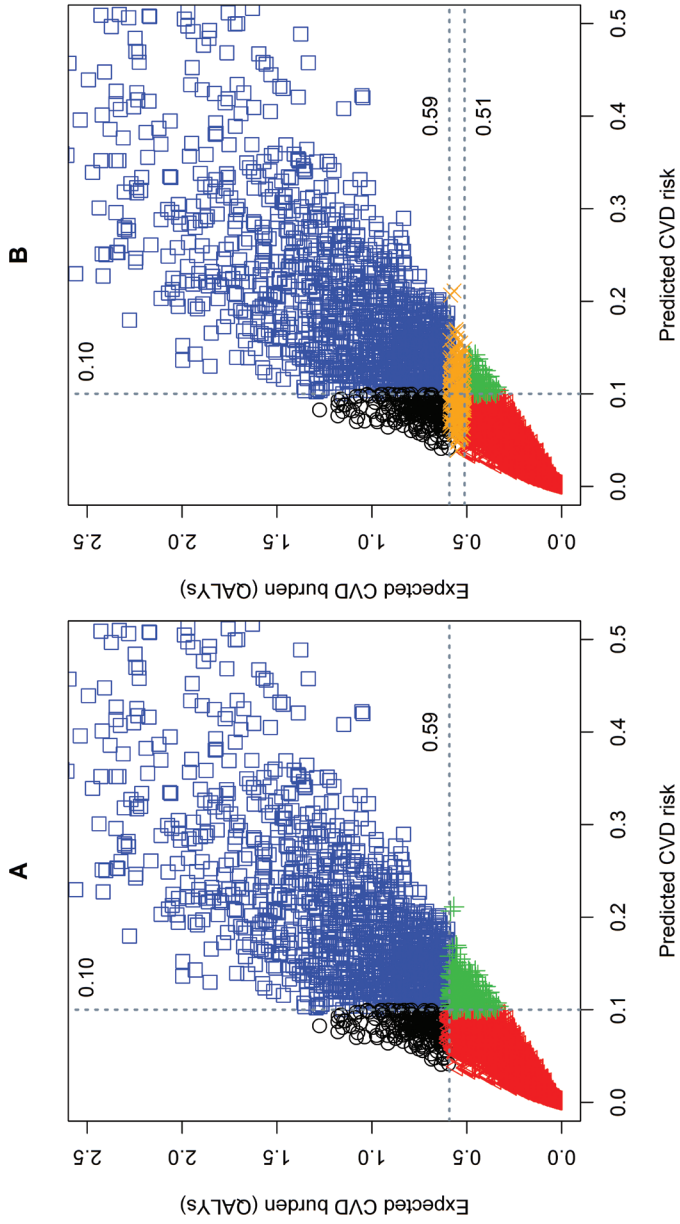
E. CVD RISK AND CVD BURDEN

The variation in CVD burden increases with CVD risk. Individuals with an absolute small CVD risk are similar, where high risk individuals have a large variation in CVD burden due to the health consequences of CVD events.



Appendix E - Figure 1: The average CVD burden estimates per risk group where all predicted risk estimates are divided in groups of 2% and plotted on 1%, 3%, etc. The dotted lines present the 5th- and 95th-percentile values of the burden estimates per risk group risks in each group and not the confidence intervals for the expected mean CVD risk estimates.

F. INVESTIGATED SCENARIOS



Appendix F - Figure 1: Scatterplot of the predicted CVD risk and expected CVD burden. For clarity, only 10% of the individuals were plotted (random sample). Plot A shows the scenario 1 and 2, with individuals selected according to scenario 1 presented by the plus (green +) and squared (blue □) signs, and individuals selected according to scenario 2 presented by the squared (blue □) and circle (black o) signs (# individuals = 15.263). Plot B shows scenario 3 and 4, with individuals considered for scenario 3 presented by the plus (green +), squared (blue □), and circle (black o) signs, and individuals considered for scenario 4 presented by the cross (orange x), circle (black o), and squared (blue □) signs. The grey dotted lines represent the risk threshold of 10%, and a burden threshold of 0.59 and 0.51 QALYs (only plot B).

G. GENDER DIFFERENCES

Results on the selection and impact part, separate for men and women, are shown in Appendix G - Table 1 and Appendix G - Table 2. The numbers in these tables correspond with Figure 1, where columns “average risk” and “average CVD burden” corresponds with the values on the y axis of the bars in Figure 1(upper and lower part). For example, men from the age group “35-45 years” have an average risk and burden of 0.15 and 1.25, based on scenario 1. Moreover, the increase in risk over age, shown in Figure 1, can also been seen in the tables below, together with the difference in predicted risk for men and women. Men have on average a higher predicted CVD risk and burden for all age groups compared to women. Furthermore, Appendix G - Table 1 and Appendix G - Table 2 show that the percentage of selected individuals per age group increases with age, hence, selection of high risk individuals in general means selection of older individuals. Additionally, more older men are selected for preventive treatment compared to women from the same age group, for example, 82% vs 33% for age group “55-65” years of the men and women, respectively.

For scenario 2, i.e. burden based selection, the same number of individuals were selected but the percentage of man versus women and the percentages of selected individuals per age groups changes. There is an increase in the proportion of younger individuals and a decrease in the proportion of older individuals selected. The switch in selected individuals results in higher burden estimates for each age group and thus a health gain of 217 QALYs according to scenario 2 compared to scenario 1.

Results on the selection and impact part, separate for men and women, are shown in Appendix G - Table 1 and Appendix G - Table 2. The numbers in these tables correspond with Figure 1, where columns “average risk” and “average CVD burden” corresponds with the values on the y axis of the bars in Figure 1(upper and lower part). For example, men from the age group “35-45 years” have an average risk and burden of 0.15 and 1.25, based on scenario 1. Moreover, the increase in risk over age, shown in Figure 1, can also been seen in the tables below, together with the difference in predicted risk for men and women. Men have on average a higher predicted CVD risk and burden for all age groups compared to women. Furthermore, Appendix G - Table 1 and Appendix G - Table 2 show that the percentage of selected individuals per age group increases with age, hence, selection of

high risk individuals in general means selection of older individuals. Additionally, more older men are selected for preventive treatment compared to women from the same age group, for example, 82% vs 33% for age group "55-65" years of the men and women, respectively.

For scenario 2, i.e. burden based selection, the same number of individuals were selected but the percentage of man versus women and the percentages of selected individuals per age groups changes. There is an increase in the proportion of younger individuals and a decrease in the proportion of older individuals selected. The switch in selected individuals results in higher burden estimates for each age group and thus a health gain of 217 QALYs according to scenario 2 compared to scenario 1.

Apart from differences in CVD risk, and in the distribution of CVD event types, experienced by men and women, the reduction in health-related quality of life (i.e. disutility) when experiencing a particular CVD event may also be different for men and women. Technically, it is straightforward to account for different utilities of men and women in models such as the one used here. The same goes for gender-specific burden threshold values. However, the societal acceptance of such gender-specific threshold values for treatment requires further investigation of ethical and social considerations.

Appendix G - Table 1: Impact of statin treatment when selecting male individuals based on A) risk threshold of 10% and B) burden threshold of 0.59 QALYs according to FRS.

MEN	Selection			Impact			Selection			Impact	
A: Risk based selection	Total selected individuals (%)	Average CVD risk	Average estimated CVD burden (QALYs loss)	With preventive treatment (QALYs loss)	Gain in QALYs	**B: Burden based selection**	Total selected individuals (%)	Average CVD risk	Average estimated CVD burden (QALYs loss)	With preventive treatment (QALYs loss)	Gain in QALYs
15-25	34 (4.9%)	0.19	1.95	1.27	0.68		51 (7.4%)	0.15	1.57	1.02	0.55
25-35	122 (6.6%)	0.16	1.42	0.92	0.50		207 (11.2%)	0.13	1.13	0.74	0.4
35-45	485 (15.2%)	0.15	1.25	0.81	0.44		910 (28.6%)	0.12	1.00	0.65	0.35
45-55	2,221 (49.2%)	0.17	1.13	0.73	0.39		2,521 (55.8%)	0.16	1.09	0.71	0.38
55-65	3,127 (82.4%)	0.21	1.26	0.82	0.44		3,019 (79.6%)	0.21	1.3	0.85	0.45
65-75	1,777 (99.0%)	0.34	1.78	1.16	0.62		1,762 (98.1%)	0.34	1.8	1.17	0.63
75-85	417 (100%)	0.54	2.07	1.34	0.72		417 (100%)	0.54	2.07	1.34	0.72
Total	*8,182 (50.4%)*	*0.24*	*11.320 (1.38)*	*7.362 (0.90)*	*3.958 (0.48)*		*8,887 (54.7%)*	*0.23*	*11.937 (1.34)*	*7.764 (0.87)*	*4.174 (0.47)*

Appendix G - Table 2: Impact of preventive treatment when selecting female individuals based on A) risk threshold of 10% and B) burden threshold of 0.59 QALYs according to FRS.

WOMEN	Selection			Impact			Selection			Impact	
A: Risk based selection	Total selected individuals (%)	Average CVD risk	Average estimated CVD burden (QALYs loss)	With preventive treatment (QALYs loss)	Gain in QALYs	**B: Burden based selection**	Total selected individuals (%)	Average CVD risk	Average estimated CVD burden (QALYs loss)	With preventive treatment (QALYs loss)	Gain in QALYs
15-25	30 (3.1%)	0.18	2.1	1.37	0.74		65 (6.7%)	0.12	1.46	0.95	0.51
20-35	59 (2.4%)	0.16	1.55	1.01	0.54		163 (6.5%)	0.11	1.05	0.68	0.37
35-45	154 (4.4%)	0.15	1.95	1.27	0.68		491 (13.9%)	0.10	1.22	0.80	0.43
45-55	12,127 (11.3%)	0.16	1.15	0.75	0.40		1,586 (14.7%)	0.14	1.08	0.70	0.38
55-65	3,249 (33.1%)	0.17	0.95	0.62	0.33		2,486 (25.4%)	0.19	1.11	0.72	0.39
65-75	2,188 (64.3%)	0.21	0.92	0.6	0.32		1,414 (41.5%)	0.26	1.17	0.76	0.41
75-85	184 (99.5%)	0.41	1.19	0.77	0.42		171 (92.4%)	0.43	1.24	0.81	0.43
Total	*7,081 (22.7%)*	*0.19*	*7,191 (1.02)*	*4,675 (0.66)*	*2,516 (0.36)*		*6,376 (20.4%)*	*0.19*	*7,195 (1.13)*	*4,678 (0.73)*	*2,517 (0.39)*

H. DIFFERENCES IN COHORTS

Results on the selection and impact, separate for cohorts, are shown in Appendix H - Table 1, Appendix H - Table 2, and Appendix H - Table 3. The different combined cohorts have a large variation in risk estimates hence we used a relative risk threshold of 10% rather than an absolute risk threshold. In other words, individuals with the highest 10% risk and highest 10% burden estimates were selected and compared among each other.

Across all cohorts, the selection of high risk individuals, i.e. scenario 1, in general means selection of older individuals. Additionally, more men are selected for preventive treatment compared to women. For scenario 2, i.e. relative burden threshold, exactly the same number of individuals was selected but the percentage of man versus women and the average age of the selected individuals changes. There is an increase in the proportion of younger individuals and women selected resulting in higher burden estimates for each cohort and thus a health gain of 109 (8.6%), 20 (2.6%), and 15 (2.4%) QALYs according to scenario 2 compared to scenario 1, according to the SMART, Morgen, and Prospect cohort.

Appendix H - Table 1: Impact of preventive treatment when selecting individuals from SMART cohort based on A) relative risk threshold of 10% and B) relative burden threshold of according to SMART risk score.

A: Scenario 1 - risk based strategy

	Selection					**Impact**		
	Total selected individuals (%)	Average age (SD)	Average CVD risk	Expected number of events	Estimated CVD burden (QALYs lost)	With preventive treatment (QALYs lost)	Expected number of events	Gain in QALYs
Men	868 (81.5%)	70.6 (7.1)	0.63	545	2,698 (3.1)	1,754 (2.0)	354	944 (1.1)
Women	197 (18.5%)	70.1 (9.1)	0.61	120	598 (3.0)	389 (2.0)	79	209 (1.1)
Men and Women	1,065 (100%)	70.5 (7.5)	0.62	666	3,296 (3.1)	2,143 (2.0)	433	1,153 (1.1)

B: Scenario 2 - burden based selection

	Selection					**Impact**		
	Total selected individuals (%)	Average age (SD)	Average CVD risk	Expected number of events	Estimated CVD burden (QALYs lost)	With preventive treatment (QALYs lost)	Expected number of events	Gain in QALYs
Men	841 (79.0%)	65.0 (10.2)	0.59	497	2,810 (3.3)	1,827 (2.1)	323	983 (1.2)
Women	224 (21.0%)	57.7 (12.1)	0.48	107	796 (3.6)	518 (2.3)	69	279 (1.2)
Men and Women	1,065 (100%)	63.5 (11.0)	0.53	603	3,606 (3.4)	2,345 (2.2)	392	1,261 (1.2)
Total difference strategies								**109 (8.6%)**

Appendix H - Table 2: Impact of preventive treatment when selecting individuals from Morgen cohort based on A) relative risk threshold and B) relative burden threshold according to FRS.

A: Scenario 1 - risk based strategy

	Selection					Impact		
	Total selected individuals	Average age (SD)	Average CVD risk	Expected number of events	Estimated CVD burden (QALYs lost)	With preventive treatment (QALYs lost)	Expected number of events	Gain in QALYs
Men	1,558 (76.2%)	55.1 (5.3)	0.23	351	1,656 (1.1)	1,077 (0.7)	228	580 (0.4)
Women	485 (23.7%)	56.8 (4.7)	0.21	102	500 (1.0)	325 (0.7)	67	175 (0.4)
Men and Women	2,043 (100%)	55.5 (5.2)	0.22	454	2,156 (1.1)	1,402 (0.7)	295	755 (0.4)

B: Scenario 2 - burden based selection

	Selection					Impact		
	Total selected individuals	Average age (SD)	Average CVD risk	Expected number of events	Estimated CVD burden (QALYs lost)	With preventive treatment (QALYs lost)	Expected number of events	Gain in QALYs
Men	1,543 (75.5%)	52.6 (6.4)	0.22	340	1,687 (1.1)	1,096 (0.7)	221	590 (0.4)
Women	500 (24.5%)	54.2 (6.2)	0.20	101	528 (1.1)	343 (0.7)	66	185 (0.4)
Men and Women	2,043 (100%)	53.0 (6.3)	0.21	441	2,215 (1.1)	1,496 (0.7)	287	775 (0.4)
Total difference								**20 (2.6%)**

Appendix H - Table 3: Impact of preventive treatment when selecting individuals from Prospect cohort based on A) relative risk threshold and B) relative burden threshold according to FRS.

A: Scenario 1 - risk based strategy

	Selection					Impact		
	Total selected individuals	Average age (SD)	Average CVD risk	Expected number of events	Estimated CVD burden (QALYs lost)	With preventive treatment (QALYs lost)	Expected number of events	Gain in QALYs
Women	1,641 (100%)	62.9 (5.2)	0.24	390	1,712 (1.0)	1,112 (0.68)	253	599 (0.37)

B: Scenario 2 - burden based selection

	Selection					Impact		
	Total selected individuals	Average age (SD)	Average CVD risk	Expected number of events	Estimated CVD burden (QALYs lost)	With preventive treatment (QALYs lost)	Expected number of events	Gain in QALYs
Women	1,641 (100%)	60.6 (5.7)	0.23	380	1,753 (1.1)	1,139 (0.69)	247	614 (0.37)
Total difference								**15 (2.4%)**

4.

Interpretation of CVD risk predictions in clinical practice: mission impossible?

G.R. Lagerweij
K.G.M. Moons
G.A. de Wit
H. Koffijberg

PLOS ONE; 2019

ABSTRACT

Background: Cardiovascular disease (CVD) risk prediction models are often used to identify individuals at high risk of CVD events. Providing preventive treatment to these individuals may then reduce the CVD burden at population level. However, different prediction models may predict different (sets of) CVD outcomes which may lead to variation in selection of high risk individuals. Here, it is investigated if the use of different prediction models may actually lead to different treatment recommendations in clinical practice.

Methods: The exact definition of and the event types included in the predicted outcomes of four widely used CVD risk prediction models (ATP-III, Framingham (FRS), Pooled Cohort Equations (PCE) and SCORE) was determined according to ICD-10 codes. The models were applied to a Dutch population cohort (n=18,137) to predict the 10-year CVD risks. Finally, treatment recommendations, based on predicted risks and the treatment threshold associated with each model, were investigated and compared across models.

Results: Due to the different definitions of predicted outcomes, the predicted risks varied widely, with an average 10-year CVD risk of 1.2% (ATP), 5.2% (FRS), 1.9% (PCE), and 0.7% (SCORE). Given the variation in predicted risks and recommended treatment thresholds, preventive drugs would be prescribed for 0.2%, 14.9%, 4.4%, and 2.0% of all individuals when using ATP, FRS, PCE and SCORE, respectively.

Conclusion: Widely used CVD prediction models vary substantially regarding their outcomes and associated absolute risk estimates. Consequently, absolute predicted 10-year risks from different prediction models cannot be compared directly. Furthermore, treatment decisions often depend on which prediction model is applied and its recommended risk threshold, introducing unwanted practice variation into risk-based preventive strategies for CVD.

1. INTRODUCTION

Reduction of cardiovascular disease (CVD) burden, i.e. at population level, is commonly accomplished using preventive strategies (like lifestyle and dietary advice or pre-emptive drug treatment) in individuals with marked elevations in risk factors, e.g. low-density lipoprotein (LDL), or a high predicted CVD risk based on a combination of risk factors [1]. Identification of high risk individuals is often achieved using CVD risk prediction models of which over 360 different variants have been published as of 2016 [2]. However, different models may predict multiple and often different CVD outcomes or sets of outcomes (as is the case in model with composite endpoints) [2-4]. These differences in predicted outcomes may result in large variation in CVD risk estimates. Consequently, it is unclear to what extent the predicted CVD risks obtained from different prediction models are comparable and can be interpreted similarly in clinical practice [4-7].

The large variation in CVD risk estimates combined with different recommended risk thresholds for each prediction model, may lead to different definitions of high-risk individuals. For example, the Pooled Cohort Equation stratifies individuals with a > 7.5% 10-year CVD risk as high-risk whereas the recommended threshold for the Framingham risk equation is 10% [8, 9]. Different definitions of high-risk individuals may, in turn, lead to different treatment recommendations. Furthermore, the expected health benefits of treatment may also be different since the impact on quality of life differs per CVD event type and severity. For example, the expected health loss due to a stroke is expected to be higher than the health loss due to a myocardial infarction [10].

Since the implication of different treatment recommendations could be large, the aim of this paper is to assess if the use of different prediction models leads to different treatment recommendations in clinical practice. Therefore, four widely used CVD risk prediction models were investigated regarding their comparability and interpretation, after applying them to a large population cohort. Additionally, we discuss the usefulness of such models based on the comprehensiveness of their composite endpoint and provide a recommendation for the development of new prediction models in order to enhance their usefulness in clinical practice. This paper does not focus specifically on Dutch clinical practice and does not provide guidance on preferred prediction models for the Dutch context.

2. METHODS

Adult Treatment Panel III (ATP), Framingham Global Risk Score (FRS), Pooled Cohort Equations (PCE), and SCORE-low (SCORE) are four widely used CVD risk prediction models for primary prevention [11-14]. All are derived from general population cohort data. Hence, they include (often similar) predictors that are easy to measure in everyday clinical practice, such as gender, age and systolic blood pressure. The exact definition of the included risk factors in the risk equation can be found in the original publication [11-14]. Furthermore, the probability estimate of each model reflects the absolute risk of the composite endpoint occurring within 10 years. In order to compare these four models, we first identified the exact definition of each composite endpoint from the original publication describing the development of the model [11-14]. We then, standardized the composite endpoints using ICD-10 codes. This was necessary since the published articles often only described the outcomes in words, e.g. “coronary heart disease” or “ischemic heart disease”.

To compare the composite endpoints, we used the MORGEN cohort. The MORGEN cohort is a large Dutch general population cohort which includes men and women aged 20 to 74 years at baseline, recruited from the general population between 1993 and 1997 [15]. Participant information on vital status, cause of death and comorbidity was obtained from Statistics Netherlands and the National Medical Registry (NMR). The follow-up period of the MORGEN cohort was 10 to 15 years with a mean follow-up time of 12 years. To apply the prediction models, information both from baseline and from follow-up was required, leaving 19,484 (72%) individuals with adequate data for the analysis from the original cohort [16, 17]. To further investigate the constitution of the composite endpoint, we determined the observed rates and distributions of the individual components, i.e. included CVD event type according to the associated ICD-10 code(s), for each model *separately*.

As the indication for statin therapy is also LDL-dependent and we aim to illustrate the complexity of CVD risk predictions by comparing results of different prediction models, individuals with an elevated level of LDL and/or diabetes were excluded for further analysis. We focused on individuals in whom preventive intervention was indicated based on predicted CVD risk rather than on elevated LDL levels and/or diabetes. After excluding 231 individuals with diabetes, 1,141 individuals with elevated LDL levels and 25 individuals with both risk factors at baseline, this resulted in a cohort size of 18,137 individuals (mean age is 42.4

years, range 20.1-73.7 years, and 45% men).The MORGEN cohort was also used to compare the predicted CVD risks by estimating every individual’s 10-year CVD risk with each of the four prediction models. Implementation of a prediction model typically follows updating or recalibration of the model in the target setting, as the target cohort may differ from the original development cohort [18]. Therefore, we first recalibrated the four prediction models using the MORGEN cohort to ensure that the models provide accurate risk estimates in this cohort. For the survival data (time-to-event data) considered in this study, recalibrating a prediction model typically involves updating the baseline hazard and centering each predictor around the mean value of all patient characteristics in our empirical cohort, correcting for men and women separately [19, 20]. Furthermore, we incorporated an additional correction factor to ensure that the updated baseline hazards actually reflect the observed probability of survival after 10 years.

Many clinical guidelines advocate the use of prediction models to select individuals with a predicted risk above a certain threshold for pre-emptive lipid or blood pressure lowering drug treatment. Different recommendations for absolute 10-year risk thresholds were identified for each model: 10% (ATP), 10% (FRS), 7.5% (PCE), and 5% (SCORE) [9,12,21]. By doing this, we were able to further explore and compare the varying treatment decisions according to the four models. Finally, we first assigned individuals to treatment based on their FRS risk and the FRS risk threshold and then reassigned individuals according to their ATP, PCE, and SCORE risks, and the corresponding thresholds.

The aim of this paper was to illustrate the complexity of comparing predicted risks. This paper does not focus specifically on Dutch clinical practice and does not provide guidance on preferred prediction models for the Dutch context.

3. RESULTS

Although the predictors of the four prediction models are similar, the composite endpoints vary widely and include different CVD event types (Table 1, column 1-6). Myocardial infarction (MI) is included in all four composite endpoints, either alone or in combination with other CVD event types. The endpoint defined for FRS includes the largest range of fatal and non-fatal CVD event types, whereas the endpoint defined for SCORE only includes fatal event types.

Table 1 (column 4, 6, 8, and 10) shows the incidence of each CVD event type as observed in the MORGEN cohort for the four different prediction models. Due to different composite endpoints, individuals with an earlier CVD event which was not included in the considered endpoint were not censored. Therefore, the observed event rates for a specific CVD event vary per prediction model. Definition of a first and secondary event within individuals thus depends on whether the observed CVD events for individuals are included in the composite endpoint of each prediction model. Due to the different definitions of the composite endpoint of the four prediction models, the total number of observed events for SCORE (n=105) is almost nine times smaller than for FRS (n=928). These differences in composite endpoints also affect the absolute number of observed events per prediction model, due to different censoring mechanisms. For example, the absolute number of fatal MIs varies per prediction model because secondary fatal MIs may be censored due to occurrence of another primary event present in the composite endpoint. To illustrate: a fatal MI following a non-fatal stroke event would be accounted for (*not* censored) in the SCORE and ATP model and not accounted for (censored) in the FRS and PCE model. The relative incidence of CVD event types within composite endpoints also varies substantially. For example, of the 105 events observed according to SCORE, 48 (46%) were fatal MIs, whereas the relative incidence of fatal MIs is 17%, 4%, and 11%, for ATP, FRS, and PCE, respectively. This means that the burden, or health loss, associated with the incidence of each composite endpoint varies with a) the included event types, and b) the relative incidence of these event types.

The performance of the recalibrated prediction models was good and quite similar; the c-index was 0.81, 0.78, 0.78, and 0.81 for ATP, FRS, PCE, and SCORE respectively. Appendix A - Table 1 shows an overview of the observed and predicted number of CVD events for each of the four models. Following from this table, it is apparent that the events are well captured by the models and that the number of predicted events closely matches the observed number of events.

Figure 1 shows that the dissimilarities in composite endpoints lead to substantial variation in predicted 10-year CVD risks. Since predicted risks increase with the inclusion of more CVD event types in the composite endpoint, to an extent that depends on their absolute incidence. For example, in our cohort the broad composite endpoint used in the FRS model, covering a large range of CVD event types, yields higher risk predictions than the ATP, PCE, and SCORE models. Similarly, the narrow composite endpoint of SCORE (only fatal events), and its inherent low incidence of included event types yields the lowest risk predictions of all models considered. The average predicted risks in the MORGEN cohort are 1.2% (ATP), 5.1% (FRS), 1.9% (PCE), and 0.6% (SCORE). Hence, the differences in composite endpoints between prediction models (shown in Table 1) result in large variation in predicted CVD risks across prediction models.

Considering that the largest set of CVD event types was included by the FRS composite endpoint, we compared FRS risk estimates with risk estimates from the other three models using more narrow composite endpoints. Figure 1 shows the comparison in CVD risks and reveals an association between these risk estimates. This association indicates that, *in this cohort*, individuals who have the highest risk according to FRS typically also have the highest risk according to ATP, PCE, and SCORE. However, while the *relative* risks are similar the *absolute* risks are clearly different. Furthermore, the vertical spread of points in Figure 1 shows how individuals with a certain FRS risk estimate may have varying risk estimates according to the other models, due to the effect of different risk factor combinations in each model. For example, the group of individuals with an average predicted FRS risk of 10% had an average PCE risk of 3.9%, with a 95% percentile range of 2.2%-5.1% (Figure 1, plot B).

Given the variation in composite endpoints and the subsequent variations in risk predictions from the four models, selecting high risk individuals based on the corresponding recommended risk thresholds results in highly different high risk groups, identified per model. Unfortunately, the fact that each prediction model has its own associated risk threshold further complicates the interpretation and comparison of absolute predicted risks between prediction models. Consequently, treatment decisions may vary with the prediction model that is used. For example, in the MORGEN cohort these thresholds would possibly lead to a seventy-fold difference in prescription of preventive drug treatment in 0.2%, 14.4%, 4.3%, and 1.4% of all individuals, when using ATP, FRS, PCE and SCORE, respectively. To illustrate the implications of these differences, we determined the CVD risks and the consequences on

treatment decisions according to the four prediction models for one individual in our cohort. Indeed, using FRS for this individual implies both a greater necessity to consider preventive drug treatment and a larger potential benefit of such treatment, compared to ATP, PCE, and SCORE (see Clinical example).

The treatment decisions based on the four risk thresholds are shown in Table 2. We found that the treatment decisions based on the different models vary widely, which is undesirable from a public health point of view. When using FRS, 2618 individuals have an estimated risk exceeding the FRS threshold and would thus be eligible for medical treatment. Of these individuals, only 32 (1.2%), 725 (27.7%), and 56 (2.1%) individuals would be considered eligible for medical treatment using the estimated risk and corresponding threshold when applying ATP, PCE, and SCORE respectively.

These different decisions may be due to either the different estimated risks or due to the use of different risk thresholds for classifying individuals as high risk and thus eligible for medical treatment. In our cohort, we observed that mostly the same individuals were assigned a *relatively* high risk according to each of the four prediction models (Figure 1). For example, of the individuals with the highest 20% predicted risks according to FRS (n=3621), 3106 (85.8%), 3131 (86.5%), and 861 (23.8%) of individuals were also classified as relatively high risk (top 20%) according to ATP, PCE, and SCORE, respectively. This relatively high risk group had an average CVD risk of 14.2% according to FRS, and average risks of 3.9%, 5.6%, and 0.7% according to ATP, PCE, and SCORE, respectively. None of the individuals within the top 20% risk group according to FRS had a relatively low risk (bottom 20%) according to the other models. Hence, the expected differences in treatment decisions across prediction models is mainly due to the different corresponding treatment thresholds, and their relation to predicted risks, and not due to the different classification of individuals.

4. DISCUSSION

CVD risk prediction is key in providing preventive medication to large groups of individuals at intermediate or high risk of future CVD events, despite absence of specific elevated risk factors. Although PCE is often used, contemporary decision making and CVD management in the US, FRS is also applied, for example to guide pharmacotherapy for LDL-C lowering in women [9].

This paper illustrates the complexities of interpreting and comparing predicted 10-year CVD risks from four widely used CVD risk prediction models. We showed that the models vary substantially regarding their composite endpoints, and therefore also regarding their predicted absolute risks. As a result, absolute predicted 10-year risks from different prediction models cannot be compared directly and treatment decisions depend on the applied prediction model and its associated risk threshold. For example, of the high-risk individuals considered for preventive treatment according to FRS, only 1%, 28%, and 2% were eligible according to ATP, PCE, and SCORE, respectively (Table 2). Hence, the choice for a specific prediction model is very likely to impact treatment decisions in a large group of assessed individuals. Fortunately, the variation in relative predicted CVD risks is limited, implying that these prediction models rank individuals similarly regarding their CVD risk.

Consequences of difference in composite endpoints on clinical utility

Previous research has indicated that the use of composite endpoints instead of single endpoints *in clinical trials* may have benefits, e.g. improved power or wider coverage of the disease [22]. However, the overall usefulness of composite endpoints in clinical trials is still debated due to the difficulty of interpreting differences in 'set of outcomes' [22, 23]. The interpretation of the associated consequences of predicted CVD risks is also directly affected by the different composite endpoints. For example, communicating to a patient that he/she has a 10-year CVD risk of 3% according to SCORE, compared to a 10-year CVD risk of 6% according to FRS, may affect understanding and adherence of patients to any recommended preventive treatment. A 3% SCORE risk could indicate that the patient is part of the group with the 20% highest absolute risk according to SCORE whereas the patient could be part of the group with the 20% lowest predicted absolute risk with a 6% risk according to FRS (Figure 1).

In addition, the expected health loss due to events predicted by SCORE is expected to be higher than the health burden or health loss due to events predicted by FRS due to how all

included events in SCORE are fatal, but can fatal or non-fatal in FRS. This issue also affects the evaluation of benefits from preventive interventions. For example, when preventive statin treatment is assumed to reduce the risk of a "composite" endpoint with a certain fraction (relative risk < 1), estimates of the corresponding health benefits will be highly dependent on the (constitution of) the composite endpoint of the prediction model used [24].

Even for a single prediction model, the impact of experiencing a predicted composite event is likely to depend on age, since a) the proportion of fatal events increases with age, and b) the actual health loss due to CVD events decreases with age (i.e. with decreasing life expectancy). Hence, even if the distribution of events included in a composite endpoint is known, the expected health impact of a specific risk estimate, for example a 10-year FRS risk of 8%, and therefore the potential benefits of preventive intervention, may differ between groups of individuals [25].

Given the adequate performance of the CVD prediction models considered, and roughly similar relative risk classification, it is recommended that models are applied that have a broad rather than narrow composite endpoint, i.e. models covering a large range of CVD event types. For example, ATP and SCORE may be less useful in this context than FRS and PCE, as the latter cover more manifestations of the underlying cardiovascular disease process. This results in higher predicted risks, which may then be communicated as the 'total risk' of any (type of) CVD event to the patient, to facilitate understanding and improve adherence to preventive medication [26]. However, understanding the "total (high) risk" is only an aspect of adherence and should not replace informed choice and shared decision making.

Implications for development of new prediction models

Regarding prediction model development and research, it is recommended that any newly developed clinically relevant risk prediction model also use a broad composite endpoint, with each included event type uniquely defined, e.g. using ICD-10 codes. A clear definition of a) the composite endpoint and b) the observed incidence of each event type in the development cohort is critical to enable correct interpretation of the predicted risks. This will allow for more transparent and direct comparison of predicted risks and statistical performance of prediction models as well as more standardized evaluations of the health impact of risk-based preventive interventions.

Conclusion

Current CVD risk prediction models vary widely in predicted outcomes, which directly impact their usefulness in clinical practice. Furthermore, this renders estimates of the population burden of CVD, and of the impact of risk-based CVD intervention strategies that highly depend on the prediction model used. Physicians, patients and health policy makers may benefit from a broader and more standardized method of defining outcomes and classification thresholds in prediction model studies.

Table 1: Constitution of the composite endpoints according to ATP, FRS, PCE, and SCORE and incidence of CVD events in MORGEN cohort

		ATP		FRS		PCE		SCORE	
Individual components	**ICD-10 code**		**#**		**#**		**#**		**#**
Morbidity									
Myocardial infarction (MI)*	I21,I22	X	183	X	164	X	176		
Other Coronary heart disease (OCHD)	I20,I23,I24,I25			X	348				
Cardiac arrest, sudden death	I46,R96			X	3				
Haemorrhagic stroke (CVAH)	I60,I61,I62			X	39	X	39		
Ischemic stroke (CVAI)	I63,I65			X	56	X	58		
Other stroke (OCVA)	I64,I66			X	29	X	29		
Other Cardiovascular diseases (OCVD)	G45,I67,I69,I70-I74,I50			X	222				
Total observed events			***183***		***861***		***302***		***0***
Mortality									
Myocardial infarction (MI)	I21,I22	X	38	X	33	X	38	X	48
Other Coronary heart disease (OCHD)	I20,I23,I24			X	3			X	12
Cardiac arrest, sudden death	I46,R96			X	7			X	8
Haemorrhagic stroke (CVAH)	I60,I61,I62			X	5	X	5	X	12
Ischemic stroke (CVAI)	I63,I65			X	2	X	2	X	4
Other stroke (OCVA)	I64,I66			X	1	X	3	X	2
Other Cardiovascular diseases (OCVD)	G45,I67,I69,I70-I74,I50			X	16			X	19
Total observed events			***38***		***67***		***48***		***105***
Composite endpoints (morbidity + mortality)									
Ischemic Heart disease (IHD)	I20-I25								
Coronary heart disease (CHD)	I20-I25,I46,R96								
Cerebrovascular accident (CVA)	I60-I66					X			
Cardiovascular disease (CVD)	I20-I26,I46,R96,G45, I60-I67,I69,I70-I74,I50			X				X (only fatal events)	
Overall observed events			***221***		***928***		***350***		***105***

* The primary endpoint for ATP III is 'hard CHD', however model ATP III was based on the previously developed Framingham risk score with total CHD as primary endpoint [27]. For this study, the endpoint defined in the original ATP III paper is followed, i.e. endpoint 'hard CHD' is used.

Table 2: Reclassification table where all individuals are classified and considered for treatment, according to the FRS risk threshold (10%), and reassigned for treatment according to the thresholds according to ATP (10%), PCE (7.5%), and SCORE (5%)

		Framingham risk prediction (percentiles)	
		Below (risk < 10%) N = 15,519	Above (risk ≥ 10%) N = 2618
	Mean predicted risk	3.26%	16.15%
	Observed events	489	439
Reclassification			
ATP	Mean predicted risk	0.68%	4.42%
	Observed events	209	12
		N (%)	N (%)
	Below (risk < 10%)	15,519 (100)	2,586 (98.78)
	Above (risk ≥ 10%)	0 (0)	32 (1.22)
PCE	Mean predicted risk	1.17%	6.44%
	Observed events	172	178
		N (%)	N (%)
	Below (risk < 7.5%)	15,469 (99.68)	1,894 (72.35)
	Above (risk ≥7.5%)	50 (0.32)	724 (27.65)
SCORE	Mean predicted risk	0.55%	0.76%
	Observed events	87	18
		N (%)	N (%)
	Below (risk < 5%)	15,330 (98.78)	2,562 (97.86)
	Above (risk ≥ 5%)	189 (1.22)	56 (2.14)

All individuals are separated into two subgroups "below" and "above" based on the FRS risk threshold, with the following definitions; ***below*** - individuals with a predicted risk < 10% (no treatment), and ***above*** - individuals with a predicted risk ≥ 10% (treatment). For each (FRS-)subgroup (column 3-4), the number of individuals present (N) and their average predicted FRS risk (%) is shown. For each FRS-subgroup, individuals are reassigned into two (sub-) subgroups below or above according to ATP (row 5-8), PCE (row 10-13), and SCORE (row 15-18). The green highlighted cells indicate *concordance* and blue highlighted cells indicate *discordance* on the classification of individuals.

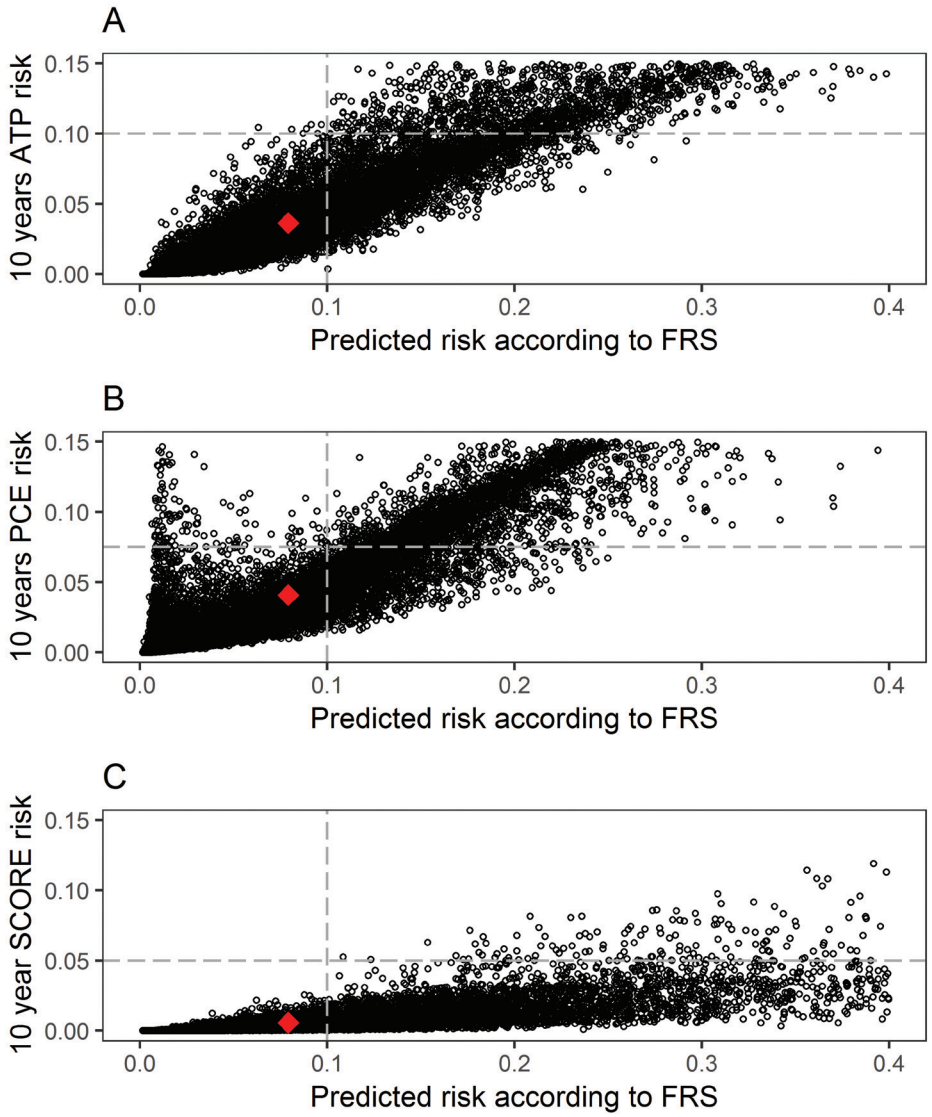


Figure 1: Predicted (absolute) CVD risk according to FRS and A) ATP, B) PCE, and C) SCORE.
The red marker is the estimate of the mean predicted risk according to FRS and ATP, PCE, or SCORE.
The grey lines (raster lines) represent the different risk thresholds and reveal the fraction of individuals eligible for treatment.

REFERENCES

1. Nayor, M. and R.S. Vasan, *Recent Update to the US Cholesterol Treatment Guidelines: A Comparison With International Guidelines.* Circulation, 2016. 133(18): p. 1795-806.
2. Damen, J.A., et al., *Prediction models for cardiovascular disease risk in the general population: systematic review.* BMJ, 2016. 353: p. i2416.
3. Beswick, A.D., et al., in *A Systematic Review of Risk Scoring Methods and Clinical Decision Aids Used in the Primary Prevention of Coronary Heart Disease (Supplement)*. 2008: London.
4. Allan, G.M., et al., *Agreement among cardiovascular disease risk calculators.* Circulation, 2013. 127(19): p. 1948-56.
5. Kent, D.M. and N.D. Shah, *Risk models and patient-centered evidence: should physicians expect one right answer?* JAMA, 2012. 307(15): p. 1585-6.
6. Cooney, M.T., et al., *Cardiovascular risk-estimation systems in primary prevention: do they differ? Do they make a difference? Can we see the future?* Circulation, 2010. 122(3): p. 300-10.
7. Jackson, R., A. Kerr, and S. Wells, *Vascular risk calculators: essential but flawed clinical tools?* Circulation, 2013. 127(19): p. 1929-31.
8. Goff, D.C., Jr., 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, 2014. 63(25 Pt B): p. 2935-59.
9. 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, 2011. 57(12): p. 1404-23.
10. *Lipid Modification: Cardiovascular Risk Assessment and the Modification of Blood Lipids for the Primary and Secondary Prevention of Cardiovascular Disease*, 2014, NICE: London.
11. D'Agostino, R.B., Sr., et al., *General cardiovascular risk profile for use in primary care: the Framingham Heart Study.* Circulation, 2008. 117(6): p. 743-53.
12. Goff, D.C., Jr., 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.* Circulation, 2014. 129(25 Suppl 2): p. S49-73.
13. Conroy, R.M., et al., *Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project.* Eur Heart J, 2003. 24(11): p. 987-1003.
14. *Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III).* JAMA, 2001. 285(19): p. 2486-97.
15. Beulens, J.W., et al., *Cohort profile: the EPIC-NL study.* Int J Epidemiol, 2010. 39(5): p. 1170-8.
16. Verschuren, W.M., et al., *Cohort profile: the Doetinchem Cohort Study.* Int J Epidemiol, 2008. 37(6): p. 1236-41.
17. Blokstra, A., Smit, H.A., Bueno de Mesquita H.B., Seidell J.C., Verschuren W.M.M. , *Monitoring of risk factors and health in the Netherlands (MORGEN-cohort), 1993–1997*, 2005, RIVM: Bilthoven, The Netherlands.
18. Nederlands Huisartsen Genootschap, *Multidisciplinaire Richtlijn Cardiovasculair Risicomanagement 2011*. Utrecht: Bohn Stafleu van Loghum.
19. Steyerberg, E.W., et al., *Assessing the performance of prediction models: a framework for traditional and novel measures.* Epidemiology, 2010. 21(1): p. 128-38.
20. Harrell, F., *Regression Modeling Strategies*. 2001, New York: Springer.
21. 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 & Rehabilitation (EACPR).* Eur J Prev Cardiol, 2016. 23(11): p. NP1-NP96.
22. Cordoba, G., et al., *Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review.* BMJ, 2010. 341: p. c3920.
23. Kip, K.E., et al., *The problem with composite end points in cardiovascular studies: the story of major adverse cardiac events and percutaneous coronary intervention.* J Am Coll Cardiol, 2008. 51(7): p. 701-7.
24. Heeg, B.M. and B.A. van Hout, *Assessing uncertainties surrounding combined endpoints for use in economic models.* Med Decis Making, 2014. 34(3): p. 300-10.
25. Kent, D.M., et al., *Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal.* Trials, 2010. 11: p. 85.
26. Usher-Smith, J.A., et al., *Impact of provision of cardiovascular disease risk estimates to healthcare professionals and patients: a systematic review.* BMJ Open, 2015. 5(10): p. e008717.
27. Wilson, P.W.F., et al., *Prediction of coronary heart disease using risk factor categories.* Circulation, 1998. 97(18): p. 1837-1847.
28. Taylor, F., et al., *Statins for the primary prevention of cardiovascular disease.* Cochrane Database Syst Rev, 2013(1): p. CD004816.

CLINICAL EXAMPLE

Consider a 57 year-old female, with total cholesterol, low-density-lipoprotein (LDL) and high-density-lipoprotein (HDL) levels of 193 mg/dL, 91 mg/dL, and 54 mg/dL, respectively, a systolic blood pressure (SBP) of 186, no use of blood pressure lowering drugs, who does not smoke and does not have diabetes. This woman has an estimated 10-year CVD risk of 1.9%, 14.1%, 3.2%, and 0.4%, according to ATP, FRS, PCE, and SCORE, respectively. This would indicate a risk above the respective treatment threshold for FRS (10%), but not for ATP (10%), PCE (7.5%) and SCORE (5%).

The use of these different risk prediction models may not only lead to different treatment decisions, but also to different estimates of the expected benefit from preventive treatment. When this woman would receive preventive statin treatment, expected to reduce the overall CVD risk with 30%, this would reduce the risk of non-fatal and fatal MI by 0.6% according to ATP, whereas according to SCORE, the risk of fatal CVD events would decrease by 0.1% [28]. Similarly, the risk of non-fatal and fatal MI and stroke would decrease by 1.0% according to PCE, whereas according to FRS, the risk of the broad range of CVD events included in FRS, both non-fatal and fatal, would be reduced by 4.2%. Apparently, for this woman CVD risk prediction using FRS implies both a greater necessity to consider preventive drug treatment and a larger potential benefit of such treatment, compared to ATP, PCE, and SCORE.

SUPPLEMENTARY

Supplementary – Table 1: Observed and predicted number of CVD events

	# Observed	**# Expected CVD events**		
	CVD events	*Original model*	*Recalibrated model*	*Recalibrated model with correction factor*
ATP	221	581	672	221
FRS	928	1,147	1,456	928
PCE	350	577	751	350
SCORE	105	119	105	105

5.

Assessing the impact of CVD prevention strategies at young age: exploring key parameters and choices

G.R. Lagerweij
G.A. de Wit
Y.T. van der Schouw
B.C.J.M. Fauser
E. Boersma
K.G.M. Moons
H. Koffijberg
on behalf of the CREW consortium

Submitted

ABSTRACT

Background: Manifestations of common and/or severe diseases at old(er) age can sometimes be traced back to elevated risk factor levels at young(er) age. Intervening at young(er) age to reduce the burden of disease may be beneficial, but comes at a cost. Assessing the long-term health benefits in a trial is hardly feasible, conversely, modelling these benefits is likely to be more feasible. As evidence on the impact of prevention at young age is very limited, it is important to study which parameters and assumptions influence the expected long-term benefits of preventive strategies at young age most.

Methods: A micro-simulation model with a lifetime horizon is developed to explore the influence of key parameters on the long-term impact assessment of preventive strategies for cardiovascular diseases (CVD). Women with hypertensive pregnancy disorder (HPD) are used as a case study and screening at 30, 40, or 50 years is investigated as preventive strategy. Different preventive strategies are compared in terms of total number of CVD events, costs, and quality-adjusted-life-years. Uncertainty is investigated by Monte-Carlo simulation. Additionally, a Value of Information analysis and net benefit regression are performed to identify parameters that warrant future further research and that have a large influence on the health benefits.

Results: Parameters associated with treatment effectiveness, adherence to medication and average 10-year CVD risk at 80 years have the largest impact on model outcomes. All investigated preventive strategies for CVD in women with HPD are cost-effective, with screening every 10 years at the age of 30, 40, and 50 having the highest probability to be cost-effective.

Conclusion: Decision analytic models can determine those parameters that impact the long-term impact of preventive interventions most. In the case of prevention of CVD in women with HPD, it would be crucial to further study risk reduction of preventive treatment and the long-term adherence rate of medication in women.

1. INTRODUCTION

The development of a disease can take decades before first manifestations emerge, many cardiovascular diseases (CVDs) have such long incubation period being a progressive consequence of atherosclerosis which often begins early in life [1]. In 2016, 15.2 million people died due to CVD which represents 27% of all global deaths [2]. Risk factors for CVD, e.g. obesity, smoking, high blood pressure and lipids, mainly result from a prolonged unhealthy lifestyle, e.g. inadequate nutrition and physical inactivity. Primary prevention of CVD is possible by changing lifestyle, e.g. by losing weight, increasing physical activity, quitting smoking, or administering lipid and blood pressure lowering drugs [3].

Several guidelines recommend the use of CVD risk prediction models to stratify individuals into CVD risk categories, and based on absolute CVD risks that exceed a certain threshold, preventive treatment strategies are administered. For example, a lipid lowering drug prescription is advocated in case of a 10-year CVD risk estimate above the threshold of 7.5% according to prediction model Pooled Cohort Equations (PCE) [4, 5]. As age is an important risk factor for CVD, current risk-based selection of individuals mainly targets the more elderly, even though the process of developing CVD already starts at a young(er) age. Intervening already in the twenties, thirties or forties could be a substantial component of the global effort to reduce the burden of CVD worldwide, and may be supported by a burden-based rather than risk-based selection approach [6].

Over the last decades evidence has grown that other, non-traditional and non-modifiable, risk factors also affect the manifestation of CVD. Quantifying benefits from CVD prevention requires taking into account such new CVD risk factors since they also contribute to CVD risk. Examples of non-modifiable risk factors are preterm delivery, hypertensive pregnancy, and autoimmune diseases [7-9]. Of the non-modifiable CVD risk factors, pregnancy disorders are mostly described in the literature as early manifestation of CVD.

The long-time horizon over which health benefits from CVD prevention may accrue makes observing and measuring them infeasible in randomized trials or prospective longitudinal cohort studies. Another way of quantifying long-term benefits from early CVD prevention strategies is to perform a model-based analysis [10]. However, developing a valid decision analytic model to assess prevention at young age, with all required evidence on CVD risk,

effectiveness of the preventive strategies and adherence to these strategies over time, is challenging. For example, adherence to prescribed medication (such as blood pressure or lipid lowering medication) is highly variable over time, treatment effectiveness of such medication (as a function of age and other risk factors) may change over time, as well as patterns of risk progression [11, 12]. All these aspects are likely to be affected also by gender and baseline risk factor levels. A trustworthy model-based analysis of early prevention strategies needs to parameterize these aspects, and incorporate the uncertainty associated with the evidence used to establish the parameter estimates.

We have developed a micro simulation model to explore which key (sets of) parameters have the largest influence on the expected long-term benefits of CVD prevention in young women with pregnancy related CVD risk factors. Rather than accurately quantifying the expected benefits for a particular early prevention strategy to be implemented in actual daily management of young women with pregnancy related CVD risk factors, this paper aims to study the influence of parameters and other crucial evidence in the assessment of the long-term impact of early CVD prevention strategies.

2. METHODS

2.1 General remarks

We developed a discrete time micro simulation state transition model as this allows inclusion of CVD risk factors, simulated events and outcomes on an individual level [13]. The cycle length of the model is one year and all individuals are followed until death (i.e. a lifetime time horizon is applied). Then, outcomes over time are aggregated at population level, i.e. total number of CVD events, total costs and total health outcomes, expressed in Quality-Adjusted Life Years (QALYs).

Appendix A shows the flow chart of the micro simulation model and Appendix B - Table 1 shows an overview of all input parameters that are varied in the modelling. Given the scarcity of evidence we divide parameters into two categories: a) parameters with an uncertain average and/or unknown range (row 3-38), and b) parameters related to policy decisions and therefore with fixed values (row 39-44). Examples of parameters with an uncertain average and known range are average costs and utilities of CVD event types, whereas parameters with unknown

average and range are for example, the long-term adherence rate of medication. This last group may play a large role in the cost-effectiveness of preventive screening. Therefore, we will focus in more detail on these five parameters, i.e. parameters with uncertain average and unknown range.

Below we first explain the natural history of the individuals who are followed in the micro simulation model. Second, the early prevention strategies for cardiovascular disease are explained. Third, the impact analysis including the reduced incidence of CVD events and cost effectiveness analysis of these strategies is specified. Last, we describe the value of information (VOI) analysis and Net Benefit Regression (NBR) which are used to identify parameters with large influence on the cost-effectiveness estimates of these early CVD prevention strategies [14].

2.2 Natural history of CVD in targeted individuals

An example of non-traditional and non-modifiable CVD risk factors are pregnancy related and reproductive disorders in women [15]. Some pregnancy and reproductive related disorders can be viewed as early manifestation of CVD. For example, women with a hypertensive pregnancy disorder (HPD) have a two-fold increased risk of CVD [16, 17]. In this paper, we focus on the long-term benefits of early prevention in women with HPD as a case study to identify key parameters in this context.

Survival

The survival of young women is simulated based on sex- and age-specific life tables from Statistics Netherland [18].

CVD risk and correlation

It is assumed that the population of women with HPD has an average age of 30 years, i.e. the average age of women at a first pregnancy in the Netherlands [18]. To estimate annual and 10-year CVD risks, a hypothetical dataset is populated with general and easily measurable risk factors, e.g. age, cholesterol levels and systolic blood pressure.

In the absence of evidence on change over time, for example in CVD risk factor levels, we predict individualized CVD risk estimates before running the simulation. Based on literature, the average 10-year CVD risk at young and older age, in this study 30 and 80 years, are determined

[19]. It is assumed that after the age of 80 years, the 10-year CVD risk stays constant. The 10-year CVD risk estimates for the years in between 30 and 80 years are determined by smooth exponential interpolation. These 10-year CVD risk estimates per age decade are used to calculate corresponding annual CVD risk for use in the cycles of the simulation model. For more details see Appendix C

On average risk estimates of CVD increase with age. Unfortunately there is no reliable data available on the potential change in the risk profiles within women over time. In the absence of evidence on the correlation between 10-year CVD risk estimates at young age and at older age, a single correlation coefficient is used to correlate the risk values within women over time. A small value indicates small correlation between the risks, i.e. women have substantial chance to move from a relative high predicted risk to a relative low risk and vice versa. More details are shown in Appendix C.

CVD primary prevention strategies

Preventive screening for CVD is common (certainly in western countries) but is often not systematically provided and conducted across the entire population. It may vary between care providers and young women are often not included in CVD prevention [20, 21]. The natural history (i.e. the CVD occurrence without any primary prevention) is therefore hard to define.

There is no nationwide primary prevention program in the Netherlands. However, the general practitioner may identify women at high CVD risk, for example based on early symptoms or complaints. These women then are classified into risk categories based on their predicted 10-year CVD risk according to the SCORE model [22-24]. Women at high risk, i.e. a risk estimate exceeding the risk threshold of 10%, receive preventive medical treatment.

The expected risk reduction (based on statin use) is assigned to each woman who adheres to their medication [25]. For the medication adherence rate, it is assumed that the annual adherence rate decrease over time. Evidence on the proportion of women who are adherent to their medication for a long time horizon is not available. Therefore, we assume all high risk women start with medication and a proportion of women adhere to their medication after 10 years, and derive the annual rate accordingly, i.e. exponential interpolation. The 10-year adherence rate is varied in the analyses, average of 0.25 (95% CI 0.01-0.50).

Risk of a CVD event

For simplicity, we distinguish three CVD event categories; coronary heart disease (CHD), cerebrovascular disease (CVA), and other cardiovascular disease (OCVD) events. All three CVD event categories can vary in severity, i.e. non-fatal and fatal, resulting in six CVD event types being incorporated in the model. The event distribution (i.e. relative occurrence) of these six CVD event types is based on literature and cohort studies, and is dependent on age (Appendix B - Table 2). For all women who experience a non-fatal event in the current cycle, the annual CVD risk estimates are proportionally increased since experiencing a CVD event is a risk factor for a recurrent CVD event (see Appendix B - Table 1 for further details).

Quality of life and costs

Quality of life values (utilities) of women with HPD are based on evidence from Dutch studies and the National Institute for Health and Care Excellence (NICE) [26, 27]. Quality of life is adjusted for age [28, 29]. Women, who experience a non-fatal CVD event in the previous or current cycle, have their quality of life proportionally decreased. CVD events can occur multiple times in this model. The proportional reduction in quality of life due to a CVD event is higher in the first year compared to the consecutive years to reflect recovery from the CVD event towards a normal quality of life. For non-fatal first CVD events, the utilities vary over the six CVD event types. Side-effects of medication after a CVD event are not taken into account separately but are assumed to be incorporated in the disutility of CVD events. Utilities of recurrent CVD event do not vary over event type.

Costs of CVD events are based on Dutch studies and evidence from NICE, and vary over the six event types. Costs of recurrent CVD events are set at the same value. Higher costs for the first year after the occurrence of a CVD event are incorporated. For screened women, disutility due to the screening is not taken into account. In women receiving preventive treatment, an average disutility of the treatment (due to minor side-effects) is taken into account for medication adherent women. For all screened women, costs are updated with the costs due to screening, i.e. visit to the GP and laboratory tests. For women with a CVD risk exceeding the treatment threshold, the costs for medication are added regardless of their adherence.

All values of the utilities and costs can be found in Appendix B - Table 1 (row 8-37). Following Dutch guidelines, a discount rate of 4% for costs and 1.5% for health outcomes is applied [30]. It is assumed that preventive screening and CVD events can occur halfway through the cycle, i.e. year. Therefore, a mid-year discount and half-cycle correction is applied for the costs and utilities which are accumulated at the end of each cycle.

2.3 Preventive screening strategies under study

In a simulated preventive strategy, we assume that women with HPD are invited for early screening for CVD. However, only a proportion of these women, varied in the simulation between 0.6 and 0.9, will participate in the preventive screening. This preventive screening is based on 10-years predicted CVD risks, and stratification of screened women into a risk category based on this absolute risk. Current guidelines on CVD management across the globe commonly advocate to use only one risk threshold, i.e. 7.5% - 20% depending on the country, for administering preventive medication, which is then applied to all age groups. Since the average 10-year CVD risk is very low in women at 30 years (i.e. around 2.5%), these recommended risk thresholds are not directly suitable for our case study. Therefore, a lower absolute risk threshold of 2% was chosen for the purpose of our study.

It is unknown how young women would follow-up on the advice to use preventive medication following detection of an increased risk of CVD. Therefore, we varied the adherence rate, relative to the adherence rate that is observed in older women with an increased CVD risk (see natural history, section 2.2.3). We use an adherence rate varying between 0.5 and 1.5, relative to adherence in older women with “classical” CVD risks.

Table 1 shows an overview of the different preventive scenarios considered. To investigate the potential effect of key parameters, i.e. parameters with an unknown average and range, and different preventive screening scenarios for young women with HPD, we simulate a cohort of 2,000 women and compare the resulting number of simulated CVD events.

2.4 Cost effectiveness analyses

A cost-effectiveness analysis is performed with an incremental cost-effectiveness ratio (ICER) as outcome. This ICER represents the ratio of the difference in lifetime costs between early preventive screening and no screening (natural history) divided by the difference in effectiveness (i.e. health outcomes) between both scenarios. The effectiveness is measured in quality adjusted life years (QALYs). A commonly applied Dutch Willingness-to-Pay (WTP) threshold of € 20,000 per QALY gained is used to determine whether a screening strategy is cost-effective or not, and to calculate the incremental Net Health Benefits (INHB).

To explore the influence of parameters with uncertain values and range, probabilistic sensitivity analysis is performed using 2,000 Monte Carlo simulations. Last, the probability that a screening strategy is cost-effective compared to all alternative strategies (including no screening) is estimated, as a function of the WTP, and visualized in cost-effectiveness acceptability curves.

2.5 Value of Information

A value of information (VOI) analysis is performed using the online Sheffield Accelerated Value of Information (SAVI) tool [14]. VOI can be used to investigate the value of collecting additional data to reduce uncertainty in the input parameters of the simulation model, in order to reduce uncertainty in cost-effectiveness outcomes, and thereby the risk of incorrect decisions, for example regarding reimbursement [31]. The expected value of perfect (EVPI) and partial perfect information (EVPPI) is calculated. The value of hypothetically resolving all uncertainty is reflected by the EVPI, whereas, the EVPPI indicates what the value is of resolving all uncertainty in one parameter or a group of parameters [31, 32].

Since costs of screening based on cardiovascular risk prediction are low, and preventive drug therapy with blood pressure and lipid lowering drugs is relatively cheap and effective, there may be limited uncertainty regarding the cost-effectiveness of early screening compared to no screening. Therefore, we also perform a Net Benefit Regression (NBR) to investigate which parameters have a substantial influence on the benefits of early screening strategies, using a multivariable linear regression model [33] .

3. RESULTS

3.1 Impact of key parameters

The impact of the five key parameters is investigated in terms of CVD events before the age of 60 and lifelong, in a cohort with 2,000 women. The results are presented in Appendix D, together with the initial parameter settings used for the simulations. The relation between one key parameter, i.e. probability to participate in preventive screening, and the actual number of (expected) CVD events is shown in Figure 1. The effect of this parameter on the number of CVD event is substantial for the preventive screening scenarios. For example, for screening starting at 30 years, the number of lifelong CVD events decreases by 9% and 13% for a probability to participate in screening of 0.60 and 0.75 respectively.

3.2 Cost effectiveness analyses

It is expected that the main driver of the impact of CVD preventive strategies is the reduced number of CVD events. Table 2 shows the number of CVD events for no screening and the five screening strategies of Table 1. A screening strategy with 3 screening moments at 10 year intervals results in the largest decrease in CVD events, i.e. 35% of all life-long CVD events can be prevented by screening at 30-40-50.

Table 3 shows the results of the cost-effectiveness analysis for five different screening scenarios. No screening is more costly than all five preventive screening scenarios; the mean cost for no screening is €6,376 per women. Cost savings of preventive screening versus no screening is largest when preventive screening at 30-40-50 is applied: mean incremental costs are -€ 543. Moreover, preventive screening at 30-40- also has the largest health benefit; the mean incremental benefit compared to no screening is 0.20 QALYs.

Figure 2 shows the incremental cost-effectiveness plane with a WTP threshold of €20,000/QALY. Despite the substantial uncertainty present in the model input parameters, in almost all simulations preventive screening is less costly and more effective compared with no screening. All preventive screening strategies dominate no screening. To compare the preventive screening strategies with each other, we used incremental net health benefit (INHB) estimates with a WTP threshold of €20,000/QALY. Preventive screening at 30-40-50 has the largest INHB estimate (Table 3–column 7). The cost-effectiveness plane comparing the preventive

screening strategies with no screening can be found in Appendix E, together with the associated ICERs (Appendix E - Table 1). Figure 3 shows the cost-effectiveness acceptability curve. Preventive screening at 30, 40, and 50 years has the largest probability to be cost-effective for different WTP thresholds, increasing from 0.75 for a WTP of €0/QALY to 0.94 for a WTP of €100,000/QALY.

3.3 Value of information

The value of information (VOI) analysis indicates that the overall EVPI per person affected is €4.69 per person, or 0.0002 QALYs per person. Furthermore, the analysis indicates that treatment effectiveness, in terms of relative risk reduction, has the largest EVPI value per person (i.e. €0.22). Further investigation of the uncertainty of single parameters shows that collecting more information on the five key parameters, i.e. parameters with an unknown average and range, has no added value (see Appendix F - Table 1). Assuming 2,000 women present with HPD in the Netherlands per year amounts to a population EVPI of € 9,380. When considering only decision uncertainty with respect to cost-effectiveness, this low value of EVPI would not merit investing in further research, as the cost of research would surpass the costs of further reducing uncertainty.

Table 4 shows an overview of the groups of associated parameters that are used to estimate the group EVPPI. Collecting additional information on all parameters related to treatment (i.e. set 3), has a limited value with an EVPPI of €0.71 per person. Additionally, collecting additional information on all other parameters has no value (see Appendix F - Table 1). To make sure individual patient variation (i.e. first order uncertainty) is excluded from the VOI analyses, these were repeated based on 2,500 simulation runs of a cohort of 50,000 women. The result is an overall EVPI of € 0 per person (i.e. none of the 2,500 PSA samples lie above the WTP threshold) implying that there is near perfect certainty that preventive screening at 30, 40, and 50 years is cost-effective, for the applied WTP of €20,000/QALYs.

The results of the cost-effectiveness analysis show that preventive screening strategies are (almost always) cost saving and beneficial (Figure 2). However, there is some variation in the size of the health benefits. The incremental net health benefit (INHB) estimates, i.e. preventive screening at 30, 40, and 50 years compared with usual care, is used as outcome in the NBR.

Table 5 (column 2-4) shows the results of the NBR, where most key parameters have a linear regression coefficient which is significant. The associated R^2 value of the estimated NBR is 0.84 which is acceptable for using the estimated model [34]. Utility of preventive medication has the largest coefficient (12.89). In other words, the change in INHB is almost 13 if the utility of medication increases with 1. However, in this study the utility of medication varies between 0.997 and 0.999, hence the influence of this parameter on the INHB is very small. Therefore, we determine the 2.5 and 97.5 percentile values for all parameters, based on their respective distributions, and use these estimate of the low-INHB (2.5%) and high-INHB (97.5%) value per parameter, while keeping all other parameters fixed at their mean value.

Table 5 (column 5-6) shows the results for the estimation of the low-INHB and high-INHB values. For example, the expected INHB of 0.22 QALYs (Table 3) is predicted to increase to 0.41 if the relative risk reduction value changes from 0.70 to 0.49 (i.e. 2.5 percentile value) and to decrease to 0.06 if this value changes from 0.70 to 0.88 (i.e. 97.5 percentile value). Parameters with substantial influence on the benefits of screening are: treatment effectiveness, 10-year adherence rate to medication (for all women), average 10-year CVD risk at 80 years and the proportion of women that participate in screening. When, regardless of cost-effectiveness and VOI outcomes, more accurate estimates of these screening benefits are desired future studies should focus on collecting further evidence on these aspects.

4. DISCUSSION

Modelling may be the only feasible method to estimate the long-term benefits of early prevention but it requires a large number of parameters for which evidence is likely to be lacking at the time of modelling. In this study, different (sets of) parameters for the long-term benefits of early CVD prevention strategies are investigated with regard to their relevance to cost-effectiveness outcomes and decision making. Investigating which (sets of) parameters are most relevant allows to better focus further research. In our illustration that focuses at CVD prevention in women with HPD, the VOI and NBR analyses indicated that the most relevant parameters are intervention (treatment) effectiveness and adherence rate to intervention. Despite including substantial and realistic uncertainty in model input parameters, the results of our analysis tentatively suggest that early CVD prevention in women with HPD may have the potential to improve health outcomes and reduce health care costs.

Strengths of the study

We explicitly included parameters that are commonly excluded in health economic evaluations of primary screening for CVD, e.g. correlation of risk profiles in individuals over time, and different levels of adherence to the preventive intervention (here lipid lowering medication) in younger and older women. Including such parameters captures more of the uncertainty in the health economic outcomes, whereas (implicitly) assuming a correlation of exactly 1, and equal medication adherence across age ranges ignores uncertainty in these parameters, and obviously is distant from what is being observed in real life.

Few studies on model-based impact assessment of CVD prevention strategies can be found in the literature [35, 36]. A quick research in PubMed showed approximately 340 results of which only four studies modelled a lifelong time horizon. Many model-based impact assessments of CVD prevention strategies have a time horizon between 10 and 30 years, i.e. follow-up ends before death, whereas our study models the entire life-time of women entering the model at young age [37]. Furthermore, we used a patient-level state transition model rather than a cohort state transition model, to accurately reflect individual risks, treatment decisions, adherence, and patient pathways.

Limitations of the study

For the parameters in the model with an unknown average and/or range (e.g. probability to participate in preventive screening), a very wide distribution was defined, deliberately, to not underestimate uncertainty in these parameters. Even so, not all model aspects were parameterized and it was necessary to make a limited number of assumptions. We assumed that the relative risk of CVD when taking medication versus without medication was similar for all CVD event types and also across age categories. Evidence demonstrating otherwise is lacking. Furthermore, data on CVD risk after 80 years of age was lacking and was therefore kept constant beyond this age in the model. As the vast majority of CVD events in our model occurs before the age of 80, the influence of this assumption is however limited.

The aim of this study was to explore influence of different (sets of) parameters on the long-term impact of preventive strategies rather than to conclusively demonstrate what is the best preventive strategy in women with HPD. The structure of the model was chosen to be as simple as possible while still realistic enough to provide relevant insights. An example of the “simplistic” structure of the model is that only the occurrence of CVD events was

modelled; occurrence of other diseases or comorbidities, like chronic kidney disease, or migraine, or inflammation, were not taken into account [38, 39]. Also, unrelated medical costs in life-years gained were not taken into account. Last, the practical feasibility and acceptability of the proposed preventive screening strategies and associated preventive medication for individuals and care providers, still needs to be determined. In daily practice, young women may not be willing to take medication for prolonged periods due to the risk of side-effects. However, the opposite may also be true since these women are increasingly aware of their (high) CVD risk after experiencing a complicated pregnancy, which may increase their adherence to drug therapy.

Furthermore, with an absolute risk threshold of 2% for treatment selection, preventive screening was profitable in terms of health benefits, cost savings, and cost-effective (even almost dominant). However, the absolute risk threshold of 2% was chosen and may not be the optimal threshold [40].

Conclusion

Decision analytic modelling is crucial in assessing the expected long-term impact of preventive interventions [10]. Moreover, it can play an important role in determining which evidence is required for reliable impact assessments to be accurate and thus valuable for decision making. Collecting evidence over long periods is time consuming and costly, therefore it is important to prioritize research in this area which then contributes to efficient allocation of healthcare resources. Finally, after using decision analytic models to explore relevant evidence gaps, reusing and updating these models to assess long-term impact is likely to be feasible and an efficient use of research time and budgets.

Table 1: Preventive screening strategies in young women with HPD

	Screening moment	Absolute risk threshold
1	50	2%
2	30	2%
3	40 & 50	2%
4	30 & 40	2%
5	30 & 40 & 50	2%

Table 2: Incremental number of CVD events in preventive screening compared to no screening

No screening (N=1,178 CVD events life-long in a cohort of 2,000 women)		
		Incremental events CVD (N, %)
Preventive screening (risk threshold of 2%)	50 years	-141 (11.9%)
	30 years	-157 (13.2%)
	40 and 50 years	-300 (25.2%)
	30 and 40 years	-319 (26.9%)
	30, 40, and 50 years	-411 (34.6%)

* the probability to participate in preventive screening is 75%, other value of the initial screening settings can be found in Appendix D.

Table 3: Results from the cost effectiveness analyses

	Average costs (€)	**Average health benefits (QALY)**				
No screening	6,396	29.71				
Preventive screening	Average costs (€)	Average health benefits (QALY)	Incremental cost* (€) (95% CI)	Incremental health benefits* (QALY) (95% CI)	ICER (€/QALY)	INHB (95% CI)
50 years	6,223	29,80	-173 (-531;32)	0.086 (-0.01;0.21)	Dominant	0.09 (0.01;0.24)
30 years	6,115	29,80	-281 (-981;96)	0.084 (0.00;0.26)	Dominant	0.10 (0.00;0.31)
40 and 50 years	6,044	29,86	-352 (-1,026;46)	0.147 (-0.02;0.36)	Dominant	0.16 (0.02;0.42)
30 and 40 years	5,950	29,86	-446 (-1,355;107)	0.146 (-0.01;0.39)	Dominant	0.17 (0.01;0.47)
30, 40, and 50 years	5,852	29,91	-543 (-1,512;89)	0.198 (0.03;0.47)	Dominant	0.22 (0.02;0.56)

* compared with no screening. ICER: Incremental Cost-Effectiveness Ratio. QALY: Quality Adjusted Life Year. INHB: Incremental Net Health Benefit

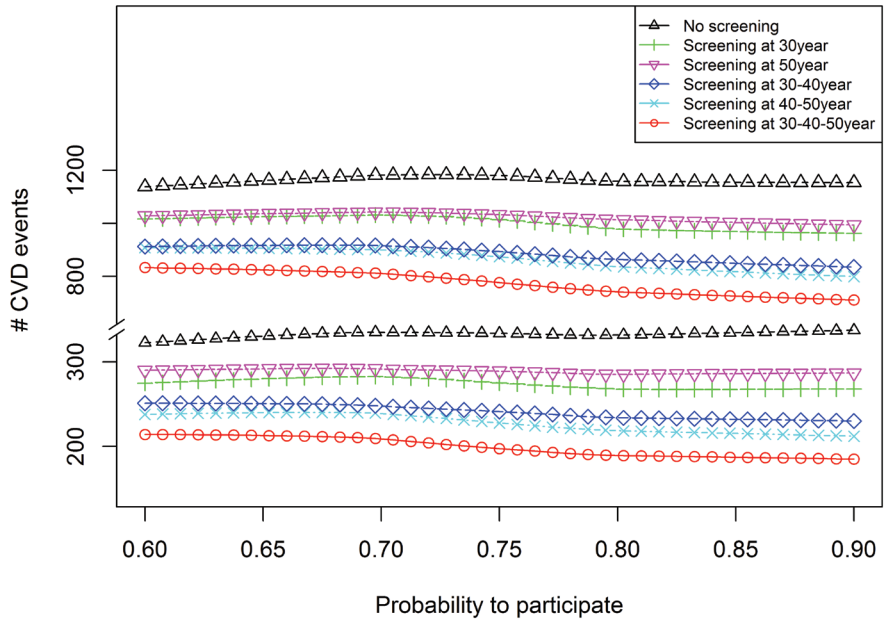


Figure 1: Effect of unknown probability to participate in preventive screening in women below the age of 60. A line in the **lower** part presents the number of CVD events in women below the age of 60, and a line in the **upper** part presents the number of predicted events over the entire life span of women with HPD

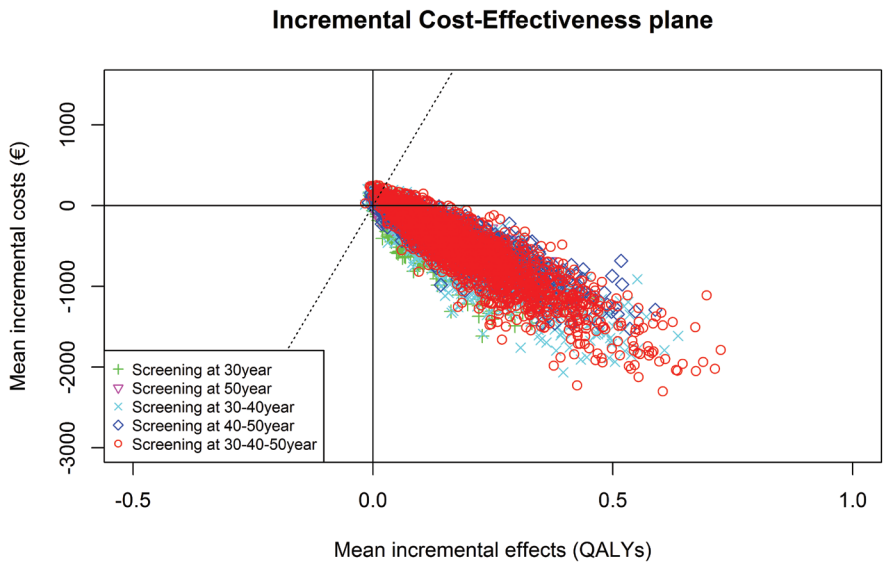


Figure 2: PSA results for 2,000 simulations and a risk threshold of 2% and the dotted line is the WTP threshold of €20,000/QALY. The points in the figure present the difference in effects and costs of a preventive screening versus no screening.

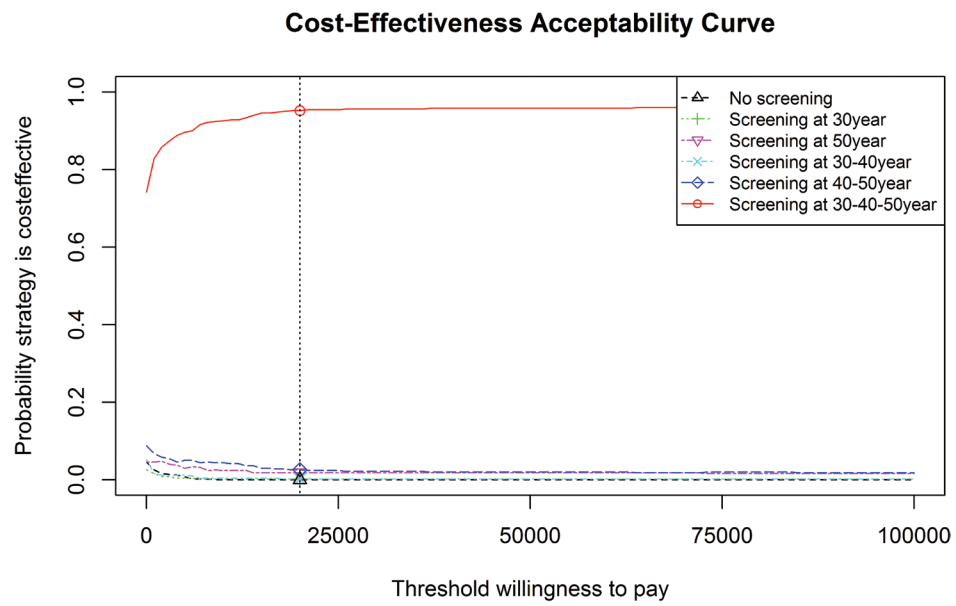


Figure 3: Cost-effectiveness acceptability curve for different screening scenarios; the dotted line represents the WTP threshold of €20,000/QALY.

Table 4: Defined sets of parameters for VOI analysis

Set 1 – Predicted CVD risk

- Average 10-year CVD risk at young age, i.e. 30 years
- Average 10-year CVD risk at older age, i.e. 80 years
- Correlation coefficient between 10-year CVD risk estimates

Set 2 – Probability to start preventive medication and stay adherent

- 10-year adherence for women ≥ 60 years
- Proportion of young women that participate in early preventive screening
- Relative adherence change (in women < 60 years compared with > 60 years)
- Annual proportion of women ≥ 60 years at high CVD risk identified by the GP

Set 3 – Screening and treatment

- Relative risk of CVD when taking medication versus without medication
- Disutility due to preventive medication
- Cost of preventive medication
- Cost of early preventive screening

Set 4 – Costs

- Cost of CHD event (first year)
- Cost of CVA event (first year)
- Cost of other CVD event (first year)
- Cost of CVD death
- Cost of CHD event (sequential years)
- Cost of CVA event (sequential years)
- Cost of other CVD event (sequential years)
- Cost of recurrent CVD event (first year)
- Cost of recurrent CVD event (sequential year)

Set 5 – Utilities

- Utility of CHD event (first year)
- Utility of CVA event (first year)
- Utility of other CVD event (first year)
- Utility of CHD event (sequential year)
- Utility of CVA event (sequential year)
- Utility of other CVD event (sequential year)
- Utility of recurrent CVD event (first year)
- Utility of recurrent CVD event (sequential years)

Set 6 – Relative risk after CVD event

- Relative risk of recurrent CVD event (first year)
- Relative risk of recurrent CVD event (sequential year)

Table 5: Net benefit regression based on all parameters with a willingness-to-pay threshold of €20,000/QALY

	Estimate	Standard error	p-value	Low INHB	High INHB
Relative risk of CVD when taking medication versus without medication	*-0.91*	0.01	0.00	*0.41*	*0.06*
10-years adherence for women ≥ 60 years	*0.50*	0.01	0.00	*0.11*	*0.34*
Relative adherence change (in women < 60 years compared with > 60 years)	*0.13*	0.00	0.00	*0.16*	*0.28*
Average 10-year CVD risk at 80 years	*0.83*	0.04	0.00	*0.18*	*0.26*
Utility of CHD (sequential year)	*-0.16*	0.01	0.00	*0.27*	*0.21*
Proportion of young women that participate in early preventive screening	*0.19*	0.02	0.00	*0.20*	*0.25*
Utility of CVA (sequential year)	*-0.12*	0.01	0.00	*0.25*	*0.20*
Annual proportion of women ≥ 60 years at high CVD risk identified by the GP	*-0.36*	0.04	0.00	*0.24*	*0.21*
Utility of preventive medication	*12.89*	2.27	0.00	*0.21*	*0.24*
Utility of OCVD (sequential year)	-0.03	0.01	0.04	0.23	0.22
Relative risk recurrent event (first year)	0.01	0.01	0.08	0.22	0.23
Cost of CVD death	0.00	0.00	0.09	0.23	0.22
Correlation coefficient between 10-year CVD risk estimates	-0.01	0.01	0.06	0.23	0.22
Cost of preventive medication	0.00	0.00	0.12	0.23	0.22
Cost of CVA (first year)	0.00	0.00	0.14	0.23	0.22
Cost of recurrent event (first year)	0.00	0.00	0.15	0.22	0.23
Cost of CHD (sequential year)	0.00	0.00	0.16	0.23	0.22
Cost of CHD (first year)	0.00	0.00	0.21	0.23	0.22
Cost of CVA (sequential year)	0.00	0.00	0.30	0.22	0.23
Utility of recurrent event (first year)	-0.02	0.02	0.35	0.23	0.22
Relative risk recurrent event (sequential year)	0.00	0.01	0.37	0.22	0.23
Cost of recurrent event (sequential year)	0.00	0.00	0.48	0.22	0.23
Average 10-year CVD risk at 30 years	-0.36	0.45	0.42	0.23	0.22
Utility of CHD (first year)	0.01	0.02	0.54	0.22	0.23
Cost of OCVD (first year)	0.00	0.00	0.59	0.22	0.23
Cost of preventive screening	0.00	0.00	0.74	0.22	0.22
Utility of CVA (first year)	0.00	0.02	0.79	0.22	0.22
Cost of OCVD (sequential year)	0.00	0.00	0.88	0.22	0.22
Utility of OCVD (first year)	0.00	0.02	0.89	0.22	0.22
Utility of recurrent event (sequential year)	0.00	0.01	0.93	0.22	0.22
Intercept	-11.40	2.37	0.00		
R^2	0.837				

REFERENCES

1. Fifth Joint Task Force of the European Society of, C., et al., *European Guidelines on cardiovascular disease prevention in clinical practice (version 2012): the Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts).* Eur J Prev Cardiol, 2012. 19(4): p. 585-667.
2. *Key facts of Cardiovascular Disease*. [cited 2018 05/29]; Available from: http://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).
3. *Key Facts of Cardiovascular Disease*. Available from: http://www.who.int/mediacentre/factsheets/fs317/en/.
4. Goff, D.C., Jr., 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*. Circulation, 2014. 129(25 Suppl 2): p. S49-73.
5. 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, 2011. 57(12): p. 1404-23.
6. Lagerweij, G.R., et al., *A new selection method to increase the health benefits of CVD prevention strategies*. Eur J Prev Cardiol, 2018. 25(6): p. 642-650.
7. Matsuura, E., et al., *Is atherosclerosis an autoimmune disease?* BMC Med, 2014. 12: p. 47.
8. Frostegard, J., *Atherosclerosis in patients with autoimmune disorders.* Arterioscler Thromb Vasc Biol, 2005. 25(9): p. 1776-85.
9. Brown, M.C., et al., *Cardiovascular disease risk in women with pre-eclampsia: systematic review and meta-analysis.* Eur J Epidemiol, 2013. 28(1): p. 1-19.
10. Buxton, M.J., et al., *Modelling in economic evaluation: an unavoidable fact of life.* Health Econ, 1997. 6(3): p. 217-27.
11. Gupta, P., et al., *Risk Factors for Nonadherence to Antihypertensive Treatment.* Hypertension, 2017. 69(6): p. 1113-1120.
12. Molloy, G.J., et al., *Intentional and unintentional non-adherence to medications following an acute coronary syndrome: a longitudinal study.* J Psychosom Res, 2014. 76(5): p. 430-2.
13. Rutter, C.M., A.M. Zaslavsky, and E.J. Feuer, *Dynamic microsimulation models for health outcomes: a review.* Med Decis Making, 2011. 31(1): p. 10-8.
14. Strong, M., J.E. Oakley, and A. Brennan, *Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach*. Med Decis Making, 2014. 34(3): p. 311-26.
15. Appelman, Y., et al., *Sex differences in cardiovascular risk factors and disease prevention*. Atherosclerosis, 2015. 241(1): p. 211-8.
16. Garcia, M., et al., *Cardiovascular Disease in Women: Clinical Perspectives*. Circ Res, 2016. 118(8): p. 1273-93.

17. Bouthoorn, S.H., et al., *Ethnic differences in blood pressure and hypertensive complications during pregnancy: the Generation R study.* Hypertension, 2012. 60(1): p. 198-205.
18. Statistics, N. 08/05/2018 [cited 2018 08/24]; Available from: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37422ned/table?ts=1535105738006.
19. Blokstra, A., Smit, H.A., Bueno de Mesquita H.B., Seidell J.C., Verschuren W.M.M. , *Monitoring of risk factors and health in the Netherlands (MORGEN-cohort), 1993–1997*, 2005, RIVM: Bilthoven, The Netherlands.
20. Nederlands Huisartsen Genootschap. *Cardiovascular risk management*. 2012; Available from: https://www.nhg.org/standaarden/volledig/cardiovasculair-risicomanagement.
21. Authors/Task Force, M., 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 & Rehabilitation (EACPR).* Eur J Prev Cardiol, 2016. 23(11): p. NP1-NP96.
22. Nederlands Huisartsen Genootschap, *Multidisciplinaire Richtlijn Cardiovasculair Risicomanagement 2011*. Utrecht: Bohn Stafleu van Loghum.
23. van Dis, I., et al., *Evaluation of cardiovascular risk predicted by different SCORE equations: the Netherlands as an example.* Eur J Cardiovasc Prev Rehabil, 2010. 17(2): p. 244-9.
24. Conroy, R.M., et al., *Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project.* Eur Heart J, 2003. 24(11): p. 987-1003.
25. Taylor, F., et al., *Statins for the primary prevention of cardiovascular disease.* Cochrane Database Syst Rev, 2013(1): p. CD004816.
26. *Lipid Modification: Cardiovascular Risk Assessment and the Modification of Blood Lipids for the Primary and Secondary Prevention of Cardiovascular Disease*, 2014, NICE: London.
27. Stam-Slob, M.C., et al., *Cost-effectiveness of PCSK9 inhibition in addition to standard lipid-lowering therapy in patients at high risk for vascular disease.* Int J Cardiol, 2018. 253: p. 148-154.
28. Ara, R., Brazier, J., *Health related quality of life by age, gender and history of cardiovascular disease: results from the Health Survey for England*, 2010, The University of Sheffield.
29. M, M.V., et al., *Dutch Tariff for the Five-Level Version of EQ-5D.* Value Health, 2016. 19(4): p. 343-52.
30. *Guideline for economic evaluations in healthcare* 2016, Health Care Insurance Board: The Netherlands.
31. Wilson, E.C., *A practical guide to value of information analysis.* Pharmacoeconomics, 2015. 33(2): p. 105-21.
32. Steuten, L., et al., *A systematic and critical review of the evolving methods and applications of value of information in academia and practice.* Pharmacoeconomics, 2013. 31(1): p. 25-48.

33. Furu, K., et al., *Selective serotonin reuptake inhibitors and venlafaxine in early pregnancy and risk of birth defects: population based cohort study and sibling design.* BMJ, 2015. 350: p. h1798.

34. Hoch, J.S., A.H. Briggs, and A.R. Willan, *Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis.* Health Econ, 2002. 11(5): p. 415-30.

35. den Ruijter, H.M., et al., *Long-term health benefits and costs of measurement of carotid intima-media thickness in prevention of coronary heart disease.* J Hypertens, 2013. 31(4): p. 782-90.

36. Greving, J.P., et al., *Statin treatment for primary prevention of vascular disease: whom to treat? Cost-effectiveness analysis.* BMJ, 2011. 342: p. d1672.

37. Epstein, D., et al., *Modeling the costs and long-term health benefits of screening the general population for risks of cardiovascular disease: a review of methods used in the literature.* Eur J Health Econ, 2016. 17(8): p. 1041-1053.

38. Ayansina, D., et al., *Long term effects of gestational hypertension and pre-eclampsia on kidney function: Record linkage study.* Pregnancy Hypertens, 2016. 6(4): p. 344-349.

39. Neiger, R., *Long-Term Effects of Pregnancy Complications on Maternal Health: A Review.* J Clin Med, 2017. 6(8).

40. van Giessen, A., et al., *An alternative approach identified optimal risk thresholds for treatment indication: an illustration in coronary heart disease.* J Clin Epidemiol, 2018. 94: p. 122-131.

41. van Rijn, B.B., et al., *Cardiovascular disease risk factors in women with a history of early-onset preeclampsia.* Obstet Gynecol, 2013. 121(5): p. 1040-8.

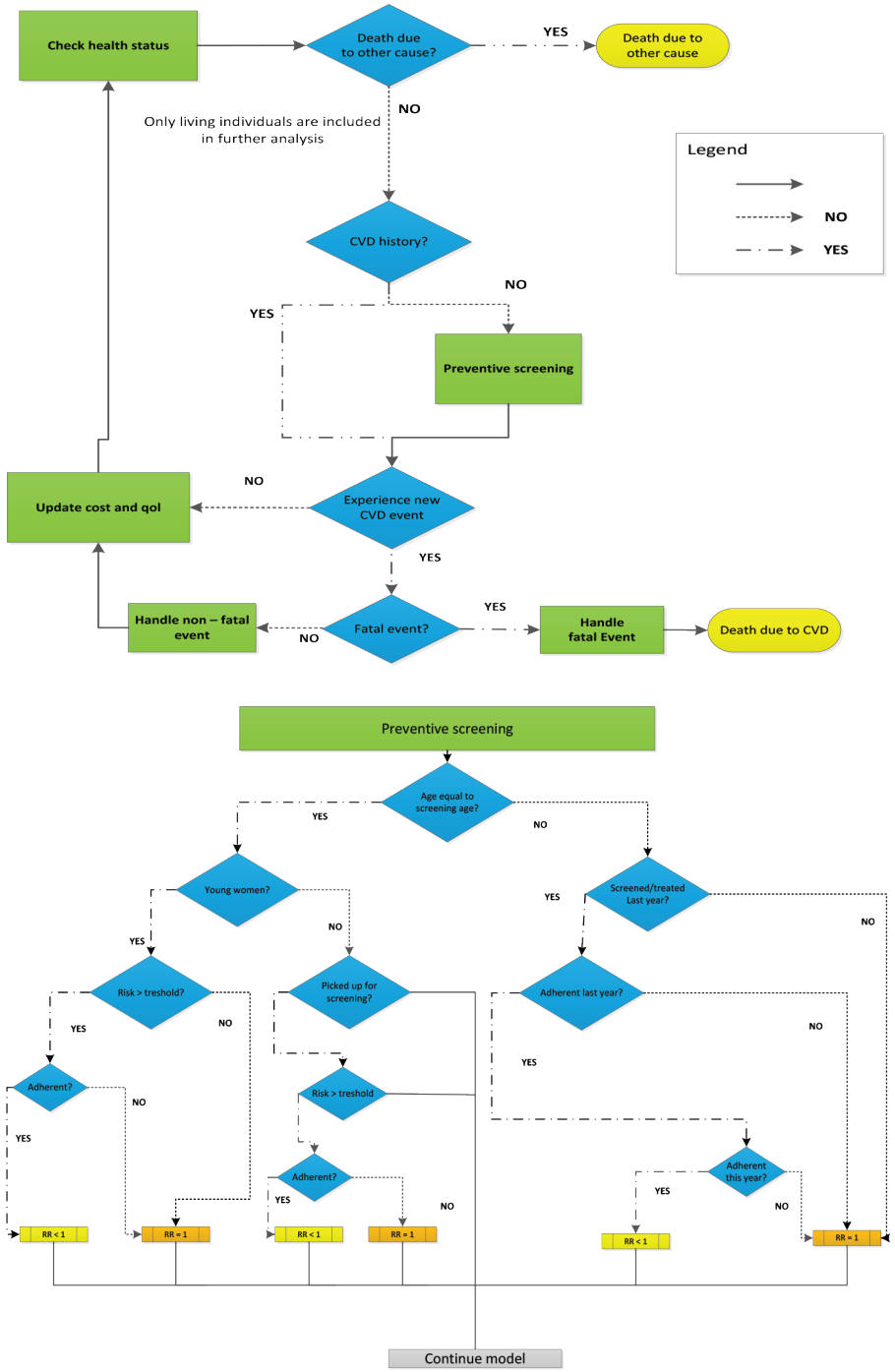
42. van Baaren, G.J., et al., *Cost-effectiveness analysis of cardiovascular risk factor screening in women who experienced hypertensive pregnancy disorders at term.* Pregnancy Hypertens, 2014. 4(4): p. 264-70.

43. Han, B.H., et al., *Effect of Statin Treatment vs Usual Care on Primary Cardiovascular Prevention Among Older Adults: The ALLHAT-LLT Randomized Clinical Trial.* JAMA Intern Med, 2017.

44. [cited 2018 05/01]; Available from: www.medicijnkosten.nl

APPENDICES

A. FLOWCHART OF THE SIMULATION MODEL



B. PARAMETERS

Appendix B - Table 1: Values for the simulation

Uncertain values with unknown range				
Name	**Mean**	**Range (95% CI)**	**Distribution**	**Source**
Marginal correlation between risk profiles (per 10 years)	0.50	0.12 - 0.88	uniform	*Assumption*
Proportion of women below the age of 60 who participate in early preventive screening	0.75	0.61 - 0.89	uniform	*Assumption*
Annual proportion of women above the age of 60 who are detected at the GP	0.05	0.00 - 0.10	uniform	*Assumption*
Probability of women who are adherent to medication after 10 years	0.25	0.01 - 0.50	uniform	*Assumption*
Relative change of adherence rate after 10 years for women who start with preventive screening, i.e. women below the age of 60	1.00	0.52 - 1.48	uniform	*Assumption*
Uncertain values but with a certain range/distribution				
Name	**Mean**	**Range (95% CI)**	**Distribution**	**Source**
Average 10-year CVD risk at age 30	0.02	0.02 – 0.03	uniform	[20, 37]
Average 10-year CVD risk at age 80	0.25	0.20 – 0.30	uniform	[20, 37]
Relative risk ratio after first CVD event (for all years)	2.14	1.71 – 2.68	gamma	[38]
Relative risk after recurrent CVD event (for all years)	2.12	1.70 – 2.63	gamma	[38]
Relative risk reduction of preventive treatment	0.71	0.49 – 0.88	beta	[26]
Mean Costs* (95%)				
Medication				
Standard statin per year (40mg/d)	9	5 – 15	gamma	[39, 40]
Early preventive screening#	143	128 - 161	gamma	[39]
Event				
CHD	5,037	4,985 – 5,085	gamma	[39]
CVA	19,473	19,324 – 19,613	gamma	[39]
Other CVD	2,980	2,921 – 3,040	gamma	[41], range assumption
Recurrent CVD event	1,235	1,198 – 1,274	gamma	[41], range assumption
Fatal CVD event	2,370	2,353 – 2,390	gamma	[42]
Post event – annual				
CHD	761	729 - 798	gamma	[41], [27]
CVA	10,055	9,959 – 10,147	gamma	[39]
Other CVD	3,369	3,304 – 3,439	gamma	[41], range assumption
Recurrent CVD event	686	659 - 715	gamma	[41], range assumption

Mean Utilities (95% CI)				
Event for whole cycle				
CHD	0.77	0.59 – 0.89	beta	[41], range assumption
CVA	0.63	0.46 – 0.78	beta	[41], range assumption
Other CVD	0.69	0.52 – 0.83	beta	[41], range assumption
Recurrent CVD event	0.44	0.29 – 0.61	beta	[41], range assumption
Post-event				
CHD	0.91	0.64 - 0.99	beta	[41], range assumption
CVA	0.63	0.43 - 0.81	beta	[41], range assumption
Other CVD	0.69	0.48 - 0.86	beta	[41], range assumption
Recurrent CVD event	0.66	0.45 - 0.83	beta	[41], range assumption
Statin use	0.998	0.997 – 0.999	uniform	[42]
Parameters reflecting choices rather than uncertainty				
Number of individuals	2,000	-	-	
Discount rate Cost	4%	-	-	[29]
Discount rate Effect	1.5%	-	-	[29]
Quality of life as function of age for general population	0.95698 – 0.00085 * age – 0.00002 * age^2	-	-	[27, 28]
Start age preventive screening – natural history	60	-	-	-
Start age preventive screening (early age)	30, 40, or 50 (with all combinations	-	-	-

* The unit of cost is euro and all costs are updated according to Dutch consumer price indices (2017). # Cost of preventive screening includes costs due to a GP visit, pharmacy and laboratory tests.

Appendix B - Table 2: Age dependent event distribution for women

Age	CHD – non fatal	CVA – non fatal	OCVD – non fatal	CHD – fatal	CVA – fatal	OCVD – fatal
20-30	0.158	0.526	0.263	0.000	0.053	0.000
30-40	0.333	0.333	0.238	0.048	0.048	0.000
40-50	0.475	0.263	0.220	0.008	0.008	0.025
50-60	0.536	0.207	0.197	0.032	0.014	0.015
60-70	0.317	0.307	0.010	0.178	0.030	0.158
70-80	0.326	0.326	0.081	0.140	0.012	0.116
80-90	0.317	0.307	0.010	0.178	0.030	0.158
90-100	0.326	0.326	0.081	0.140	0.012	0.116

C. CVD RISK ESTIMATES

In our micro simulation model, CVD risk estimates are assumed to be beta distributed. From literature, the average 10-year CVD risk and the associated variation at the ages of 30 and 80 years are determined. Using the average (μ) and variance (σ^2) values the corresponding alpha (α) and beta (β) values are calculated; see Appendix C - Equation 1 and Appendix C - Equation 2. This results in one alpha and one beta value for the risk distribution at the age of 30 and another alpha and beta value for the age of 80. The alpha and beta values for the decades between 30 and 80 years are smoothly linear interpolated and stay constant above the age of 80. In total, we have eight "decade" alpha and beta values corresponding with 8 cycles, i.e. 30 until 110.

In the simulation model, risk estimates are correlated between the decades. A single correlation is used to correlate the CVD risk at age X and age X+10. Furthermore, the correlation coefficient is varied in the analyses. This insures that CVD risk is correlated over time (age) but with a diminishing correlation (age X and age X+20). Therefore, a marginal conditional normal distribution is used to simulate the correlation values over time, per woman.

The predicted 10-year CVD risk distribution is divided into intervals where one interval represents one woman. For example, the central interval is a woman of whom the predicted CVD risk estimate is equal to the average CVD risk estimate. With the estimated correlation coefficients for all women and cycles, the intervals are randomly generated for all women.

For each decade, there is an alpha, beta, and interval which are used as input for generating a value, i.e. CVD risk estimate, from a beta distribution. A large interval value for cycle one, i.e. a relatively high predicted risk estimate, together with a high correlation coefficient, e.g. 0.90, likely results in relatively high predicted risk estimates in all subsequent decades. Finally, there are eight 10-year CVD risk estimates for all women. The 10-year CVD risk estimates are converted into 10 annual risk estimates such that the annual risk estimates are increasing over time, see Appendix C - Equation 3.

Appendix C - Figure 1 shows the predicted risk estimates of 25 randomly chosen women over age, i.e. between the ages 30 to 60 years, with three different correlation coefficients. When the correlation coefficient is 0.1, i.e. almost random, the probability of having a relative

low predicted risk after a relative high risk is random, hence the lines intersect. For almost totally correlated risk estimates within women, i.e. lowest plot with a correlation of 0.9, the lines do not intersect. Relatively high risk estimates stay relatively high in this scenario.

Appendix C - Figure 2 shows the predicted 10-year CVD risk estimates from the ages 30 to 100 years. Here, the influence of the correlation coefficients on the risk estimates is more present and the lines intersect more for a low correlation coefficient compared to a high coefficient.

Appendix C - Table 2 shows the classification of women into two risk categories based on a 2% risk threshold for different correlation coefficients. The number of women who are classified into the same risk category, i.e. low-low and high-risk, at the age of 30 and 40 years is larger when the correlation coefficient is 0.9 compared to a correlation coefficient of 0.1. For example, for a correlation coefficient of 0.1 or 0.9, 1,464 versus 1,406 women have a relatively high predicted risk at both 30 and 40 years.

Example on CVD risk estimation:

Average 10-year CVD risk at 30 years is 0.025 and the standard deviation is 0.01. At the age of 80, the average 10-year CVD risk estimate is 0.25 and the standard deviation is 0.05. The corresponding alphas are 6.07 and 18.5 for the ages of 30 and 80 respectively, and the beta values are 236.7 and 55.5 (Appendix C - Equation 1 and Appendix C - Equation 2). The increase in alpha for one decade is 2.5, i.e. 18.5 minus 6.07 divided by 5, and similarly the decrease in beta value per decade is 36.2.

The average 10-year CVD risks per age decade for a simulation are shown in column 2 (Appendix C - Table 1). With an average correlation coefficient of 0.9, i.e. average correlation over all women with the same age, the relative risk within this 10-year risk distribution for the first simulated woman are shown in column 3 (Appendix C - Table 1). With the 10-year CVD risk distribution per age decade and the relative risk within this distribution, the individualized 10-year CVD risk estimates are generated and shown in column 4 (Appendix C - Table 1). From the 10-year CVD risk estimates, the annual CVD risk estimates can be smoothly interpolated such that the risk estimates increase over time.

Appendix C - Table 1: Correlation coefficients and interval values for all decades between 30 and 100 years, for the first simulated woman.

Age (years)	Average 10-year CVD risk (95% CI)	Relative risk within the 10-year CVD risk distribution	Individualized 10-year CVD risk
30	0.03 (0.01; 0.05)	51%	0.03
40	0.04 (0.02; 0.07)	44%	0.04
50	0.07 (0.03; 0.11)	62%	0.07
60	0.10 (0.06; 0.15)	82%	0.12
70	0.16 (0.09; 0.22)	64%	0.17
80	0.25 (0.16; 0.35)	81%	0.29
90	0.25 (0.16; 0.35)	89%	0.31
100	0.25 (0.16; 0.35)	91%	0.32

$$\alpha = \left(\frac{1-\mu}{\sigma} - \frac{1}{\mu}\right) \cdot \mu^2$$

Appendix C - Equation 1: Formula to estimate alpha value from mean and variance values

$$\beta = \alpha \cdot \left(\frac{1}{\mu} - 1\right)$$

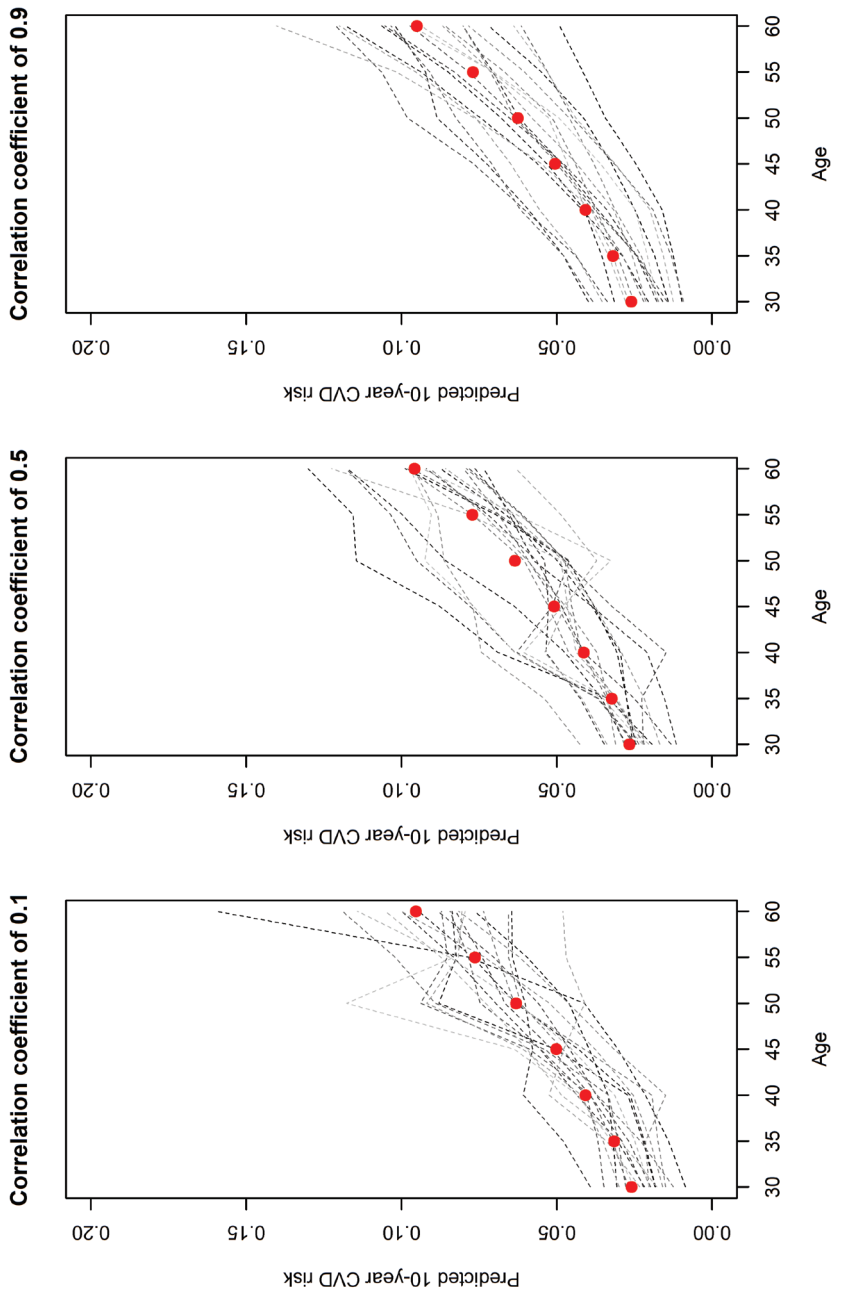
Appendix C - Equation 2: Formula to estimate beta values from mean and variance values

$$r_{1y} = \sqrt[10]{(1 + r_{10y})} - 1$$

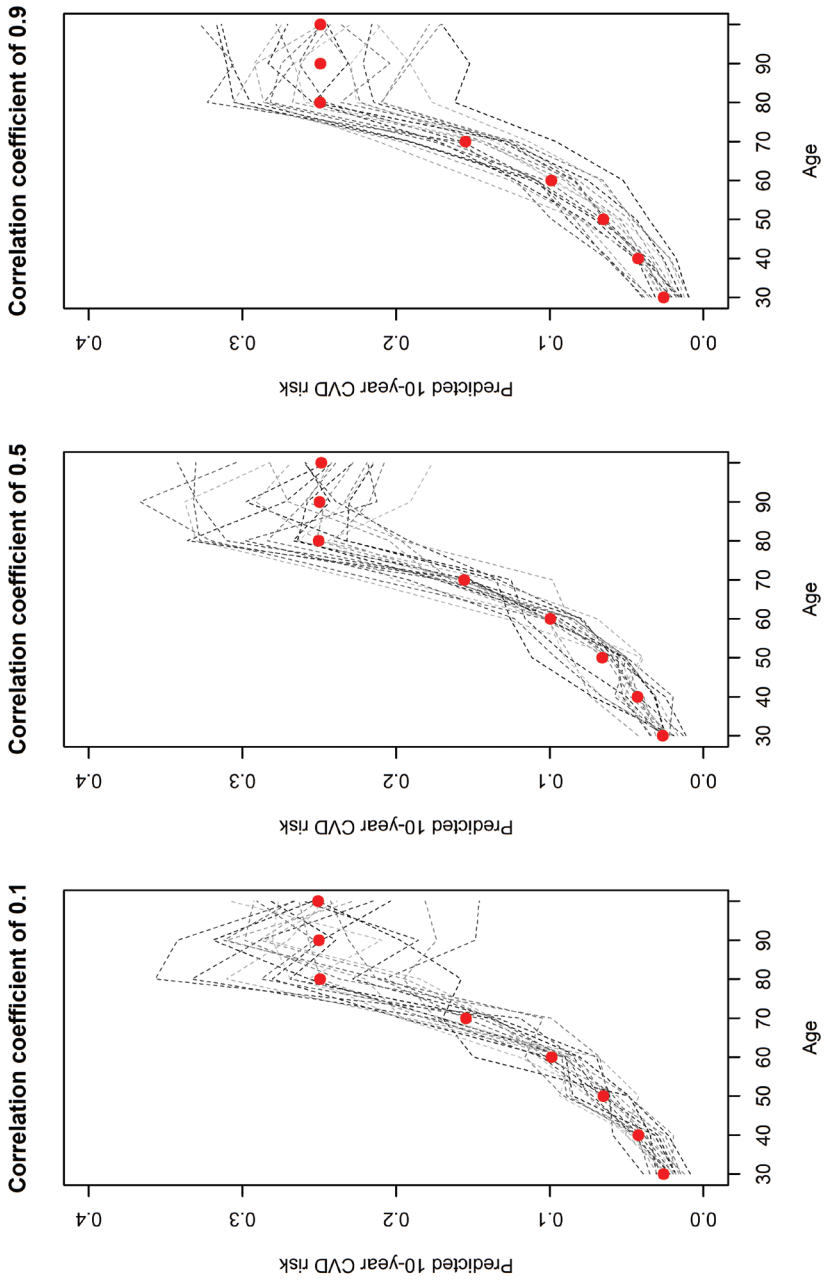
Appendix C - Equation 3: Estimation of the annual CVD risk estimate from the 10-year CVD risk estimate

Appendix C - Table 2: Classification table for 10-year CVD risk estimates at 30 and 40 years with an absolute risk threshold of 2%.

10-year CVD risk at 30 years (average 0.026; 95% CI 0.013; 0.051) \ 10-year CVD risk at 40 years (average 0.043; 95% CI 0.023; 0.072)	Low	High
	Correlation coefficient of 0.1	
Low (N=530)	3 (0.6%)	527 (99.4%)
High (N=1470)	6 (0.4%)	1,464 (99.6%)
	Correlation coefficient of 0.5	
Low (N=530)	22 (4.2%)	508 (95.8%)
High (N=1470)	5 (0.3%)	1,465 (99.7%)
	Correlation coefficient of 0.9	
Low (N=594)	48 (8.1%)	546 (91.9%)
High (N=1406)	0 (0%)	1,406 (100%)



Appendix C - Figure 1: Predicted 10-year CVD risk estimates for three correlation coefficients for 25 women from the age of 30 to 60.



Appendix C - Figure 2: Predicted 10-year CVD risk estimates for three correlation coefficients for 25 women from the age of 30 to 100.

D. IMPACT OF KEY PARAMETERS

The impact of the key parameters, i.e. the parameters with unknown mean and range, is discussed in the next paragraphs. The initial parameter settings for the simulation are shown below.

Initial parameter settings

The total number of simulated women is 2,000, the average predicted 10-year CVD risk at the age of 30 and 80 years is 2.5% and 25% respectively. The correlation coefficient applied to the 10-year CVD risks is 0.50, the annual proportion of women aged 60 year or older with high CVD risk detected at the GP is 10%. In these women, the adherence rate for preventive medication after 10 years is 25%. For the preventive screening, there are five screening moments, the probability to participate in preventive screening is 75% and the 10-year adherence rate in women up to the age of 60 is 25% (i.e. relative change of 1, exactly similar to the 10-year adherence rate in women above the age of 60, see Appendix B - Table 1, row 3-7). The relative risk reduction of taking medication is 25%.

D.1 - Correlation coefficient

Appendix D - Figure 1 shows the relation between the correlation coefficient and the number of CVD events. The number of CVD events is hardly affected by the correlation coefficient which can be explained by the distribution of CVD risk estimates in the simulated cohort. Within the cohort, the predicted risk estimates are distributed differently across women for different values of the correlation coefficient, but the same age-specific predicted risk estimates are present, and 'distributed among all women', regardless of this correlation coefficient.

D.2 - Medication adherence rate in older women

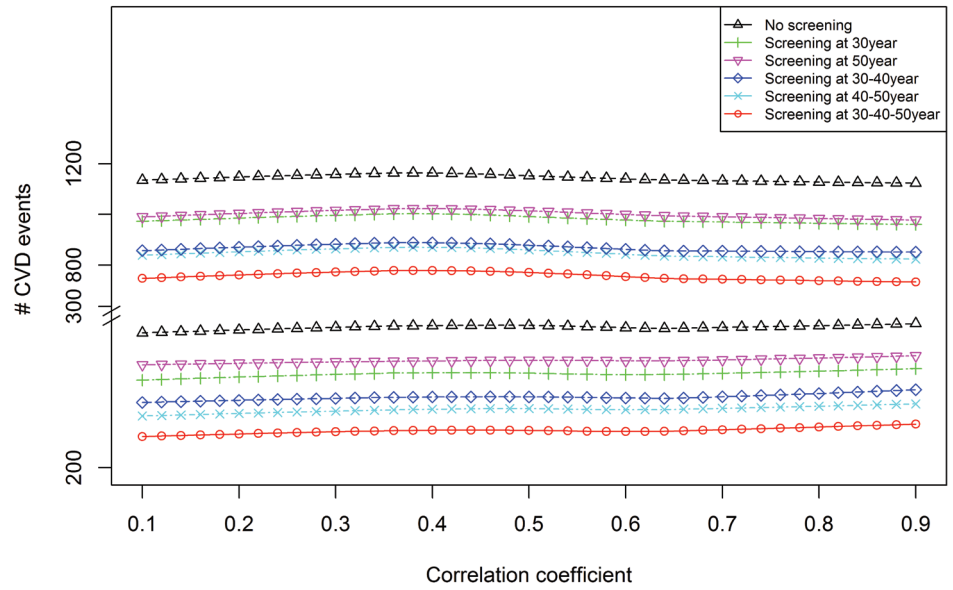
Appendix D - Figure 2 shows the relation between the medication adherence rate and the number of predicted events. A higher adherence rate over 10 years means that more women starting on preventive medication stay adherent resulting in fewer CVD events. Additionally, the preventive strategy plays a role. For a similar adherence rate, the number of CVD events is lower if there are more screening moments and/or the screening age is lower.

D.3 - Adherence rate in young women

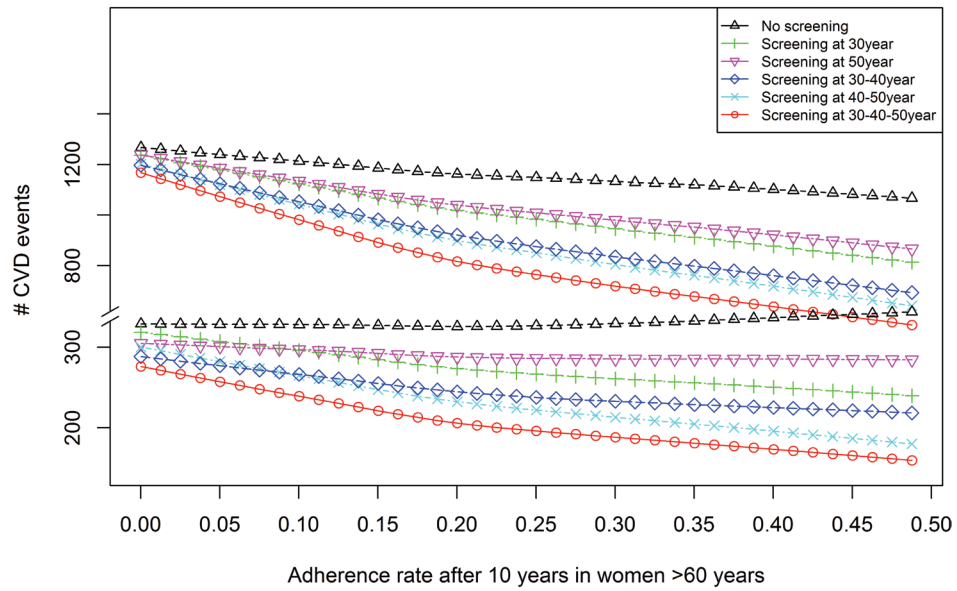
Appendix D - Figure 3 shows the relation between the medication adherence rate of young women, younger than 60 years of age, and the number of CVD events. The number of CVD events decreases when the adherence rate increases; more women adherent to medication results in fewer CVD events. The decrease in number of CVD events is larger when there are more screening moments and/or the screening age is lower.

D.4 - Probability to detect high CVD risk in older women

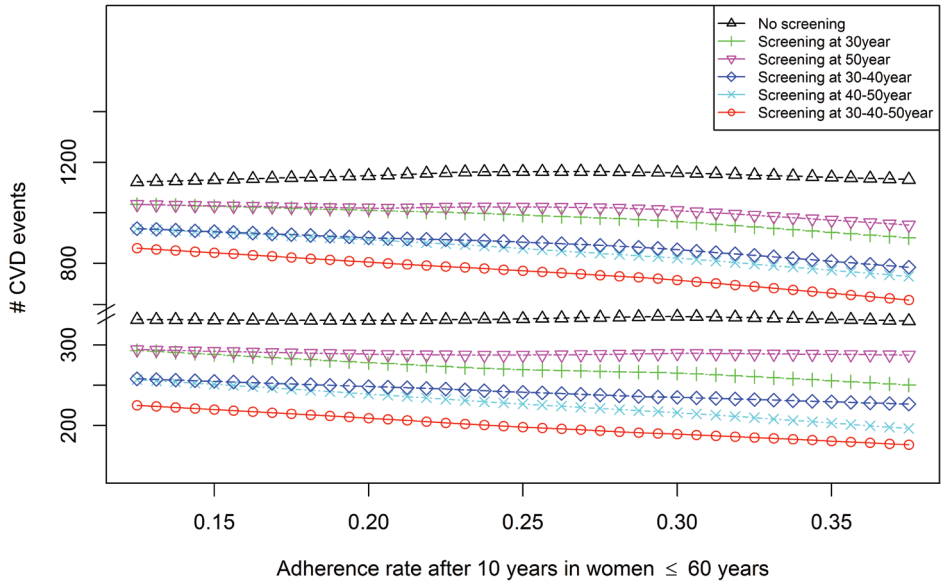
Appendix D - Figure 4 shows the relation between the annual probability of being identified as having high CVD risk in women above the age of 60 and the number of CVD events. The number of CVD events decreases if this probability increases from 0 to 0.02. However, the number of events stays almost constant for a probability between 0.02 and 0.1. Additionally, the preventive strategy influences the relation between this annual probability to be detected and number of CVD events. For all five preventive strategies, the number of CVD events decreases if the annual probability increases up to 0.02 and remains constant for higher probabilities.



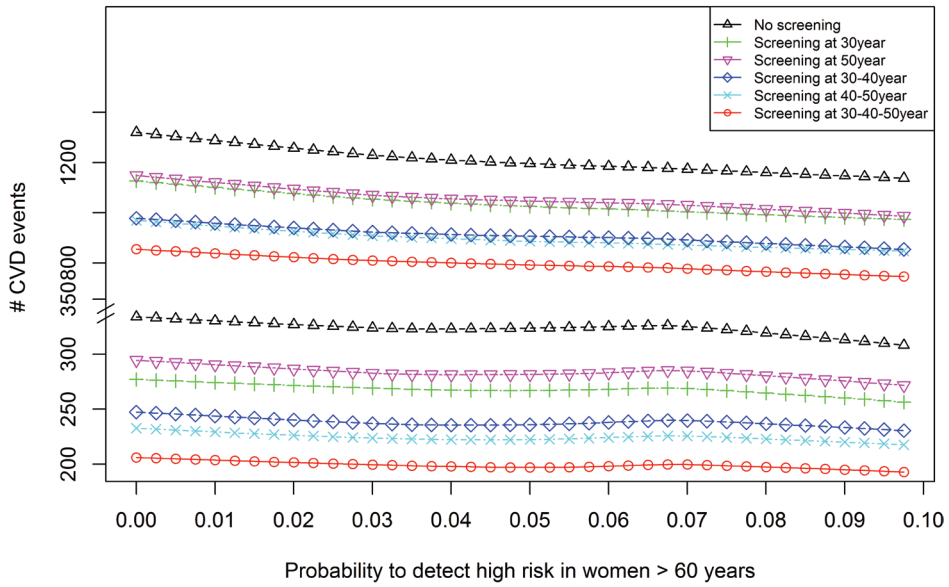
Appendix D - Figure 1: Effect of unknown correlation on the number of CVD events.
A line in the lower part presents the number of CVD events in women below the age of 60, and a line in the upper part presents the number of predicted events over the entire life span of women with HPD.



Appendix D - Figure 2: Effect of uncertain adherence rate in older women on the number of CVD events.
A line in the lower part presents the number of CVD events in women below the age of 60, and a line in the upper part presents the number of predicted events over the entire life span of women with HPD.

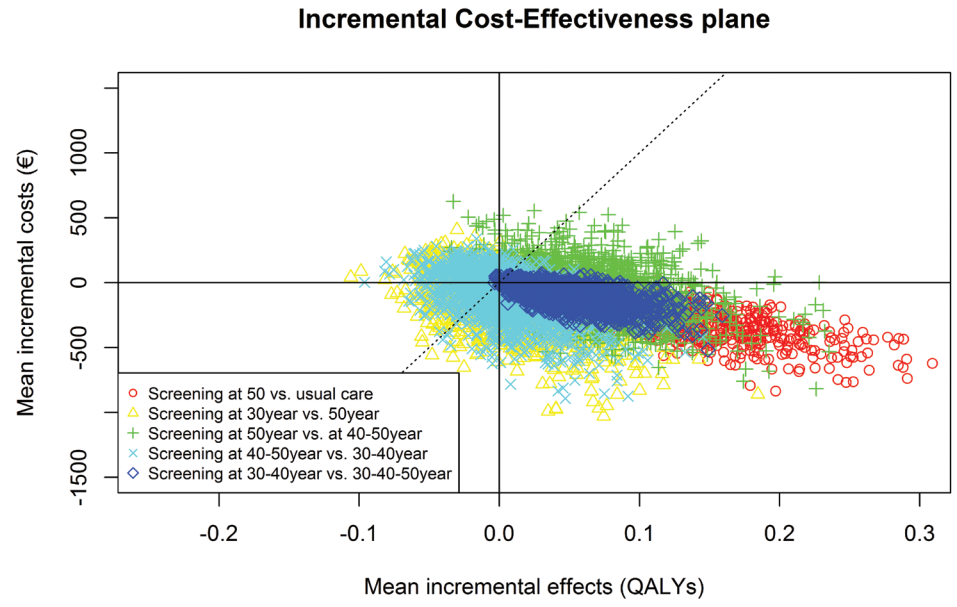


Appendix D - Figure 3: Effect of uncertain adherence rate in women below the age of 60 on the number of CVD events.
A line in the lower part presents the number of CVD events in women below the age of 60, and a line in the upper part presents the number of predicted events over the entire life span of women with HPD.



Appendix D - Figure 4: Effect of unknown probability to detect high CVD risk in women above the age of 60 on the number of CVD events.
A line in the lower part presents the number of CVD events in women below the age of 60, and a line in the upper part presents the number of predicted events over the entire life span of women with HPD.

E. COST EFFECTIVENESS ANALYSES



Appendix E - Figure 1: Incremental cost-effectiveness plane

Appendix E - Table 1: Results from the PSA with the incremental results for the preventive screening strategies

	Average Costs (€)	Average benefits (QALY)	Difference in costs* (€)	Difference in benefits* (QALY)	Incremental cost (€)	Incremental benefits (QALY)	ICER (€/QALY)	NHB
50 years	6,223	29.80	-173	0.08				0.09
30 years	6,115	29.80	-281	0.09	-108	0.00	Dominated	0.10
40 and 50 years	6,044	29.86	-352	0.15	-719	0.06	Dominated	0.16
30 and 40 years	5,950	29.86	-446	0.15	-94	0.00	Dominated	0.17
30, 40, and 50 years	5,852	29.91	-544	0.20	-197	0.05	Dominated	0.23

* Difference is defined as preventive screening compared to no screening

F. VALUE OF INFORMATION

Appendix F - Table 1: Value of information results

	Per Person EVPPI (€)	Standard Error	Indexed to Overall EVPI = 1.00	EVPPI for the Netherlands Per Year (€)	EVPPI for the Netherlands over 10 years (€)
Single parameter EVPPI					
Relative Risk Reduction	*0.22*	*0.70*	*0.05*	*443*	*4429*
Young risk	0.00	0.00	0	0.00	0.0
Old risk	0.00	0.00	0	0.00	0.0
Correlation	0.00	0.00	0	0.00	0.0
Adherence older women	0.00	0.00	0	0.00	0.0
Adherence younger women	0.00	0.00	0	0.00	0.0
Cost screen	0.00	0.00	0	0.00	0.0
Cost medication	0.00	0.00	0	0.00	0.0
Utility medication	0.00	0.00	0	0.00	0.0
Cost CHD - first year	0.00	0.00	0	0.00	0.0
Cost CHD - sequential year	0.00	0.00	0	0.00	0.0
Cost CVA - first year	0.00	0.00	0	0.00	0.0
Cost CVA - sequential year	0.00	0.00	0	0.00	0.0
Cost Other CVD - first year	0.00	0.00	0	0.00	0.0
Cost Other CVD - sequential year	0.00	0.00	0	0.00	0.0
Cost Recurrent event - first year	0.00	0.00	0	0.00	0.0
Cost Recurrent event - sequential year	0.00	0.00	0	0.00	0.0
Cost CVD death.	0.00	0.00	0	0.00	0.0
Utility CHD - first year	0.00	0.00	0	0.00	0.0
Utility CHD - sequential year	0.00	0.00	0	0.00	0.0
Utility CVA - first year	0.00	0.00	0	0.00	0.0
Utility CVA - sequential year	0.00	0.00	0	0.00	0.0
Utility OCVD - first year	0.00	0.00	0	0.00	0.0
Utility OCVD - sequential year	0.00	0.00	0	0.00	0.0
Utility Recurrent event - first year	0.00	0.00	0	0.00	0.0
Utility Recurrent event - sequential year	0.00	0.00	0	0.00	0.0
Proportion participate	0.00	0.00	0	0.00	0.0
Annual probability detect	0.00	0.00	0	0.00	0.0
Relative risk – first year	0.00	0.00	0	0.00	0.0
Relative risk – sequential year	0.00	0.00	0	0.00	0.0

	Per Person EVPPI (€)	Standard Error	Indexed to Overall EVPI = 1.00	EVPPI for the Netherlands Per Year (€)	EVPPI for the Netherlands over 10 years (€)
Group parameter EVPPI					
Set 1 – Predicted CVD risk	0.00	0.00	0.000	0.000	0.00
Set 2 – Probability to start preventive medication and stay adherent	0.00	0.36	0.001	3053.261	30532.61
Set 3 – Screening and treatment	*0.71*	*1.04*	*0.15*	*1416*	*14158*
Set 4 – Costs	0.00	1.56	0.00	0.00	0.00
Set 5 – Utilities	0.00	0.41	0.00	0.00	0.00
Set 6 – Relative risk after CVD event	0.00	0.00	0.00	0.00	0.00

6.

Impact of preventive screening and lifestyle interventions in women with a history of preeclampsia: micro-simulation model study

G.R. Lagerweij
L. Brouwers
G.A. de Wit
K.G.M. Moons
L. Benschop
A.H.E.M. Maas
A. Franx
J.E. Roeters van Lennep
B.B. van Rijn
H. Koffijberg
on behalf of the CREW consortium.

In preparation

ABSTRACT

Background: Preeclampsia is a female-specific risk factor for the development of future cardiovascular disease (CVD). Whether early preventive CVD risk screening combined with risk-based lifestyle interventions in women with previous preeclampsia are beneficial and cost-effective is unknown.

Methods: A micro-simulation model was developed to assess the life-long impact of preventive cardiovascular screening strategies initiated after women experienced preeclampsia. Screening was started at the age of 30 or 40 and was repeated every 5 years. 10-year CVD risk estimates were calculated according to Framingham Risk Score and multiple absolute risk thresholds (2% and 5%) were evaluated for treatment selection, i.e. lifestyle interventions (including smoking cessation, weight reduction, increasing physical activity). Screening benefits were assessed in terms of costs and quality-adjusted-life-years (QALYs), and incremental cost-effectiveness ratios (ICERs) compared to no screening. Probabilistic sensitivity analysis was performed by Monte-Carlo simulation and a Value of Information analysis identified parameters that warrant further research.

Results: Expected health outcomes for no screening are 27.35 QALYs and increase to 27.41 QALYs (screening at 40 years, both thresholds), and to 27.42 QALYs and 27.43 QALYs (screening at 30 years with 2% and 5% threshold respectively). The expected costs for no screening are €9,426 and around €11,600 for screening at 40 years (for both thresholds) and €13,078 and €13,881 for screening at 30 years (for a 5% and 2% threshold respectively). Preventive screening at 40 years with a 2% threshold has the most favourable ICER, i.e. €34,996/QALY, compared with other preventive screening scenarios and no screening. Probabilistic sensitivity analysis shows that no screening has the largest probability to be cost-effective up to a WTP threshold of €57,000/QALY (95% for a WTP of €0/QALY to 27% for a WTP of €57,000/QALY) Parameters associated with predicted CVD risk have a large impact on the cost-effectiveness of screening scenarios; further evidence collection would be merited for these parameters.

Conclusions: Early CVD risk screening followed by risk-based lifestyle interventions can improve long term health outcomes in women with a history of preeclampsia. However, the cost-effectiveness of establishing a lifelong cardiovascular prevention program for women starting early after experiencing preeclampsia by risk-based lifestyle advice alone is relatively unfavourable. A combination of risk-based lifestyle advice plus medical therapy may be more beneficial.

1. INTRODUCTION

Cardiovascular disease (CVD) is the most prevalent cause of death in women worldwide and is predominantly caused by long term progression of atherosclerosis [1]. The global burden of CVD is associated with lifelong exposure to traditional risk factors, such as hypertension, hypercholesterolemia, obesity, smoking and type II diabetes mellitus and is strongly associated with a prolonged unhealthy lifestyle [1, 2]. It has been estimated that up to 90% of CVD risk can be explained through traditional, and modifiable, risk factors [3]. Over the past decades, long-term population studies have identified additional female-specific risk factors. Preeclampsia is one of the strongest female-specific risk factors for CVD, associated with a two- to seven fold increased risk of developing ischemic heart disease and stroke compared to women with normotensive pregnancies [4-8].

Several international obstetric guidelines recommend screening for cardiovascular risk profiles of women who have a history of preeclampsia, These guidelines classically start cardiovascular screening in women at the age of 50 [9-11]. However, the recommendations are not yet implemented in the leading cardiovascular prevention guidelines [12-15]. Additionally, treatment recommendations are based on risk prediction models that calculate 10-year CVD risk and are strongly age-dependent. Shortly after pregnancy, women will not usually reach the current risk threshold for preventive measures recommended by these guidelines. For example, in women with mean age of 31 (SD 4.5 years), the average 10-year CVD risk according to the Framingham Global Risk score is 1.08% (95% CI of 1.04-1.12%) whereas the recommended risk threshold is 10% [16]. Current risk-based selection may therefore not be appropriate for these young women at *relatively high* but *low absolute* risk and a lifetime CVD risk-based approach may be preferable [17].

As the timeline during which benefits from preventive intervention in young women accrue is long, a randomized or cohort setting is not feasible to assess the full benefits of prevention. Here, a model-based approach is valuable, even though collecting the required evidence is challenging. Two Dutch Markov model-based studies previously showed that early CVD prevention in women with previous preeclampsia is likely to be cost-effective [18, 19]. However, authentic long-term follow up data from cardiovascular screening including multiple cardiovascular risk factors measures, e.g. blood pressure, weight, height, and blood samples, for each participant were not available at the time these studies were performed. Furthermore, previous studies

used a cohort model that is not able to include treatment decisions on an individual level, which is likely to give a less realistic representation of clinical practice.

We present a model-based patient-level simulation (i.e. micro-simulation) of early cardiovascular risk screening combined with risk-based lifestyle interventions to assess health benefits, costs, and cost-effectiveness in women with a history of preeclampsia. We incorporated individual patient data on cardiovascular risk factor measures, e.g. blood pressure level, cholesterol level and smoking status, of an initial cardiovascular screening six months after delivery in women with preeclampsia and of screening after 10-20 years follow up to estimate 10-year CVD risks. A life-long horizon was applied to capture all benefits of screening and subsequent lifestyle interventions in these women.

2. METHODS

To assess the impact of early CVD preventive screening strategies, datasets from two studies in the Netherlands were combined. Both studies measured cardiovascular risk parameters at different time intervals after preeclampsia.

The first dataset comprised initial cardiovascular screening performed in 349 women six months after a first pregnancy was complicated by early-onset preeclampsia (mean age 30.8, 95% range 22.0-39.6). The complete study design has been previously published [16]. In short, this study was performed between 1994 and 2007 and recorded BMI, blood pressure and fasting blood lipid and glucose levels, as well as the presence of diabetes and chronic hypertension.

Secondly, we used data from the CREw-IMAGO (Cardiovascular Risk Profile: Imaging and Gender-Specific Disorders) study where women were screened for cardiovascular parameters 10-20 years after pregnancy complicated by (early and late onset) preeclampsia (n=291) [20]. Women from the first study were also invited to participate in the CREw-IMAGO study. The complete design of the CREw-IMAGO study was published previously [5, 20]. In short, asymptomatic women, aged 40 to 63 (mean age 46.4, 95% range 40.2-57.8), with a history of preeclampsia were assessed for cardiovascular risk factors including BMI, waist circumference, blood pressure, lipid and glucose levels [16, 20-22]. In total, 49 women were included in both studies with a mean age of 33.2 at recruitment of the initial post-partum screening (95% range 25.7- 40.3).

2.1 Model development and parameters

A discrete time micro-simulation model was developed to assess the impact of early preventive strategies for CVD [23]. The flowchart of this model is presented in Appendix A. The time cycle of the model was one year. Women were followed until death and outcomes were aggregated at population level, i.e. total CVD events, total costs and health outcomes, expressed in Quality-Adjusted-Life-Years (QALYs). Appendix B shows an overview of all input parameters that were used in the analysis. Estimates for model parameters were based on evidence from the literature and partially on expert opinion and consensus. Relatively wide distributions were used to properly reflect any parameter uncertainty.

In total, we simulated a hypothetical cohort of 2,000 women as the incidence of early-onset preeclampsia is currently about 1-2% amongst a total of approximately 171.000 annual pregnancies in the Netherlands [24, 25]. Women entered the simulation model at the average age of a first pregnancy in the Netherlands (i.e. 30 years old) [24].

2.2 CVD risk estimates

As CVD risk estimates vary with age, we assumed that CVD risk increased over time for each woman. Published long-term data on the development of risk factors was not available for this specific group of women with previous preeclampsia. Therefore, we used 10-year CVD risk estimates from the two cohorts.

The Framingham Global Risk Score (FRS) was used to estimate 10-year CVD risk at initial post-partum screening and at follow-up (CREw-IMAGO study) [26]. Multiple imputation (with 10 datasets) was performed to handle missing predictor data using the MICE package in R [27]. Imputation was based on all other available patient characteristics, such as age, sex, blood pressure, and cholesterol levels.

Estimated CVD risk estimates and follow-up time were not the same for all women in the two cohorts due to differences in age at screening in both studies. To correct for this, 10-year CVD risk estimates were recalculated to risk estimates at the same age. First, by using the data of the 49 women with two risk estimates, it was possible to calculate annual change in CVD risk. However, using the absolute difference, i.e. linear change, was not possible because; a) risk estimates could become negative, and b) risk estimates were based on a power function rather

than a linear function. Therefore, the individual relative change in CVD risk was calculated for each of these individuals based on the two measures at the corresponding ages. Second, a statistical distribution was defined based on the observed annual individual relative change values of these 49 women. Finally, this distribution was used to draw random values as plausible annual relative changes in the women (from the initial post-partum screening and the CREw-IMAGO study) who were screened only once. For women from the post-partum screening, CVD risk estimates were re-estimated to the age of 30 and 40, and for the CREw-IMAGO study to the age of 40 and 50. For an example of this recalculation, see Appendix C. Based on the 10-year (recalculated) CVD risk estimates at 30, 40, and 50 years, beta distributions were defined and used to draw random values (i.e. 2,000 simulated women) as plausible 10-year CVD risk estimates for three age-decades. Furthermore, individual annual CVD risks were calculated by interpolation (for more details, see appendix C). The 10-year risk estimate at the age of 80 was determined by expert opinion and Dutch prevalence data [28]. The risk estimates between the age of 50 and 80 were then interpolated and after the age of 80, it was assumed that CVD risks stayed constant. To correlate the risk values for each woman over time, a single correlation estimate was estimated based on the data of 49 women with two measures and used to correlate risk profiles over time within each woman (for more details, see appendix C).

2.3 Usual care

Despite a national multidisciplinary guideline recommending that women who experienced preeclampsia should be offered CVD screening by their general practitioner at the age of 50, no nationwide primary prevention program is currently offered in the Netherlands [14]. We therefore assumed usual care for these women as follows. We presumed that annually 3% (range 2-4%) of all women above the age of 60 would get a cardiovascular screening and could then be identified as high risk. Usual care applied a risk threshold of 10% (Framingham Risk Score) to classify women as high risk [29, 30]. Lifestyle interventions (including smoking cessation, weight reduction, increasing physical activity) were recommended to high risk women as preventive intervention. Medication was not used as preventive intervention due to the young age of the women. For those women adhering to these lifestyle change, we used a risk reduction (average 0.91, range 0.84-0.96) in the model [31]. Finally, because evidence on long-term adherence rates was not available, we assumed that on average 20% of women stayed adherent up to 10 years after initiation of the intervention and derived the annual adherence rate through exponential interpolation.

2.4 Preventive strategy for CVD

The CVD prevention strategies for women after their preeclampsia were defined as cardiovascular risk screening starting at the age of 30 or 40 years, with screening repeated every five years and ending at the age of 55, followed by lifestyle advices based on these risks. As women were young at enrolment in the model, the current recommended generic risk threshold (FRS>10%) was too high, yielding hardly any women in the high risk category. We therefore had to apply lower risk thresholds for the purpose of this study (i.e. FRS>2% and >5%). We used the response rate of the women invited to participate in the CREw-IMAGO study to estimate the proportion that would participate in such a screening programme (39%, range 21-60%). Women who already experienced CVD (with one or more CVD event(s)) were not considered eligible for (primary) preventive screening, but remained in the micro-simulation, potentially experiencing sequential CVD events, until they died. Women who were assessed as low-risk at the previous screening or who did not adhere to the lifestyle changes were invited to the subsequent screening moment(s) after 5 years.

Lifestyle interventions (including smoking cessation, weight reduction, increasing physical activity) were the recommended preventive intervention for women classified as high risk (i.e. FRS>2% and >5%), consistent with usual care. As data on adherence was lacking, we assumed that the relative change that younger women were adherent to lifestyle interventions was equal to the 10-year adherence of 20% in older women (see *usual care*). However, given uncertainty regarding this adherence rate a relative change of 0.9 (lower) to 1.1 (higher) to this 20% adherence was used to define at plausible range of values.

2.5 Model parameters

All model parameters are provided in Appendix B. Three CVD event categories are distinguished in this study; coronary heart disease (CHD), cerebrovascular accident (CVA), and other cardiovascular disease (OCVD) events. The CVD events could be either fatal or non-fatal, resulting in incorporation of six total CVD event types. The relative occurrence of the six event types was age-dependent and based on previous literature (Appendix B - Table 2) [17, 32, 33]. When a cardiovascular event occurred, the CVD risk estimate was proportionally increased (relative risk ratio 2.1, range 1.7-2.6).

Although women may experience other outcomes, (e.g. mental or psychosocial problems) after preeclampsia, little follow-up data are available regarding these long-term outcomes and their effects (and relevance) on quality of life [34, 35]. Therefore, we used quality of life values (utilities) available for women from the general population and adjusted for age [36, 37]. Quality of life (QoL) was proportionally reduced after the occurrence of a CVD event [36, 38-40]. The proportional reduction in QoL after a first CVD event depended on the CVD event type, but remained the same for similar recurrent CVD events. Also, the decrease in utility after a CVD event was lower in the first year compared to consecutive years after the event. It was assumed that women with a CVD event would receive medication. We did not take side-effects of medication after the occurrence of a CVD event into account. Instead, we assumed that these are incorporated in the disutility of CVD events.

Dutch studies and evidence from NICE were used for the estimation of the costs of CVD events [36, 38-40]. Similar to the utilities, costs varied over the six different CVD types. Costs for recurrent events were assumed to be similar to costs for first-time events. Costs of the first year after a CVD event were set higher than costs the subsequent years. Costs of the screening programme included a visit to the general practitioner and laboratory tests, and were applied to all women who participated in the screening programme. Costs of preventive lifestyle interventions were applied to all women who were classified as high risk, i.e. women with CVD risk estimates that exceeded the intervention threshold of 2 and 5%, regardless of their adherence to these lifestyle interventions.

An overview of all utilities and costs together with the distribution for the sensitivity analyses is presented in Appendix B - Table 1 (row 14-44). Following Dutch guidelines, a discount rate of 4% for costs and 1.5% for health outcomes was applied [41]. As preventive screening, CVD events and death due to natural causes can occur at any time during the years (instead of only at the start or end of a year) a half-cycle correction was applied in the model.

2.6 Cost-effectiveness analysis

The cost-effectiveness analysis was performed with the incremental cost-effectiveness ratio (ICER) as outcome, using a health care perspective. This ratio represents the difference in lifetime costs divided by the difference in effectiveness, i.e. health outcomes. The difference in costs and effectiveness is defined as the difference between the four preventive strategy (i.e. screening starting at different age (i.e. 30 and 40 year) and risk levels (i.e. 2% and 5%), with subsequent lifestyle interventions) and usual care. Probabilistic sensitivity analysis was applied to assess how uncertainty in parameter values resulted in uncertainty in the effect and cost outcomes. To determine the differences between strategies, we used 4,500 Monte-Carlo simulations applied to a cohort of 2,000 hypothetical, unique women. Furthermore, the probability of a preventive screening to be cost-effective compared to alternative strategies and usual care was estimated as a function of the WTP and presented in cost-effectiveness-acceptability curves. A commonly applied Dutch Willingness-to-Pay (WTP) threshold of € 20,000 per QALY gained is used to determine whether a screening strategy is cost-effective or not, and to calculate the incremental Net Health Benefits (INHB).

Lastly, we performed a value of information (VOI) analysis to investigate the value of collecting additional information on the used parameters to reduce the uncertainty in cost-effectiveness outcomes. We used the Sheffield Accelerated Value of Information (SAVI) tool to estimate the expected value of perfect information (EVPI) and expected value of partial perfect information (EVPPI) [42]. The value of hypothetically resolving all uncertainty is reflected by the EVPI whereas the EVPPI indicates what the value is of resolving all uncertainty in one parameter or a group of parameters [43].

3. RESULTS

3.1 Study population

Appendix D - Table 1 shows the baseline table of the two cohorts and the number of missing data. Figure 1 shows the authentic risk assessment data of women included in both cardiovascular screening studies (i.e. post-partum and at follow up). The time period between the two real-life screening time points varied significantly within the 49 women who attended both screening moments (average 14.3 years, 95% range 6.6-19.2 years). Similarly, the 10-year CVD risk estimates differ substantially within the group (Figure 1A).

The annual relative risk change within the group was quite similar with an average of 1.06 (95% range 1.01- 1.14). There is a substantial variation in CVD risk estimate at similar ages, indicating substantial heterogeneity in the expected CVD risk in women with preeclampsia (Figure 1B). Furthermore, the overall correlation estimate between the risk profiles at the age of 30 and 40 was large (i.e. 0.855). For example, of all women with a predicted risk above the 75th percentile at the age of 30, 67% also had a predicted risk above the 75th percentile at the age of 40.

Figure 2 shows a histogram of the distribution of the re-estimated 10-year CVD risk evaluations for women in all screening cohorts and the projected distributions. The red line corresponds with the probability density function of the beta distribution estimated based on the re-estimated 10-year CVD risks. Average CVD risk estimates and the 95-percentrange for both cohorts are shown in Appendix D - Table 1 (row 11-12).

3.2 Intermediate outcomes

The reduced number of CVD events resulting from preventive screening and intervention may be the main driver of the expected impact of preventive screening. Table 1 shows the number of CVD events for no screening and the four preventive screening strategies. A screening strategy starting at 30 and with a 5% threshold results in the largest decrease in CVD events, i.e. 3.2% of all life-long CVD events can be prevented by screening every 5 years and starting at 30 years.

Table 2 shows the number of women needed to screen to prevent one CVD event. The number of screened women is equal for both risk thresholds when preventive screening starts at the age of 40. In these two scenarios, all women have a 10-year CVD risk ≥ 2% hence in both scenarios everybody will be selected for preventive treatment. The number needed to screen to prevent one CVD event is largest when preventive screening starts at the age of 30 with a risk threshold of 2%.

Table 3 shows the percentage of women who are classified as high risk at each screenings moment. The percentage of high risk women is almost equal at 40 years, it is almost 40% (i.e. the chosen probability of women that participate in preventive screening). Hence, all women above the age of 40 have a 10-year CVD risk ≥ 2%.

3.3 Cost-effectiveness analysis

Table 4 shows the results of the cost-effectiveness analysis using the chosen risk thresholds of 2% and 5%. No screening has slightly lower health outcomes and costs compared to all four preventive screening scenarios, i.e. 27.35 QALYs and €9,426 per woman. Screening scenarios starting at 40 years have similar health benefits (27.41 QALYs) and the scenario with a 5% threshold has slightly higher costs (€11,578 versus €11,561). Screening starting at 30 with a 2% threshold has slightly lower health effects (27.42 versus 27.43 QALYs) and higher costs (€13,881 versus €13,078) than with a 5% threshold.

When comparing the screening strategies among each other, preventive screening starting at 40 and with a 2% threshold is the 'favourable' preventive screening in terms of the ICER, i.e. €34,996/QALY. Although screening starting at 40 with a 5% threshold is less costly, it has less health benefits resulting in a slightly higher ICER. Therefore, screening starting at 40 with a 5% threshold is dominated by screening starting at 40 with a 2% threshold. However, the latter strategy would not be considered cost-effective if a WTP threshold of €20,000/QALY is applied.

Screening starting at 30 with a 5% threshold is the second 'best' screening strategy in terms of cost-effectiveness; the ICER is €101,092/QALY, compared with screening starting at 40 with a 2% threshold. Screening starting at 30 with a 2% threshold is dominated by preventive screening starting at 30 with a 5% threshold due to similar health benefits but

slightly higher costs. Screening starting at 40 with a 5% threshold and screening starting at 30 with a 2% threshold are therefore dominated by other strategies (and strikethrough in Table 4). Appendix E shows the incremental cost-effectiveness plane for screening starting at 40 with a 5% threshold and screening starting at 30 with a 2% threshold. The PSA samples of the two scenarios are almost similar.

Figure 3 shows the cost-effectiveness acceptability curve. For a WTP threshold of €20,000/QALY, no screening has the largest probability to be cost-effective, i.e. probability of 72%. For a WTP threshold above €57,000/QALY all screening strategies are more likely to be cost-effective than no screening, but no single strategy clearly outperforms the other strategies.

3.4 Value of Information analysis

Table 5 shows the main results from the Value of Information (VOI) analysis. The VOI analysis indicates that the (annual) overall EVPI per person affected is €5,023 per person, or 0.25 QALYs per person. Furthermore, the analysis indicates that the average single parameter 10-year CVD risk at 30 years has the largest EVPI value per person (i.e. €1,568). Further investigation of the uncertainty of single parameters shows that collecting more information on the treatment effectiveness, the average 10-year CVD risk for older women and cost of a CVA event in the first year has added value (see Appendix F - Table 2, row 3 to 31). Given that 2,000 women present with HPD in the Netherlands per year, the population EVPI equals €10.1 million per year.

Appendix F - Table 1 shows an overview of the groups of associated parameters that are used to estimate the group EVPPI. Collecting additional information on parameters related to predicted CVD risk (i.e. set 1), has the largest value with an EVPPI of €1,696 per person (Table 5, row 8-10).

4. DISCUSSION

A model-based simulation study enables to show whether early and long-term preventive CVD risk screening combined with risk-based lifestyle interventions, may indeed reduce CVD burden in women with previous preeclampsia. We found that early (i.e. starting at 30 or 40 years old) and repeated (5-years) CVD risk screening and risk-based lifestyle interventions after preeclampsia potentially reduces CVD risk and improve health outcomes. However, preventive CVD risk screening and risk-based lifestyle intervention alone with an absolute risk threshold of 2% or 5% are not cost-effective. Although different (intervention) strategies were compared, the conclusions of this study are somewhat different with those of the (model) development study described in chapter 4. That study showed that preventive CVD risk screening starting at age of 30, 40, or 50 combined with risk-based preventive medication based on lipid lowering drugs (instead of lifestyle interventions as studied in this chapter) led to improved health outcomes (i.e. mean incremental benefits of 0.20 QALYs) and costs saving (i.e. mean incremental saving of € 543) compared with usual care. Appendix G presents a comparison between the two studies in terms of model parameters, together with a discussion on how two model settings resulted in two different conclusions on the (cost-)effectiveness of preventive screening for CVD.

Strengths

The strength of this study is based on the incorporation of actual risk factor data from women who underwent cardiovascular screening at several time points after preeclampsia. These data gave insight in the risk distribution among women with preeclampsia for different age categories. Furthermore, these data were used to estimate the correlation between 10-year risk estimates *within* women over time. A micro-simulation model was used to assess the long-term benefits from CVD risk screening combined with risk-based lifestyle advices in young women. Using a model with a lifetime horizon is important, as age is a key factor in development of CVD. Moreover, the first manifestation of CVD may take two to four decades following preeclampsia [44]. We postulate that this model, with simple adjustments, can be applied to assess the potential benefits of early CVD risk screening combined with any subsequent risk-based intervention in other populations with (female) specific risk factors, such as women with polycystic ovarian syndrome (PCOS) or premature ovarian insufficiency (POI) [45, 46]. Use of such models may provide information to make evidence-based guidelines and decisions for establishing cardiovascular prevention programs for women with a medical history, while the evidence of intervention studies in these specific female subgroups is still lacking.

Limitations

Our analysis also has limitations. Because evidence on several parameters within the model was lacking, certain assumptions had to be made and extrapolation was required. To properly reflect the uncertainty in the parameters, we allowed relatively wide distributions for most parameters and incorporated expert opinions on behaviour, risks and benefits of interventions in this specific group of women. Also, data on CVD risk after 80 years of age was lacking and this risk was therefore kept constant beyond this age.

Furthermore, we considered only the preventive intervention of lifestyle changes for both young and older women. This approach may not be realistic in clinical practice since lifestyle modification is known to be difficult to achieve and the effectiveness is rather low [31, 47, 48]. Additionally, lifestyle interventions were not combined with any drug therapies, such as lipid lowering or antihypertensive medication, as was done in chapter 4. However, as women were young during the post-partum risk evaluation, the use of life long drug therapy from a young age onwards is perhaps unrealistic anyhow. Nevertheless, some young women may be willing to take medication when becoming aware of their CVD risk after having suffered from preeclampsia. For example, the proportion of women that answered "yes" to the question 'do you have a prescription of preventive medication' in the initial cardiovascular screening is around 19% (see Appendix D). Additionally, a notable proportion has health complaints due to hypertension shortly after pregnancy, making it more likely they would be willing to use medical therapy, even at their young age [49].

Recent CVD preventive guidelines have supported treatment of young individuals even though evidence from randomized or cohort studies for these implementations are not yet available [50]. Taking these possibilities into consideration, the assumption that women in our 'no screening' scenario are not identified, or treated, before the age of 60 may lead to an underestimation of the benefits of usual care in reality. This needs to be evaluated further and may need to be taken in to account when performing similar research in the future.

Also, we estimated CVD risk with FRS, which might not be suitable for young women with previous preeclampsia. Age is a strong contributor to this score and although women with previous preeclampsia develop CVD as soon as 10 years earlier, FRS is often not raised above the indicative 10% threshold soon after pregnancy [9, 10]. Unfortunately, there is no CVD risk score available that includes a (complicated) obstetric history as predictor.

At last, we used data from women with both late and early onset preeclampsia for this study. Although this gives a relevant overview of women with previous preeclampsia, this may underestimate the possible benefits of screening for women with a severe, or early, phenotype. Results from our analysis can therefore also not directly be extrapolated to women with other pregnancy complications (or specific phenotypes of preeclampsia), as the preventive effects are likely to differ in those women. Lastly, we were not able to consider comorbidities or the occurrence of other diseases, like auto-immune disorders or impaired memory, associated with preeclampsia and affecting the outcome and quality of life in these women [51].

Comparison with other studies

Two Dutch studies showed the potential benefits of early hypertension and metabolic syndrome detection, including medication and/or lifestyle intervention, in women with a history of preeclampsia [18, 19]. These studies concluded that CVD prevention in women with preeclampsia is likely to be cost-effective or may save costs without affection quality of life for the first 10-20 years. The opposite conclusions of the current study, i.e. CVD risk screening and subsequent risk-based lifestyle changes are not cost-effective, and the two previous studies may be related to the use of medication as intervention strategy following CVD screening in the two previous studies, as medication is much cheaper than an intervention targeting at a change in eating, drinking, smoking and physical activity habits.

Both previous studies use the level of (systolic and/or diastolic) blood pressure for treatment selection (i.e. ≥ 140/99 mmHg) whereas the treatment selection of our study was risk-based. Furthermore, the published studies used a Markov model with a number of "health states" with fixed transitions between states, whereas we use an individual patient-level model. This provides the opportunity to include CVD risk factors, simulated events and outcomes on an individual level which moves closer to individualized care. For the current study, all individual CVD risk factors were combined in one risk estimate and the change in expected risk was modelled over time. For further research, it is possible to further detail individual risk assessment by also incorporating the assessed CVD risk factor levels per individual. Furthermore, the use of real-world follow-up data of women at 10-20 years post preeclampsia to estimate the CVD risks and subsequent correlation between risk profiles likely has led to more accurate and realistic results, compared to studies making assumptions on risk development over time.

Clinical implications

Although our model estimates that early CVD risk screening and risk-based lifestyle interventions may lead to very small health benefits and is not cost-effective(with the current model settings), some aspects need to be considered for implications in clinical practice. In our experience, offering cardiovascular screening to women after (especially early-onset) preeclampsia results in relatively high percentage of women willing to participate. Unfortunately, the current cardiovascular screening for these women takes place in the hospital which may result in a lower participation rate of these women, i.e. mothers with young children who do not attend the half day of in-hospital screening. Although specific risk factors, such as familiar hypercholesterolemia, should be treated by a vascular specialist, implementation of screening and lifestyle interventions in Dutch primary care would be more efficient. The Dutch GP system is well structured and easily accessible, but such a system may not be available in some other countries.

In addition, our model only included the health effects gained by reducing cardiovascular outcomes. It is likely that the proposed lifestyle interventions (i.e. weight reduction, smoking cessation and improving physical activity patterns) have an additional health benefit in preventing other (non-cardiovascular) health problems, such as preventing joint problems in obese patients and chronic pulmonary problems in smokers. Additionally, the lowering of risk factors will likely reduce risk of other long-term events, such as hypertension and subsequent renal failure, which were currently not incorporated in the model. This may result in further lengthening of life in good health as women age. Taking facts combined, we anticipate that, in a real-world setting, more women would, and could, benefit from early cardiovascular screening and intervention when they have experienced preeclampsia.

Conclusion

Our model-based impact assessment demonstrates that CVD risk screening combined with risk-based lifestyle interventions (without preventive treatment initiation) to prevent CVD in women with a history of preeclampsia is not cost-effective. This study shows that for establishing a beneficial cardiovascular prevention program for women starting early after experiencing preeclampsia, a more effective intervention or combination of interventions may be more realistic.

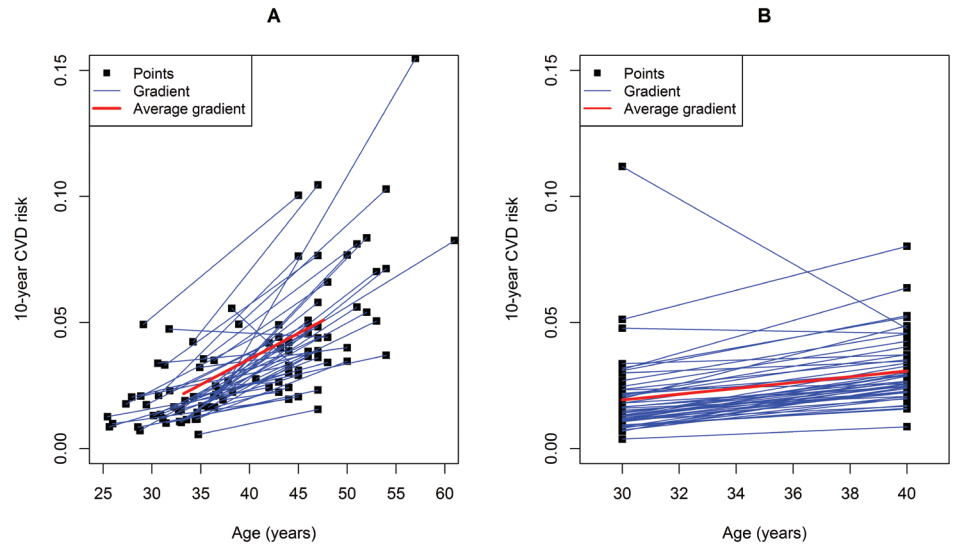


Figure 1: 10-year CVD risk estimates according to FRS for 49 women who participated in two cardiovascular screening studies at different time-points after pregnancy complicated by preeclampsia.
Plot (**A**) shows the authentic 10-year CVD risk estimates for 49 women with two CVD screening moments. Plot (**B**) shows the re-estimated CVD risk assessments to a standardized 10-years' time period.

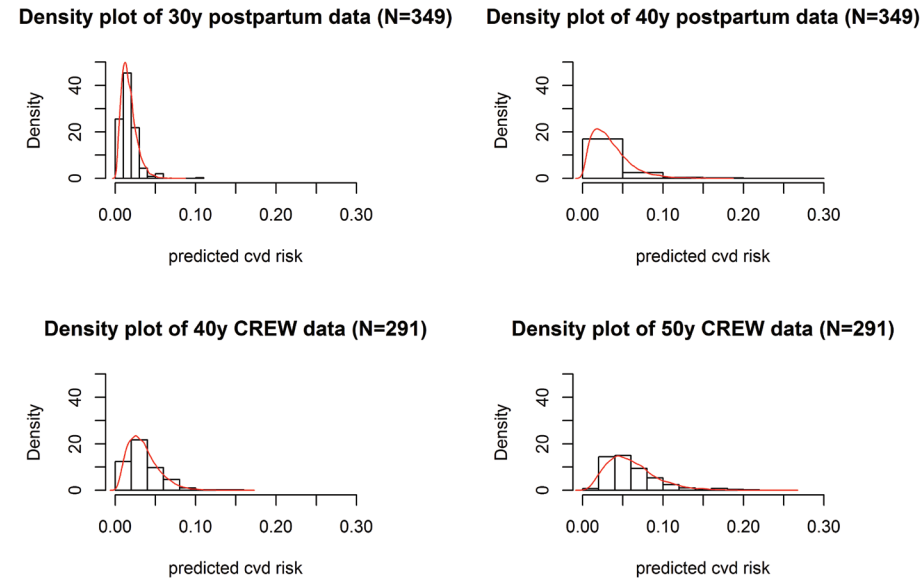


Figure 2: Histograms of the re-estimated 10-year CVD risk evaluations for all women included in both authentic cohorts.
The bars of the histogram represent the density of the 10-year CVD risk estimates of the women included in the cohorts. The red line represents the probability density function of the beta distribution that was based on 10-year CVD (re-estimated) risks of the included women.

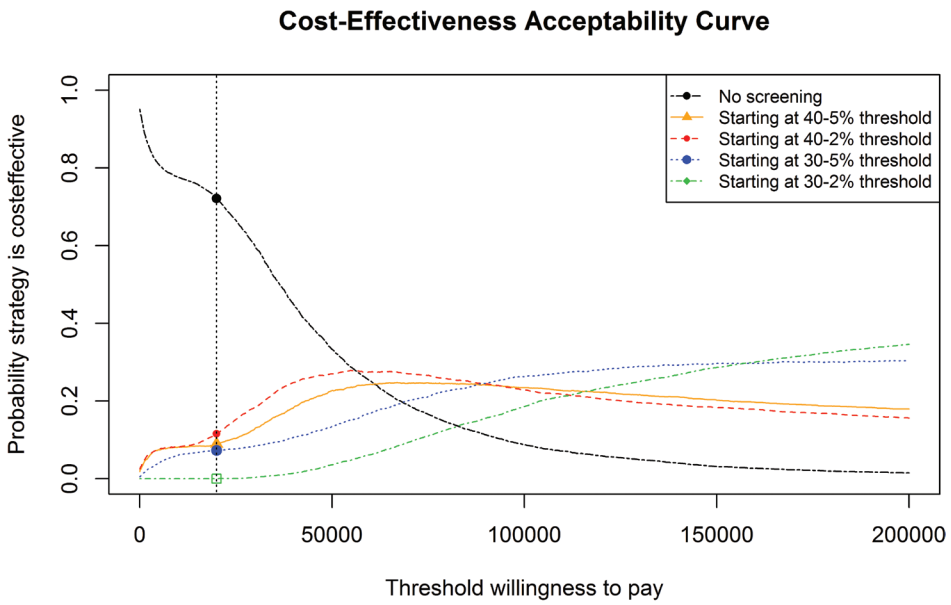


Figure 3: Cost-effectiveness acceptability curves for preventive screening and lifestyle interventions in women with a history of preeclampsia

Table 1: Number of CVD events before the age of 60 in a hypothetical cohort of 2,000 women

No screening	Number of CVD events: N=711	
		Incremental number of CVD events (N,%)
Preventive screening	Screening starting at 40+ 5% threshold	-19 (2.6%)
	Screening starting at 40 + 2% threshold	-19 (2.6%)
	Screening starting at 30 + 5% threshold	-23 (3.2%)
	Screening starting at 30 + 2% threshold	-18 (2.5%)

Table 2: Number of women needed to screen to prevent one CVD event

		Number to screen
Preventive screening	Screening starting at 40 + 5% threshold	134
	Screening starting at 40 + 2% threshold	134
	Screening starting at 30 + 5% threshold	222
	Screening starting at 30 + 2% threshold	186

Table 3: Percentage ^ of women who are classified as high risk

	Preventive screening			
	Screening starting at 40 + 5% threshold	Screening starting at 40 + 2% threshold	Screening starting at 30 + 5% threshold	Screening starting at 30 + 2% threshold
30	-	-	12 (1%)	401 (20%)
35	-	-	269 (13%)	781 (39%)
40	695 (35%)	792 (39%)	695 (35%)	792 (40%)
45	768 (39%)	785 (40%)	768 (39%)	784 (40%)
50	758 (39%)	772 (40%)	758 (39%)	772 (40%)
55	720 (39%)	732 (40%)	720 (39%)	731 (40%)

^: percentage of high risk women is estimated through determining the number of high risk women divided by the number of women alive at that moment

Table 4: Impact of 5-annual cardiovascular screening and lifestyle interventions in women with previous preeclampsia

		Average costs (€)	Average health benefits (QALY)					
No screening		9,426	27.35					
Preventive screening		Average costs (€)	Average health benefits (QALY)	Comparing	Incremental cost (€) (95% CI)	Incremental health benefits (QALY) (95% CI)	ICER (€/QALY)	INHB (95% CI)
~~**A**~~	*Screening starting 40 + 5% threshold*	11,578	27.41	A to 0* (Dominated by B)	~~2,152 (-168;4,394)~~	~~0.06 (-1.57;1.84)~~	~~37,098~~	-0.05
B	*Screening starting 40 + 2% threshold*	11,561	27.41	B to 0*	2,135 (-378;4,405)	0.06 (-1.71;1.80)	34,996	-0.05
~~**C**~~	*Screening starting 30 + 2% threshold*	13,881	27.42	C to B (Dominated by D)	~~2,320 (-325;5,170)~~	~~0.01 (-1.72;1.76)~~	~~210,894~~	-0.15
D	*Screening starting 30 + 5% threshold*	13,078	27.43	D to B	1,526 (-1,134;4,363)	0.01 (-1.68; 1.77)	101,092	-0.11

QALY: Quality-Adjusted-Life-Year; ICER: Incremental Cost-Effectiveness Ratio; INHB: Incremental Net Health Benefit. *; scenario 0 represents usual care

Table 5: Summary of main results from Value of Information analysis (see Appendix F - **Table 1** for the complete set of results)

	Per Person EVPPI (€)	Standard Error	Indexed to Overall EVPI = 1.00	EVPPI for the Netherlands Per Year (€)	EVPPI for the Netherlands over 5 years (€)
Single parameter EVPPI					
Average 10-year CVD risk at age 30	*1,567.6*	129.4	0.3	3,135,000	*15,680,000*
Average 10-year CVD risk at age 80	*259.7*	98.1	0.1	519,400	*2,597,000*
Cost of CVA event (first year)	44.3	40.1	0.0	88,630	443,100
Relative risk of CVD with preventive intervention versus without intervention	21.0	37.1	0.0	41,930	209,700
Utility of other CVD event (sequential year)	12.2	37.0	0.0	24,300	121,500
Group parameter EVPPI					
	Per Person EVPPI (€)	**Standard Error**	**Indexed to Overall EVPI = 1.00**	**EVPPI for the Netherlands Per Year (€)**	**EVPPI for the Netherlands over 5 years (€)**
Set 1 – Predicted CVD risk The following parameters were grouped; • Average 10-year CVD risk at 30 years • Average 10-year CVD risk at 80 years • Marginal correlation between risk profiles (per 10 years)	*1,696.3*	134.3	0.3	3,392,611	*16,963,053*

REFERENCES

1. WHO. *World Health Organization, women's health. Fact sheet N°334.* 2013 September 2013 [cited 2017 July 11th 2017]; Available from: http://www.who.int/mediacentre/factsheets/fs334/en/.
2. WHO. *World Health Organization, Cardiovascular diseases (CVDs). Factsheet* 2017 May 2017 [cited 2017 July 11th 2017]; Available from: http://www.who.int/mediacentre/factsheets/fs317/en/.
3. Yusuf, S., et al., *Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study.* Lancet, 2004. 364(9438): p. 937-52.
4. Mol, B.W.J., et al., *Pre-eclampsia.* Lancet, 2016. 387(10022): p. 999-1011.
5. 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, 2018. 137(8): p. 877-879.
6. Wu, P., et al., *Preeclampsia and Future Cardiovascular Health: A Systematic Review and Meta-Analysis.* Circ Cardiovasc Qual Outcomes, 2017. 10(2).
7. Bellamy, L., et al., *Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis.* BMJ, 2007. 335(7627): p. 974.
8. Brown, M.C., et al., *Cardiovascular disease risk in women with pre-eclampsia: systematic review and meta-analysis.* Eur J Epidemiol, 2013. 28(1): p. 1-19.
9. Behrens, I., et al., *Risk of post-pregnancy hypertension in women with a history of hypertensive disorders of pregnancy: nationwide cohort study.* BMJ, 2017. 358: p. j3078.
10. Engeland, A., et al., *Preeclampsia in pregnancy and later use of antihypertensive drugs.* Eur J Epidemiol, 2015. 30(6): p. 501-8.
11. Sattar, N. and I.A. Greer, *Pregnancy complications and maternal cardiovascular risk: opportunities for intervention and screening?* BMJ, 2002. 325(7356): p. 157-60.
12. Nijdam, M.E., et al., *Cardiovascular risk factor assessment after pre-eclampsia in primary care.* BMC Fam Pract, 2009. 10: p. 77.
13. Bushnell, C. and L. McCullough, *Stroke prevention in women: synopsis of the 2014 American Heart Association/American Stroke Association guideline.* Ann Intern Med, 2014. 160(12): p. 853-7.
14. Heida, K.Y., et al., *Cardiovascular risk management after reproductive and pregnancy-related disorders: A Dutch multidisciplinary evidence-based guideline.* Eur J Prev Cardiol, 2016. 23(17): p. 1863-1879.
15. 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 & Rehabilitation (EACPR).* Eur Heart J, 2016. 37(29): p. 2315-2381.

16. van Rijn, B.B., et al., *Cardiovascular disease risk factors in women with a history of early-onset preeclampsia.* Obstet Gynecol, 2013. 121(5): p. 1040-8.
17. Lagerweij, G.R., et al., *A new selection method to increase the health benefits of CVD prevention strategies.* Eur J Prev Cardiol, 2018. 25(6): p. 642-650.
18. Drost, J.T., et al., *Yearly hypertension screening in women with a history of pre-eclampsia: a cost-effectiveness analysis.* Neth Heart J, 2015. 23(12): p. 585-91.
19. van Baaren, G.J., et al., *Cost-effectiveness analysis of cardiovascular risk factor screening in women who experienced hypertensive pregnancy disorders at term.* Pregnancy Hypertens, 2014. 4(4): p. 264-70.
20. Zoet, G.A., et al., *Cardiovascular RiskprofilE - IMaging and gender-specific disOrders (CREw-IMAGO): rationale and design of a multicenter cohort study.* BMC Womens Health, 2017. 17(1): p. 60.
21. Drost, J.T., et al., *Cardiovascular risk factors in women 10 years post early preeclampsia: the Preeclampsia Risk EValuation in FEMales study (PREVFEM).* Eur J Prev Cardiol, 2012. 19(5): p. 1138-44.
22. Hermes, W., et al., C*ardiovascular risk estimation in women with a history of hypertensive pregnancy disorders at term: a longitudinal follow-up study.* BMC Pregnancy Childbirth, 2013. 13: p. 126.
23. Lagerweij, G.R., de Wit, G.A., van der Schouw, Y.T., Velthuis, B.K.,Fauser, B.C.J.M., Moons, K.G.M., Koffijberg, H., *Assessing the impact of CVD prevention strategies at young age: exploring key parameters and choices,* submitted.
24. Statistics Netherlands, 2018; Available from: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37422ned/table?ts=1535105738006.
25. Bouthoorn, S.H., et al., *Ethnic differences in blood pressure and hypertensive complications during pregnancy: the Generation R study.* Hypertension, 2012. 60(1): p. 198-205.
26. D'Agostino, R.B., Sr., et al., *General cardiovascular risk profile for use in primary care: the Framingham Heart Study.* Circulation, 2008. 117(6): p. 743-53.
27. van Buuren, S. and K. Groothuis-Oudshoorn, *mice: Multivariate Imputation by Chained Equations in R.* 2011, 2011. 45(3): p. 67.
28. RIVM. *Prevalentie hart- en vaatziekten naar leeftijd en geslacht.* 2016 [cited 2018 10-09-2018]; Available from: https://www.volksgezondheidenzorg.info/onderwerp/hart-en-vaatziekten/cijfers-context/huidige-situatie#bronverantwoording.
29. NHG. *Nederlands Huisartsen genootschap. Cardiovascular risk management.* 2012; Available from: https://www.nhg.org/standaarden/volledig/cadriovasculair-risicomanagement.
30. 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, 2011. 57(12): p. 1404-23.

31. Berks, D., et al., *Risk of cardiovascular disease after pre-eclampsia and the effect of lifestyle interventions: a literature-based study.* BJOG, 2013. 120(8): p. 924-31.

32. Beulens, J.W., et al., *Cohort profile: the EPIC-NL study.* Int J Epidemiol, 2010. 39(5): p. 1170-8.

33. Lagerweij, G.R., et al., *Predicted burden could replace predicted risk in preventive strategies for cardiovascular disease.* J Clin Epidemiol, 2018. 93: p. 103-111.

34. Mommersteeg, P.M., et al., *Long-term follow-up of psychosocial distress after early onset preeclampsia: the Preeclampsia Risk EValuation in FEMales cohort study.* J Psychosom Obstet Gynaecol, 2016. 37(3): p. 101-9.

35. Stramrood, C.A., et al., *Posttraumatic stress disorder following preeclampsia and PPROM: a prospective study with 15 months follow-up.* Reprod Sci, 2011. 18(7): p. 645-53.

36. Ara, R.B., J., *Health related quality of life by age, gender and history of cardiovascular disease: results from the Health Survey for England,* 2009, The university of sheffield.

37. Zhou, X., et al., *Precision test for precision medicine: opportunities, challenges and perspectives regarding pre-eclampsia as an intervention window for future cardiovascular disease.* Am J Transl Res, 2016. 8(5): p. 1920-34.

38. in Lipid Modification: *Cardiovascular Risk Assessment and the Modification of Blood Lipids for the Primary and Secondary Prevention of Cardiovascular Disease.* 2014: London.

39. Greving, J.P., et al., *Statin treatment for primary prevention of vascular disease: whom to treat?* Cost-effectiveness analysis. BMJ, 2011. 342: p. d1672.

40. Stam-Slob, M.C., et al., *Cost-effectiveness of PCSK9 inhibition in addition to standard lipid-lowering therapy in patients at high risk for vascular disease.* Int J Cardiol, 2018. 253: p. 148-154.

41. *Guideline for economic evaluations in healthcare* The Netherlands: Health Care Insurance Board 2016.

42. Strong, M., J.E. Oakley, and A. Brennan, *Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach.* Med Decis Making, 2014. 34(3): p. 311-26.

43. Wilson, E.C., *A practical guide to value of information analysis.* Pharmacoeconomics, 2015. 33(2): p. 105-21.

44. Perk, J., et al., *European Guidelines on cardiovascular disease prevention in clinical practice (version 2012). The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts).* Eur Heart J, 2012. 33(13): p. 1635-701.

45. de Groot, P.C., et al., *PCOS, coronary heart disease, stroke and the influence of obesity: a systematic review and meta-analysis.* Hum Reprod Update, 2011. 17(4): p. 495-500.

46. Roeters van Lennep, J.E., et al., *Cardiovascular disease risk in women with premature ovarian insufficiency: A systematic review and meta-analysis.* Eur J Prev Cardiol, 2016. 23(2): p. 178-86.

47. *Gezondheid, leefstijl, zorggebruik en -aanbod, doodsoorzaken; kerncijfers*. 2018 [cited 2018 11/12]; Available from: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/81628NED/table?ts=1542020790751.

48. Paauw, N.D., et al., *Long-term renal and cardiovascular risk after preeclampsia: towards screening and prevention.* Clin Sci (Lond), 2016. 130(4): p. 239-46.

49. Benschop, L., et al., *Blood Pressure Profile 1 Year After Severe Preeclampsia.* Hypertension, 2018. 71(3): p. 491-498.

50. Mancia, G., et al., *2013 ESH/ESC Guidelines for the management of arterial hypertension: the Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC).* J Hypertens, 2013. 31(7): p. 1281-357.

51. Williams, D., *Long-term complications of preeclampsia.* Semin Nephrol, 2011. 31(1): p. 111-22.

52. Han, B.H., et al., *Effect of Statin Treatment vs Usual Care on Primary Cardiovascular Prevention Among Older Adults: The ALLHAT-LLT Randomized Clinical Trial.* JAMA Intern Med, 2017. 177(7): p. 955-965.

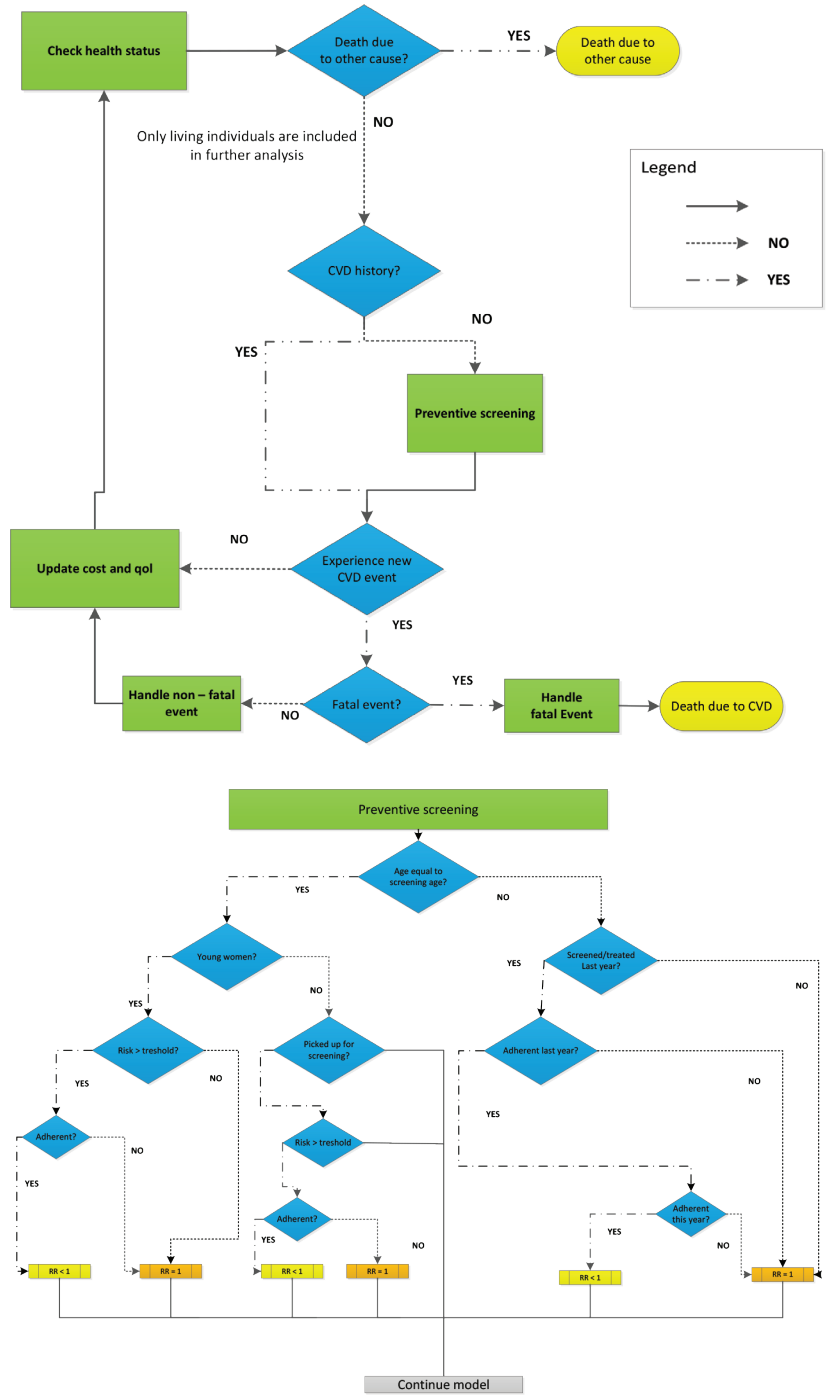
53. Jackson, S.L., et al., *Reduced Cardiovascular Disease Incidence With a National Lifestyle Change Program.* Am J Prev Med, 2017. 52(4): p. 459-468.

54. Groeneveld, I.F., et al., *Cost-effectiveness and cost-benefit of a lifestyle intervention for workers in the construction industry at risk for cardiovascular disease.* J Occup Environ Med, 2011. 53(6): p. 610-7.

55. Lipid Modification: *Cardiovascular Risk Assessment and the Modification of Blood Lipids for the Primary and Secondary Prevention of Cardiovascular Disease,* 2014, NICE: London.

APPENDICES

A. FLOWCHART OF THE SIMULATION MODEL



B. PARAMETERS

Appendix B - Table 1: Parameter estimates and values used for the micro simulation model

Uncertain values with unknown range				
Name	***Value***	***95% CI***	***Distribution***	***Source***
Marginal correlation between risk profiles (per 10 years)	0.88	0.80-0.96	uniform	*Data CREw -UMC*
Proportion of women below the age of 60 who participate in early preventive screening	0.40	0.21-0.59	uniform	Data CREw-UMC
Annual proportion of women above the age of 60 who are detected at the GP	0.03	0.02-0.04	uniform	Expert opinion^
Probability of women who are adherent to lifestyle advice after 10 years	0.20	0.10-0.30	uniform	*Expert opinion^*
Relative change of adherence rate after 10 years for women who start with lifestyle intervention, i.e. women below the age of 60	1.00	0.91-1.10	uniform	*Expert opinion^*
Uncertain values but with a certain range/distribution				
Name	***Value***	***95% CI***	***Distribution***	***Source***
Average 10-year CVD risk at age 30	0.02	0.00-0.04	uniform	*Data CREW - UMC*
Average 10-year CVD risk at age 80	0.85	0.80-0.90	uniform	*Expert opinion^*
Relative risk ratio after first CVD event (for all years)	2.14	1.72-2.66	gamma	[50]
Relative risk ratio after recurrent CVD event (for all years)	2.14	1.72-2.62	gamma	[50]
Relative risk preventive treatment	0.91	0.84-0.96	beta	[31, 51]
Mean Costs†				
Lifestyle intervention	733	688-779	gamma	[52]
Early preventive screening#	143	127-160	gamma	[40]
Event – first year				
Coronary artery disease (CAD)	5,037	4,989-5,087	gamma	[40]
Cerebrovascular accident (CVA)	19,471	19,337-19,606	gamma	[40]
Other CVD (OCVD)	2,982	2,922-3,043	gamma	[38]
Recurrent CVD event	1,235	1,199-1,274	gamma	[38]
Death due to CVD	2,371	2,353-2,390	gamma	[39]
Post event – annual				
Coronary artery disease (CAD)	763	730-796	gamma	[36, 38, 53]
Cerebrovascular accident (CVA)	10,055	9,966-10,144	gamma	[40]
Other CVD (OCVD)	3,369	3,304-3,434	gamma	[38]
Recurrent CVD event	687	657-715	gamma	[38]

Mean Utilities				
Event for whole cycle				
Coronary artery disease (CAD)	0.76	0.72-0.80	beta	[38]
Cerebrovascular accident (CVA)	0.63	0.55-0.70	beta	[38]
Other CVD (OCVD)	0.68	0.54-0.72	beta	[38]
Recurrent CVD event	0.45	0.37-0.53	beta	[38], range assumption
Post-event				
Coronary artery disease (CAD)	0.88	0.85-0.91	beta	[38]
Cerebrovascular accident (CVA)	0.63	0.55-0.71	beta	[38]
Other CVD (OCVD)	0.68	0.65-0.72	beta	[40]
Recurrent CVD event	0.66	0.58-0.74	beta	[40]
Modelling choices				
Number of individuals	2000	-	-	[26, 27]
Discount rate Cost	4%	-	-	[41]
Discount rate Effect	1.5%	-	-	[41]
Decrease quality of life over age	0.95698 -0.00085*Age -0.00002*Age^2		-	[36, 37]

^Expert opinion was formed by 5 of the main authors (GL, LB, AF, AM, BR) † The unit of cost is euro and all costs are updated according to Dutch consumer price indices (2017) and rounded to whole euros. # Cost of preventive screening includes costs due to a GP visit, pharmacy and laboratory tests.
CI, confidence interval; GP, general practitioner; CVD, cardiovascular disease.

Appendix B - Table 2: Age dependent event distribution for women

Age	**CHD – non fatal**	**CVA – non fatal**	**OCVD – non fatal**	**CHD – fatal**	**CVA – fatal**	**OCVD – fatal**
20-30	0.158	0.526	0.263	0.000	0.053	0.000
30-40	0.333	0.333	0.238	0.048	0.048	0.000
40-50	0.475	0.263	0.220	0.008	0.008	0.025
50-60	0.536	0.207	0.197	0.032	0.014	0.015
60-70	0.317	0.307	0.010	0.178	0.030	0.158
70-80	0.326	0.326	0.081	0.140	0.012	0.116
80-90	0.317	0.307	0.010	0.178	0.030	0.158
90-100	0.326	0.326	0.081	0.140	0.012	0.116

CAD, coronary artery disease; CVA, cerebrovascular accident; OCVD, other cardiovascular disease.

C. ESTIMATION OF CVD RISK ESTIMATES

Appendix C - Table 1: Example of recalculation of 10-year CVD risk estimates.

Box	Description	Value	Formula
1	Age at inclusion initial CVD screening (years)	34.27	
2	Age at inclusion in CREw study (years)	54.00	
3	Difference in time (years)	19.73	box 2 – box 1
4	10-year CVD risk at inclusion in initial CVD screening (%)	1.651	
5	10-year CVD risk at inclusion in CREw-IMAGO (%)	3.701	
6	Absolute difference in 10-year CVD risk	2.050	= box 5 – box 4
7	Relative change in 10-year CVD risk for follow years	2.242	=(box 5)/(box 4)
8	Relative annual change in 10-year CVD risk	1.042	= box 7^(1/box3)
9	New 10-year CVD risk at 30 years (recalculated) (%)	1.386	= box 4*((box8) ^(30-box1))
10	New 10-year CVD risk at 40 years (recalculated) (%)	2.087	= box5*((box8)^(40-box2))

In our micro simulation model, CVD risk estimates are assumed to be beta distributed. From literature, the average 10-year CVD risk and the associated variation at the ages of 30 and 80 years are determined. Using the average (μ) and variance (σ^2) values the corresponding alpha (α) and beta (β) values are calculated; see Appendix C - Equation 1 and Appendix C - Equation 2. This results in one alpha and one beta value for the risk distribution at the age of 30 and another alpha and beta value for the age of 80. The alpha and beta values for the decades between 30 and 80 years are smoothly linear interpolated and stay constant above the age of 80. In total, we have eight "decade" alpha and beta values corresponding with 8 cycles, i.e. 30 until 110.

In the simulation model, risk estimates are correlated between the decades. A single correlation is used to correlate the CVD risk at age X and age X+10. Furthermore, the correlation coefficient is varied in the analyses. This insures that CVD risk is correlated over time (age) but with a diminishing correlation (age X and age X+20). Therefore, a marginal conditional normal distribution is used to simulate the correlation values over time, per woman.

The predicted 10-year CVD risk distribution is divided into intervals where one interval represents one woman. For example, the central interval is a woman of whom the predicted CVD risk estimate is equal to the average CVD risk estimate. With the estimated correlation coefficients for all women and cycles, the intervals are randomly generated for all women.

For each decade, there is an alpha, beta, and interval which are used as input for generating a value, i.e. CVD risk estimate, from a beta distribution. A large interval value for cycle one, i.e. a relatively high predicted risk estimate, together with a high correlation coefficient, e.g. 0.90, likely results in relatively high predicted risk estimates in all subsequent decades. Finally, there are eight 10-year CVD risk estimates for all women. The 10-year CVD risk estimates are converted into 10 annual risk estimates such that the annual risk estimates are increasing over time, see Appendix C - Equation 3.

Example on CVD risk estimation:

Average 10-year CVD risk at 30 years is 0.025 and the standard deviation is 0.01. At the age of 80, the average 10-year CVD risk estimate is 0.25 and the standard deviation is 0.05. The corresponding alphas are 6.07 and 18.5 for the ages of 30 and 80 respectively, and the beta values are 236.7 and 55.5 (Appendix C - Equation 1 and Appendix C - Equation 2). The increase in alpha for one decade is 2.5, i.e. 18.5 minus 6.07 divided by 5, and similarly the decrease in beta value per decade is 36.2.

The average 10-year CVD risks per age decade for a simulation are shown in column 2 (Appendix C - Table 2). With an average correlation coefficient of 0.9, i.e. average correlation over all women with the same age, the relative risk within this 10-year risk distribution for the first simulated woman are shown in column 3 (Appendix C - Table 2). With the 10-year CVD risk distribution per age decade and the relative risk within this distribution, the individualized 10-year CVD risk estimates are generated and shown in column 4 (Appendix C - Table 2). From the 10-year CVD risk estimates, the annual CVD risk estimates can be smoothly interpolated such that the risk estimates increase over time.

Appendix C - Table 2: Correlation coefficients and interval values for all decades between 30 and 100 years, for the first simulated woman.

Age (years)	Average 10-year CVD risk (95% CI)	Relative risk within the 10-year CVD risk distribution	Individualized 10-year CVD risk
30	0.03 (0.01; 0.05)	51%	0.03
40	0.04 (0.02; 0.07)	44%	0.04
50	0.07 (0.03; 0.11)	62%	0.07
60	0.10 (0.06; 0.15)	82%	0.12
70	0.16 (0.09; 0.22)	64%	0.17
80	0.25 (0.16; 0.35)	81%	0.29
90	0.25 (0.16; 0.35)	89%	0.31
100	0.25 (0.16; 0.35)	91%	0.32

Appendix C - Equation 1: Formula to estimate alpha value from mean and variance values

$$\alpha = \left(\frac{1-\mu}{\sigma} - \frac{1}{\mu}\right) \cdot \mu^2$$

Appendix C - Equation 2: Formula to estimate beta values from mean and variance values

$$\beta = \alpha \cdot \left(\frac{1}{\mu} - 1\right)$$

Appendix C - Equation 3: Estimation of the annual CVD risk estimate from the 10-year CVD risk estimate

$$r_{1y} = \sqrt[10]{(1 + r_{10y})} - 1$$

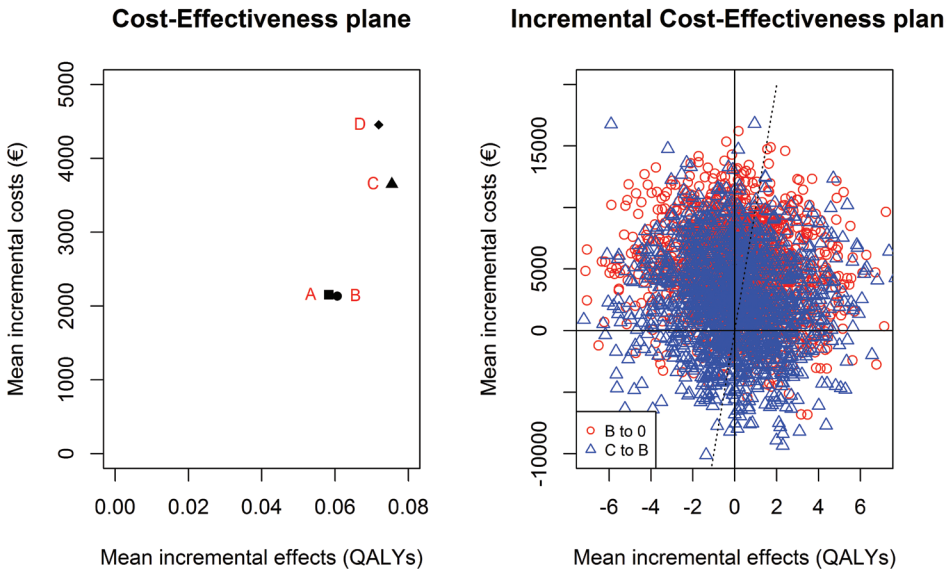
D. BACKGROUND INFORMATION ON DATASETS

Appendix D - Table 1: Baseline characteristics of the initial cardiovascular screening cohort and the CREw-IMAGO cohort

Postpartum			**CREw-IMAGO**		
	Mean (95% CI)	Missing (N, %)		Mean (95% CI)	Missing (N, %)
Number of individuals	349		Number of individuals	291	
Age (years)	30.75 (22 - 39.64)	-	Age (years)	46.69 (40.45 - 57.82)	3 (1%)
Total cholesterol (mg/dL)	196.74 (131.48 - 276.3)	6 (2%)	Total cholesterol (mg/dL)	203.67 (135.34 - 280.84)	35 (12%)
High-density-lipoprotein (mg/dL)	53.94 (30.76 - 86.37)	6 (2%)	High-density-lipoprotein (mg/dL)	58.93 (37.51 - 87.64)	35 (12%)
Treatment	19%	2 (1%)	Treatment	33%	10 (3%)
Systolic blood pressure (mmHg)	125.22 (100 - 150)	143 (41%)	Systolic blood pressure (mmHg)	131.34 (106 - 167.9)	9 (3%)
Diabetes	17%	1 (0%)	Diabetes	6%	14 (5%)
Smoking	1%	2 (1%)	Smoking	3%	12 (4%)
Re-estimated 10-year CVD risk according to FRS at the age of 30 (%)	1.8 (1.0 – 5.0)		Re-estimated 10-year CVD risk according to FRS at the age of 40 (%)	3.5 (1.0 – 8.4)	
Re-estimated 10-year CVD risk according to FRS at the age of 40 (%)	3.4 (1.0 – 10.9)		Re-estimated 10-year CVD risk according to FRS at the age of 50 (%)	6.0 (2.0 – 16.0)	

E. COST EFFECTIVENESS ANALYSIS

Appendix E - Figure 1 shows the average cost-effectiveness plane with the average health effects and costs for the four preventive screening scenarios. Furthermore, Appendix E - Figure 1 shows the incremental cost-effectiveness plane where scenario B (preventive screening starting at 40 years and a 2% threshold) is compared with no screening (scenario 0), and preventive screening starting at 30 years and a 5% threshold (scenario D). The dotted line in the figure is the WTP threshold of €20,000/QALY.



Appendix E - Figure 1: Results of the probabilistic sensitivity analyses (PSA).
The average health benefits and costs are shown in the cost-effectiveness plane for all four screening scenarios (A). The difference in benefits and costs of promising screening scenarios are shown in the incremental cost-effectiveness plane with a WTP threshold of €20,000/QALY (B).
WTP: Willingness-To-Pay; QALY: Quality-Adjusted-Life-Year.

F. VALUE OF INFORMATION

Appendix F - Table 1: Defined sets of parameters for VOI analysis

Set 1 – Predicted CVD risk

- Average 10-year CVD risk at young age, i.e. 30 years
- Average 10-year CVD risk at older age, i.e. 80 years
- Correlation coefficient between 10-year CVD risk estimates

Set 2 – Probability to start preventive intervention and stay adherent

- 10-year adherence for women ≥ 60 years
- Proportion of young women that participate in early preventive screening
- Relative adherence change (in women < 60 years compared with > 60 years)
- Annual proportion of women ≥ 60 years at high CVD risk identified by the GP

Set 3 – Screening and treatment

- Relative risk of CVD when following lifestyle advice versus no lifestyle advice
- Cost of preventive medication
- Cost of early preventive screening

Set 4 – Costs

- Cost of CHD event (first year)
- Cost of CVA event (first year)
- Cost of other CVD event (first year)
- Cost of CVD death
- Cost of CHD event (sequential years)
- Cost of CVA event (sequential years)
- Cost of other CVD event (sequential years)
- Cost of recurrent CVD event (first year)
- Cost of recurrent CVD event (sequential year)

Set 5 – Utilities

- Utility of CHD event (first year)
- Utility of CVA event (first year)
- Utility of other CVD event (first year)
- Utility of CHD event (sequential year)
- Utility of CVA event (sequential year)
- Utility of other CVD event (sequential year)
- Utility of recurrent CVD event (first year)
- Utility of recurrent CVD event (sequential years)

Set 6 – Relative risk after CVD event

- Relative risk of recurrent CVD event (first year)
- Relative risk of recurrent CVD event (sequential year)

Appendix F - Table 2: Value of information results

	Per Person EVPPI (€)	Standard Error	Indexed to Overall EVPI = 1.00	EVPPI for the Netherlands Per Year (€)	EVPPI for the Netherlands over 5 years (€)
Single parameter EVPPI					
Young risk	*1,567.6*	129.4	0.3	3,135,000	*15,680,000*
Old risk	*259.7*	98.1	0.1	519,400	2,597,000
Cost CVA - first year	44.3	40.1	0.0	88,630	443,100
Relative Risk Reduction	21.0	37.1	0.0	41,930	209,700
Utility OCVD - sequential year	12.2	37.0	0.0	24,300	121,500
Proportion participate	9.4	52.5	0.0	18,870	94,360
Cost Other CVD - sequential year	6.7	28.0	0.0	13,430	67,140
Cost screen	3.4	23.9	0.0	6,858	34,290
Cost Other CVD - first year	1.3	25.7	0.0	2,673	13,370
Utility Recurrent event - first year	1.0	38.8	0.0	1,956	9,779
Cost Recurrent event - sequential year	0.5	22.4	0.0	1,067	5,334
Adherence younger women	0.3	39.9	0.0	618	3,088
Cost Recurrent event - first year	0.1	17.4	0.0	218	1,092
Cost CHD - first year	0.0	13.4	0.0	65	325
Relative risk – first year	0.0	13.2	0.0	42	209
Correlation	0.0	20.7	0.0	0	0
Adherence older women	0.0	48.7	0.0	0	0
Cost medication	0.0	10.3	0.0	0	0
Cost CHD - sequential year	0.0	8.9	0.0	0	0
Cost CVA - sequential year	0.0	7.5	0.0	0	0
Cost CVD death,	0.0	6.6	0.0	0	0
Utility CHD - first year	0.0	11.1	0.0	0	0
Utility CHD - sequential year	0.0	11.5	0.0	0	0
Utility CVA - sequential year	0.0	11.8	0.0	0	0
Utility OCVD - first year	0.0	10.1	0.0	0	0
Utility Recurrent event - sequential year	0.0	13.7	0.0	0	0
Annual probability detect	0.0	28.5	0.0	0	0
Relative risk – sequential year	0.0	10.3	0.0	0	0
Relative change in 10-year adherence (young versus older women)	0.0	7.1	0.0	0	0
Group parameter EVPPI					
	Per Person EVPPI (€)	**Standard Error**	**Indexed to Overall EVPI = 1.00**	**EVPPI for the Netherlands Per Year (€)**	**EVPPI for the Netherlands over 5 years (€)**
Set 1 – Predicted CVD risk	*1,696.3*	134.3	0.3	3,392,611	*16,963,053*
Set 2 – Probability to start preventive intervention and stay adherent	142.7	88.7	0.0	285,500	1,427,499
Set 3 – Screening and treatment	71.5	58.1	0.0	143,012	715,062
Set 4 – Costs	242.8	100.8	0.0	485,638	2,428,192
Set 5 – Utilities	296.4	101.8	0.1	592,702	2,963,508
Set 6 – Relative risk after CVD event	12.1	38.7	0.0	24,267	121,334

7.

Optimizing impact studies for risk prediction models: the value of decision analytical modelling

G.R. Lagerweij*
K. Jenniskens*
C.A. Naaktgeboren
L. Hooft
K.G.M. Moons
H. Koffijberg
J.B. Reitsma

*Shared first author

In preparation

ABSTRACT

Introduction: The impact of a prediction model on health outcomes of patients is not only determined by the performance of the model, but also by care providers' compliance with management recommendations based on the predictions. We explored how a decision analytic model (DAM) can be used to quantify the impact of compliance with management recommendations from a prediction model, before conducting a clinical impact study.

Methods: We built a DAM comparing the application of the HEART score, a prediction model used for stratifying patients with chest pain according to their risk of having a serious heart condition, to usual care. The outcome of interest was missed MACE, defined as MACE in patients who were discharged. Only data from the development and external validation of the prediction model, and medical consumption studies were used as input for the DAM. Costs for diagnostic tests, (re)admission days and MACE were included in the assessment. Probabilistic sensitivity analysis was used to assess robustness of impact estimates. Impact on patient outcomes and costs was assessed for scenarios in which the degree of compliance with HEART score management recommendations, and informed deviation (ID) from these recommendations, were varied.

Results: The impact of using the HEART score in a clinical setting is influenced by the interplay of compliance and informed deviation. Scenario analysis showed that a compliance of 100% (with 0% ID) reduced missed MACE compared to usual care. For a compliance of 50% (with 0% ID) there was an increase in missed MACE. When ID was increased to at least 25% (with any compliance above 50%), missed MACE were reduced. Costs per patient reduced as compliance dropped from 100%, and/or ID increased from 0%.

Conclusion: Decision analytic modelling is a useful approach to assess the potential influence of certain factors on the impact of risk prediction model, in case there is limited data available on key factors such as compliance. This approach could provide evidence for deciding whether or not to conduct a subsequent clinical impact study, or to inform the design and conduct of such studies.

1. INTRODUCTION

Diagnostic or prognostic prediction models can be used as decision rules to support management decisions such as subsequent testing or treatment. Demonstrating that a prediction model has adequate predictive performance in (ideally multiple) external validation studies is essential [1-4]. However this does not imply that implementing such a model in practice will improve management decisions by care providers, let alone that it will lead to improved health outcomes and/or reduction in healthcare costs. The impact of prediction models on patient relevant health outcomes and costs can be studied in so-called impact studies, such as comparative longitudinal studies (ideally (cluster) randomized trials), in which care directed by the prediction model is compared to usual care [5, 6].

Impact studies for prediction models are infrequent, most likely due to their complexity, long follow-up (especially when predicting long-term prognostic outcomes), associated high costs and lack of regulatory requirements [7-9]. In addition, the benefits observed in such impact studies have typically been smaller than expected, or even lacking [10, 11]. A recurring explanation for the smaller than expected benefit is the lack of compliance (or adherence) with management recommendations linked to model predictions or scores. An approach using a decision analytic model, making use of evidence that is available at the time a prediction model impact study is being considered, may provide more insight into the degree of compliance required to a model's management recommendations to result in favourable health outcomes and/or costs of care when the model is implemented.

Decision analytic modelling is a method that integrates multiple sources to assess the downstream cost-effectiveness of applying a prediction model [12, 13]. Constructing a decision analytic model (DAM) forces researchers to think about the pathway through which complex interventions, such as the interplay between the model predictions and subsequent patient management, can lead to clinical or monetary benefit. DAMs also allow for uncertainty to be taken into account on parameters related to accuracy of predicted probabilities, costs and potential side-effects of diagnostic tests, and costs and effectiveness of any subsequent intervention. Additionally, downstream effects of hypothetical scenarios can be analysed, assuming one or more values for parameters for which there is little or no evidence. This allows for flexible assessment of the potential impact of the prediction model.

For (complex) therapeutic interventions and diagnostic tests DAMs have been proposed and performed before conducting longitudinal comparative trials, [14-16], however only a few have been performed on diagnostic or prognostic prediction models [8, 17]. An explanation for the lack of DAM assessments on prediction models is that the impact of a prediction model is more complex than assessing the impact of an intervention, as such models need to include the downstream effects of those predictions, for example on benefits and harms of subsequent diagnostic tests and treatments. Additionally, compliance with management recommendations from a prediction model and informed deviation to that compliance (i.e. whether there is incremental value of a clinician's experience on top of predictions provided by a model) – both difficult parameters to estimate – also influence the impact of using a prediction model in practice. These parameters are often not available or carefully considered before a trial is performed, however this does not need to prevent the construction of a sensible DAM producing realistic estimates of impact on health effects and healthcare costs.

In this paper we demonstrate how to assess the potential impact of a prediction model on health effects and healthcare costs using a DAM approach, specifically focusing on compliance with management recommendations corresponding to prediction model estimations. We will use the HEART score prediction model for diagnosis of major adverse cardiac events (MACE) in patients with chest pain as a case study [18]. We will end this paper by providing generic guidance on how to perform a model-based assessment of an clinical decision rule or prediction model using a decision analytical approach, and how this can be used to inform the decision on whether or not to conduct a subsequent prospective comparative prediction model impact study, or to inform the design and conduct of such studies.

2. METHODS

2.1 Case study

We compared implementation of the HEART score prediction model to usual care in a DAM as an example of how compliance with management recommendations from a prediction model influence the impact of that model on patients' health outcomes, healthcare costs, and cost-effectiveness of care. A randomized impact trial has been conducted for the HEART score prediction model, however for the DAM only information from sources available before

the conduction of this trial were used [19]. This paper aims to illustrate *how* a DAM can be used to assess the effects of compliance to management recommendations from a prediction model on actual impact of the model on patients' health and healthcare costs. Its aim is not to replicate the results from the randomized impact trial.

The HEART score is a prediction model that uses routinely collected information from patient history and blood tests to predict MACE in patients presenting with chest pain at the emergency room, using an easy applicable risk score ranging from 0 to 10 [18]. Patients with a low score (i.e. ≤ 3) were considered to have a low risk for MACE within six weeks after presentation at the emergency department, and could be discharged. Those with a score of 4 or higher would need to undergo further testing. Several external validations studies have shown that the HEART score can clearly stratify patients according to their risk of having MACE [20-24].

We evaluated the HEART score purely as a diagnostic instrument for MACE, meaning that in our model the HEART score and any subsequent actions don't have an impact on the total number of MACE. The potential benefit of the HEART score lies in its ability to stratify patients according to their risk of MACE. Physicians are advised to promptly discharge low risk patients, thereby reducing utilization of healthcare resources, and to provide a more aggressive diagnostic approach in high risk patients, to prevent unnecessary delay in starting treatment. MACE found during diagnostic work-up (detected MACE) were considered a favourable outcome, whereas MACE after patients were discharged (missed MACE) were considered an unfavourable outcome.

2.2 Structure of the model

A DAM comparing usual care to using the diagnostic-management pathway HEART score can be found in Figure 1. More details on the usual care pathway can be found in Appendix A. In the usual care strategy HEART scores are not available to clinicians and are therefore not used to guide management decisions. In the HEART score strategy, we mimicked that clinicians at the emergency department would calculate the HEART score, and be given clear guidance on subsequent management based on the value of that score. Following the original HEART score development and validation studies, the following categories were used: low (HEART score 0 to 3), intermediate (HEART score 4 to 6), and high (HEART score 7

to 10). Each of the three HEART score categories was also given a subsequent management recommendation. Patients in the low, intermediate, and high HEART score categories were respectively recommended to be discharged from the hospital without any diagnostic testing performed, given non-invasive diagnostic testing (which consists of stress bicycle echocardiography, myocardial scintigraphy, coronary CT angiography, and cardiac MRI) or given invasive diagnostic testing (coronary angiography, and any of the non-invasive tests). Our model allowed for an assistive prediction model approach, meaning physicians were not forced to comply with management recommendations [25, 26]. This depicts a situation more representative of actual utilization of a prediction model in practice, hence providing more realistic assessment of its impact. Therefore, we varied the amount of compliance (i.e. in this case study the proportion of patients within a HEART score group in which the specified recommendation was followed) in several scenarios (see "Scenario analysis" paragraph).

The focus of our DAM is an evaluation of the impact of compliance with management recommendations from HEART score provided by a clinical prediction rule like the HEART score on patient relevant health outcomes and costs.

In addition, we allowed clinicians to deviate from recommended management based on additional information (e.g. additional signs and symptoms) or expertise, leading to more appropriate stratification of management given to patients. This was done by including a variable in the model to account for the degree of informed deviation (ID) from management recommendations corresponding to HEART score predictions, defined as the proportion of patients in whom the physicians correctly decided not to follow HEART score management recommendations. ID ranged from 0% (uninformative compliance; compliance is equal in patients with and without MACE) to 100% (fully informative compliance; patients with MACE follow a diagnostic pathway, patients without MACE are discharged). For further illustration, see Table 1.

Table 1. Illustration of the interplay between compliance and informed deviation from management recommendations from HEART score predictions for the low HEART score category.

	MACE		No MACE	
	Discharged (proportion of patients)	Additional testing (proportion of patients)	Discharged (proportion of patients)	Additional testing (proportion of patients)
Full compliance (100%) No ID (0%)	100%	0%	100%	0%
Mostly compliance (80%) No ID (0%)	80%	20%	80%	20%
Mostly compliance (80%) Partly ID (50%)	40% = 80% - 80% * 0.5	60% = 20% + 80% * 0.5	90% = 80% + 20% * 0.5	10% = 20% - 20% * 0.5
Mostly compliance (80%) Full ID (100%)	0% = 80% - 80% * 1.0	100% = 20% + 80% * 1.0	100% = 80% * 20% * 1.0	0% = 20% - 20% * 1.0

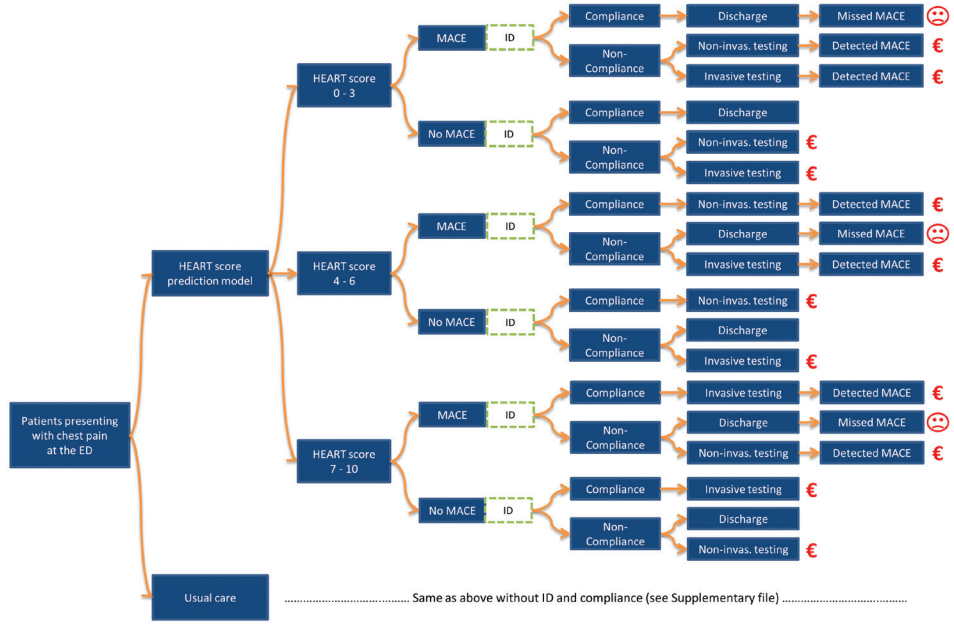


Figure 1. Decision tree for the application of the HEART score prediction model in a group of patients presenting with chest pain at the emergency department.
ID = Informed deviation of management recommendations corresponding to HEART score predictions, representing the proportion of correctly ignored HEART score recommendations. Euro signs and emoticons represent negative effects on costs and health effects respectively.

2.3 Input parameters for decision analytic model

To operationalize the DAM, each parameter requires an input value. Three types of input parameters are considered. Firstly, transition probabilities, which are the probabilities to transit from one (health) state to another, are defined (marked in Figure 1 by the orange arrows). Secondly, defining the health outcomes that the prediction model aims to prevent are set. Finally, input values for the intended and unintended effects and costs of any subsequent tests, treatments, and/or conditions need to be determined. See Appendix B for an overview of all input parameters, for instance the distribution of patients across the different HEART score categories.

Transition probabilities

The distribution of targeted patient population across the HEART score categories and MACE rates per HEART score category were derived from the development, and multiple external validation studies of the HEART score [20-22]. Values for compliance and ID were on the scenarios described in the "Scenario analysis" section of this paper. Transition probabilities for non-invasive and invasive diagnostic testing pathways, as well as the likelihood of receiving specific diagnostic tests (e.g. a stress bicycle echocardiography), were derived from a medical consumption study [27].

Health outcomes

The health outcome of interest was defined as the proportion of missed MACE, i.e. patients with a MACE who were (initially) discharged without any subsequent work-up. MACE that was detected during or occurred after diagnostic workup was not included as an adverse outcome, as this would have been found in a clinical setting, in which it could be managed accordingly. In usual care, the proportion of patients with a missed MACE was 0.4% (95% CI), 2.5% (95% CI), and 2.0% (95% CI), for the low, intermediate and high HEART score categories respectively [27]. All MACE (detected and missed) were defined as the occurrence of any of the following events or interventions: acute myocardial infarction (both ST- and non-ST-segment elevation), unstable angina, percutaneous coronary intervention (PCI), coronary artery bypass grafting (CABG), significant stenosis (>50%) managed conservatively, and death due to any cause [28].

Healthcare costs

The measurement of HEART score relies on readily available predictors, hence there is no extra cost associated with the collection of these predictors when compared to (current) usual care. The costs of MACE events were calculated based on a weighted average of costs and probability of each individual MACE component, derived from scientific literature. [27, 29-32] Costs for individual diagnostic tests were incorporated in the non-invasive and invasive testing pathways. Count data for the average number of individual diagnostic procedures per patient were extracted, meaning a patient could receive multiple diagnostic tests during the six week period of follow-up [27]. The average number of tests per patient was multiplied with their respective costs, yielding the average cost a patient generated when entering that specific diagnostic pathway. Similarly, the average number of admission and re-admission days were calculated per HEART score category for each of the diagnostic pathways were. Complication rates in non-invasive and invasive testing pathways are not explicitly included in the model. However, the expected frequency of severe complications for procedures included in the DAM is low and expected costs of complications are largely captured by the number of (re)admission days [33-35].

2.4 Analyses

We used data from previous studies as input for the parameters included in the DAM. A probabilistic sensitivity analysis was performed, in which a series of simulations were ran to take into account uncertainty surrounding the parameters in the DAM. Furthermore, hypothetical scenarios are considered in which compliance and ID were varied.

Scenario analysis

Scenario analysis focused on comparing different compliances to HEART score management recommendations, combined with varying degrees of ID with those compliances. The influence of compliance on missed MACE and costs was investigated in three different scenarios: low (50%), medium, (75%), and full (100%) compliance. Furthermore, four scenarios were defined for ID: (0%), low (25%), medium (50%), and high (75%) ID.

For each scenario, the incremental proportion of missed MACE, healthcare costs, and cost per missed MACE will be given per HEART score category, and for all HEART score categories combined, compared with usual care. Cost-effectiveness planes will be provided to give insight in the distribution of missed MACE and healthcare costs in the presence of parameter uncertainty.

Probabilistic sensitivity analysis

Monte-Carlo simulation was used to assess the robustness in health and economic outcomes based on uncertainty surrounding parameter estimates. A series of 10,000 simulations were run per scenario, each with a patient population of 200,000, reflective of the annual Dutch population visiting the emergency department with chest pain [36]. Parameter uncertainty was reflected by calculating standard errors from the data, and defining appropriate statistical distributions for each parameter. Beta and Dirichlet distributions were used to account for uncertainty in transition probabilities. Gamma distributions were used for uncertainty surrounding costs (see Appendix B).

3. RESULTS

The impact of compliance and informed deviation on the number of missed mace (i.e. effects), costs and ratio between costs and effects are investigated with pre-defined scenarios. The impact on these three different impact estimates within the different scenarios of compliance and informed deviation is shown. Because our interest is in missed MACE events, results may be slightly contraindicative as negative numbers now indicate to a more desirable outcome (less missed MACE). The values in the results are therefore coloured in green (reduction in MACE event or costs) or red (increase in missed MACE or costs). Additional to the scenario analyses is the robustness of the outcome estimates (i.e. effects and costs) investigated and presented in a cost-effectiveness plane with the most extreme scenarios (that is compliance of 50% / 100% and ID of 0% / 75%).

Missed MACE

Table 2 shows the average difference in missed MACE (per person) for each of the HEART score categories and the total patient population, compared with usual care. The low HEART score category shows an increase in the proportion of missed MACE events as compliance increases, whereas in the intermediate and high HEART categories there is an inverse relation. This can be explained by the different management recommendations associated with each HEART score category. Compliance in the low HEART score category leads to more patients being discharged, running the risk of missing MACE in these patients. On the other hand, compliance in the intermediate and high HEART score categories leads to more diagnostic testing and prolonged hospital stay, reducing the risk of missed MACE. ID can

counteract the higher proportion of missed MACE in the low HEART score category, and increase reduction of missed MACE in the intermediate and high categories, by selection of MACE patients to refer them for diagnostic testing, and selection of non-MACE patients and discharging them.

Costs

Table 3 shows the average difference in costs per patient between the HEART score strategy and usual care. Costs declined for the low and intermediate HEART score category when compliance and ID increase. A different pattern is observed when the high HEART score category is taken into consideration, where a higher compliance led to higher costs. ID counteracts the increase in this group. For the total group there is a cost increase when ID is 50% or lower. Only at higher levels did the costs decrease compared to usual care.

Missed MACE / costs ratio

To gain insight in the investment that needs to be made for reducing missed MACE, the ratio between the difference in costs and missed MACE between HEART categories and usual care is calculated. Difference estimates of compliance to management recommendations for the HEART categories may be slightly strange, e.g. combining 50% compliance with 0% ID in the low category and 75% compliance with 50% in intermediate and high categories. Therefore, the ratio between costs and missed MACE is only calculated for the total patient population. Table 4 shows the results for the different scenarios of compliance and informed deviation. The impact of introducing the HEART score strategy on cost per missed MACE event depended greatly on the interplay between compliance and ID. When the HEART score is mostly ignored (ID of 75%), the resulting decisions reduce costs and missed MACE, resulting in a promising (i.e. cost-effective) strategy. At some point (i.e. 100%, not so far from the 75% mentioned here) there is no HEART score intervention at play anymore.

Cost-effectiveness ratios are provided within a HEART score category and for the total population. HEART score strategy is inferior when there are both extra costs and more missed MACE compared to usual care. HEART score strategy is dominant when there are less costs and fewer missed MACE compared to usual care.

Uncertainty analysis

Figure 2 shows the incremental cost-effectiveness plane of four scenarios (compliance of 50% / 100% and ID of 0% / 75%) of the HEART score strategy compared to usual care (for 10,000 simulations). In a scenario where compliance to management recommendations from the HEART score is only 50% (with 0% ID), 68% of the simulations showed an inferior outcome. An inferior outcome means that there are more missed MACE events higher costs compared to usual care. When 100% compliance (and 0%) to the HEART score strategy was assumed there was no reduction in missed MACE but with costs savings per patient. For the scenarios with a positive ID of 75%, 99% and 78% of the simulations result in a reduction in missed MACE events and cost savings (for 50% and 100% compliance respectively).

Table 2. Average difference in missed MACE per person between a HEART score strategy and usual care.

		Compliance	Informed deviation (ID)			
			0%	25%	50%	75%
Usual care	HEART score 0-3	**50%**	0.006	0.004	0.001	-0.001
		75%	0.011	0.007	0.004	0.000
0.016		**100%**	0.016	0.011	0.006	0.001
	HEART score 4-6	**50%**	0.002	-0.005	-0.012	-0.018
		75%	-0.012	-0.015	-0.018	-0.021
		100%	-0.025	-0.025	-0.025	-0.025
	HEART score 7-10	**50%**	0.003	-0.003	-0.009	-0.014
		75%	-0.009	-0.012	-0.014	-0.017
		100%	-0.020	-0.020	-0.020	-0.020
	Total	**50%**	0.004	-0.001	-0.006	-0.011
		75%	-0.002	-0.005	-0.009	-0.012
		100%	-0.008	-0.010	-0.012	-0.014

A negative number represents a reduction in missed MACE.

Table 3. Average difference in costs between HEART score strategy and usual care

		Compliance	Informed deviation (ID)			
			0%	25%	50%	75%
Usual Care	HEART score 0-3	**50%**	€ 27	-€ 55	-€ 137	-€ 218
		75%	-€ 148	-€ 186	-€224	-€ 262
€ 1,456		**100%**	-€ 323	-€ 317	-€ 312	-€ 306
	HEART score 4-6	**50%**	€ 7	-€ 48	-€ 102	-€ 156
		75%	-€ 102	-€ 130	-€ 156	-€ 184
		100%	-€ 211	-€ 211	-€ 211	-€ 211
	HEART score 7-10	**50%**	€ 537	€ 159	-€ 221	-€ 599
		75%	€ 1,465	€ 1,081	€ 697	€ 311
		100%	€ 2,393	€ 2,003	€ 1,615	€ 1,223
	Total	**50%**	€ 96	-€ 19	-€ 135	-€ 251
		75%	€ 124	€ 38	-€ 50	-€ 137
		100%	€ 151	€ 94	€ 35	-€ 24

Table 4. Average difference in cost per missed MACE between HEART score strategy and usual care.

	Compliance	Informed deviation (ID)			
		0%	25%	50%	75%
Total	**50%**	€ 23,395*	**€ 14,981^**	**€ 22,307^**	**€ 23,160^**
	75%	*-€ 63,250*	*-€ 6,954*	**€ 5,681^**	**€ 11,256^**
	100%	*-€ 20,305*	*-€ 9,909*	*-€ 2,992*	**€ 1,770^**

* inferior; more missed mace and larger costs; ^ dominant; less missed mace event and cost saving.

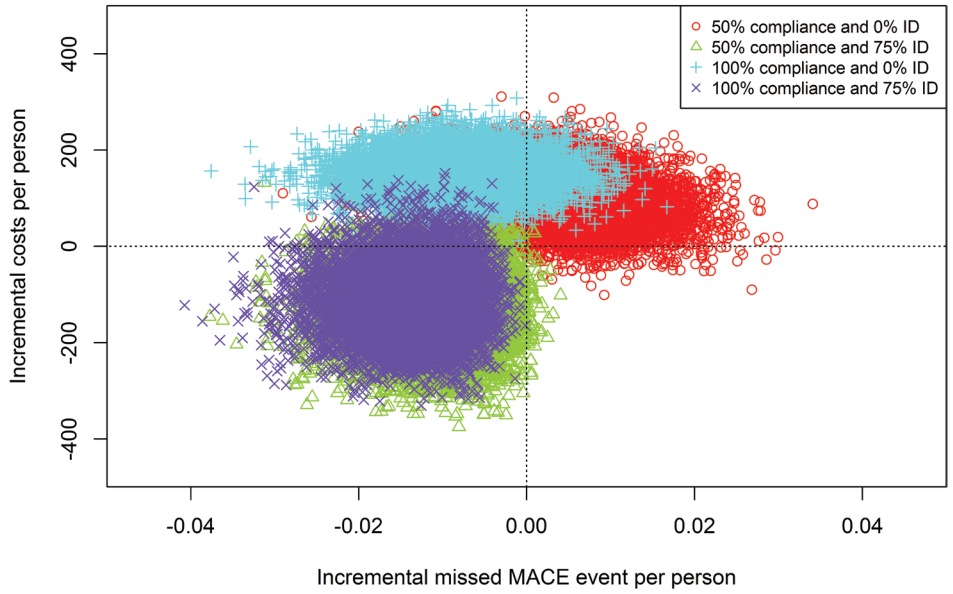


Figure 2. Incremental cost-effectiveness planes for 10,000 simulations comparing the HEART score strategy to usual care for the following scenarios
Note that negative numbers indicate a more desirable outcome (less missed MACE and/or reduction in costs).

4. DISCUSSION

In this paper we show how compliance and ID, as a measure for incremental value of experience of a clinician, affect the potential impact of using the HEART score on health effects and healthcare costs using a DAM. This illustrates how a DAM can be used to estimate the impact of a prediction model, using only data and information available before performing an impact trial.

Generating a DAM for impact assessment of a prediction model forces researchers to think about the ultimate goals of the prediction model (e.g. reducing the primary outcome, reducing side-effects, optimizing diagnostic and treatment pathways) and how it aims to achieve these goals. DAMs can help to demonstrate under which conditions a longitudinal comparative (ideally randomized) impact study is likely to have the desired impact on health effects and/or costs. If those conditions are deemed unlikely, then one should consider whether investment in a large-scale impact trial is justified [37]. Should those conditions be deemed plausible, a pilot study or qualitative assessments with experts in the field might be

considered, to gain more insight and reduce uncertainty on the parameter of interest? The DAM can be updated with additional information from for example a pilot study, allowing researchers to re-assess the potential impact of the prediction model. Ultimately the DAM can be used for optimizing the sample size calculation and the design and conduct (like the amount of attention on compliance) for an (potential) upcoming impact study [17, 38]. Additionally, the DAM can aid researchers in understanding how parameters influence the health effects and costs of an impact trial [39].

Feasibility of a DAM before an impact study has been performed depends on the availability of data on risk stratification of patients, consumption of healthcare resources, effectiveness of patient management, and potential health benefits for patients in each risk category. Table 5 provides an overview for general guidance on how to set-up a model-based impact assessment of a prediction model.

Strengths

This is one of the first examples in which a DAM was used for impact assessment of a prediction model, using solely data before implementation in clinical practice. We demonstrated methods that can be used for evaluation of a prediction model, providing insight in the influence of particular input parameters on the impact on clinical outcomes and/or costs. Compared to a clinical trial, modelling assessments such as the one demonstrated in this paper, require a fraction of the time and cost, and could help improve design and conduct of an impact trial. This paper demonstrates the feasibility of such an assessment using data from only a few publications in scientific literature (see Appendix B).

Limitations

There are a few considerations to fully appreciate the findings of the impact assessment in this paper. The use of healthcare resources in our model was based on the first 6 weeks of medical consumption [27]. It is likely that negative consequences from a MACE event will last beyond this timeframe. Markov chain modelling could account for these long-term effects, however reliable data for these effects were lacking [46]. Only in-hospital costs were incorporated in the assessment. Visits to the general practitioner, specifics on medication usage, and non-hospital based costs (e.g. labour productivity losses, traveling expenses) were not included in the assessment. Although these might influence the results when

considering a societal perspective, it is expected that the impact on incremental costs and health effects are likely to be limited.

We viewed the HEART score purely as a diagnostic tool. This implies that MACE is not prevented by using the HEART score. What the HEART score can only achieve is to stratify patients correctly and streamline subsequent management. Because MACE is not prevented, the natural outcomes is missed MACE, leading to poorer outcome and additional costs. Others have argued that the HEART score can also be used to predict MACE in the future, opening the opportunity to prevent MACE. This would of course lead to a rather different DAM. We chose not to do this because HEART was used in an acute care setting, patients with chest pain presenting at the emergency department which is a clear diagnostic setting.

Comparison with other studies

It is first worth comparing the results of our analysis to the impact trial that was eventually performed by Poldervaart et al [19]. Although not the main goal of this paper, a comparison with the results from the impact trial can provide insight in the validity of our model. The trial demonstrated 82.3% compliance to HEART score recommendations, which we could have been used as input for the DAM to compare health effects and costs. Unfortunately health effects could not be compared, because a different primary outcome was used in the impact trial compared to the outcome used in the DAM (missed MACE vs. any MACE). Furthermore, cost data were collected over a 3 month time horizon, different from the 6 week time horizon used in literature. Still, the impact of non-compliance in the DAM can be translated to the actual HEART impact trial. Non-compliance without informative deviation in patients with low HEART score has a detrimental effect on potential cost savings. This is in line with the main finding of the actual HEART impact trial: substantial non-compliance in the low HEART score category resulted in small differences in total cost between HEART and usual care.

Few model-based assessments have been previously performed that assess the potential of prediction models before an impact study has been executed. One study assessed the value of a prediction model for predicting shoulder pain in patients with early stage oral cavity squamous cell carcinoma after surgical removal of lymph nodes [40]. Although the analysis did focus on specific scenarios regarding the accuracy of the predictions, explicit consideration of compliance or additional clinical expertise on top of the prediction model,

were not evaluated. DAM assessments have also been used to calculate the potential of a novel innovation using the headroom method. Headroom analysis is a method that has been used in scientific research to assess the likelihood of cost-effectiveness of an intervention, given a willingness to pay threshold [14, 41-45]. Similar to the assessment in this paper, these analyses also make use of data before implementation of an innovation, and assess potential benefit. However, the goal of headroom analysis is to assess at a very early stage of development whether an innovation has the potential to be cost-effective. In a model-based impact analysis, costs and effects of an innovation are being estimated for a series of (realistic) scenarios. Although a headroom approach is feasible for prediction models, to our knowledge there are no articles on this topic described in literature.

Conclusion

Using DAM for impact assessment of the implementation of a prediction model, using solely data available before an impact study is conducted, can aid researchers in understanding what value of certain parameters should achieve to have a positive effect on health effects and/or costs. Based on the likelihood of observing such a value (in our example, 50% compliance, 75% ID), they can decide whether an impact trial is necessary, and if so, under what conditions such a trial is likely to reduce costs and/or reduce outcomes. Efforts can be directed at improving the design of an impact trial to prevent disappointing results. DAMs can help provide general insight in the mechanism through which the prediction model can lead to desired results, and expose potentials flaws in mechanistic pathways, allowing researchers to adapt the design of a trial beforehand. Ultimately model-based impact assessments have the potential to reduce research waste, by more efficient selection and design of impact trials.

Table 5. Guidance for a model-based impact assessment of prediction models, before data on clinical impact have become available.

Steps	Description	Sources / Methods	Examples from case-study
Design the DAM	Choose appropriate structure of DAM. Different structures are possible such as a decision tree, Markov model, or micro-simulation model.	• Current clinical pathways • Guidelines • Development/validation study of prediction model • Suggested intended use of prediction model • Expert opinion	Decision tree with risk categories, management recommendations regarding subsequent diagnostic testing, and relevant clinical outcomes (see Figure 1 and Supplementary file).
Collect parameter estimates Define uncertainty	Determine probabilities (e.g. transition probabilities between (health) states and probability of side-effects), impact estimates (e.g. health effects and costs of treatment). Uncertainty surrounding these parameters should be incorporated in the model. Uncertainty analysis can be performed in deterministic or probabilistic way.	• Medical consumption studies • Electronic Health Records • Expert opinion • IPD data or pre-defined distribution • Expert opinion	For example, define distribution of heart scores based individual patient data from development/validation study. Appendix B shows an overview of all used parameters and the incorporated uncertainties.
Compare different strategies	Determine the (hypothetical) strategies that will be compared; these can be defined by manually varying parameter values, or choosing alternative risk threshold(s) for (different) risk categories.	• Two-way sensitivity analysis • Threshold analysis • Best, reasonable, or worse case scenarios	For example, compliance with management recommendations and informed deviation (Figure 2 and Tables 1, 2 and 3).
Interpret results and define next steps for future research	Results of the impact study can be used to improve an (potential) upcoming impact trial. For example, the results can be used to argue the need for a pilot study to determine the achievable compliance.		

REFERENCES

1. Steyerberg, E.W. and F.E. Harrell, Jr., *Prediction models need appropriate internal, internal-external, and external validation.* J Clin Epidemiol, 2016. 69: p. 245-7.
2. Steyerberg, E.W., et al., *Prognosis Research Strategy (PROGRESS) 3: prognostic model research.* PLoS Med, 2013. 10(2): p. e1001381.
3. Bleeker, S.E., et al., *External validation is necessary in prediction research: a clinical example.* J Clin Epidemiol, 2003. 56(9): p. 826-32.
4. Altman, D.G., et al., *Prognosis and prognostic research: validating a prognostic model.* BMJ, 2009. 338: p. b605.
5. Dharmarajah, B., et al., *Impact of risk scoring on decision-making in symptomatic moderate carotid atherosclerosis.* Br J Surg, 2014. 101(5): p. 475-80.
6. van Montfort, P., et al., *Implementation and Effects of Risk-Dependent Obstetric Care in the Netherlands (Expect Study II): Protocol for an Impact Study.* JMIR Res Protoc, 2018. 7(5): p. e10066.
7. Ferrante di Ruffano, L., et al., *A capture-recapture analysis demonstrated that randomized controlled trials evaluating the impact of diagnostic tests on patient outcomes are rare.* J Clin Epidemiol, 2012. 65(3): p. 282-7.
8. van Giessen, A., et al., *Systematic Review of Health Economic Impact Evaluations of Risk Prediction Models: Stop Developing, Start Evaluating.* Value Health, 2017. 20(4): p. 718-726.
9. Wallace, E., et al., *Impact analysis studies of clinical prediction rules relevant to primary care: a systematic review.* BMJ Open, 2016. 6(3): p. e009957.
10. Kappen, T.H., et al., *Impact of risk assessments on prophylactic antiemetic prescription and the incidence of postoperative nausea and vomiting: a cluster-randomized trial.* Anesthesiology, 2014. 120(2): p. 343-54.
11. Huang, D.T., et al., *Procalcitonin-Guided Use of Antibiotics for Lower Respiratory Tract Infection.* N Engl J Med, 2018. 379(3): p. 236-249.
12. Drummond MF, et al., *Methods for the economic evaluation of health care programme.* 3rd ed. 2005, Oxford: Oxford University Press.
13. IJzerman, M., et al., *Emerging Use of Early Health Technology Assessment in Medical Product Development: A Scoping Review of the Literature.* Pharmacoeconomics, 2017. 35(7): p. 727-740.
14. McAteer, H., et al., *Cost-effectiveness analysis at the development phase of a potential health technology: examples based on tissue engineering of bladder and urethra.* J Tissue Eng Regen Med, 2007. 1(5): p. 343-9.
15. Greving, J.P., et al., *Cost-effectiveness of aspirin treatment in the primary prevention of cardiovascular disease events in subgroups based on age, gender, and varying cardiovascular risk.* Circulation, 2008. 117(22): p. 2875-83.
16. Miquel-Cases, A., et al., *Early stage cost-effectiveness analysis of a BRCA1-like test to detect triple negative breast cancers responsive to high dose alkylating chemotherapy.* Breast, 2015. 24(4): p. 397-405.

17. Kappen, T.H., et al., *Evaluating the impact of prediction models: lessons learned, challenges, and recommendations.* Diagnostic and Prognostic Research, 2018. 2(1): p. 11.

18. Six, A.J., B.E. Backus, and J.C. Kelder, *Chest pain in the emergency room: value of the HEART score.* Neth Heart J, 2008. 16(6): p. 191-6.

19. Poldervaart, J.M., et al., *Effect of Using the HEART Score in Patients With Chest Pain in the Emergency Department: A Stepped-Wedge, Cluster Randomized Trial.* Ann Intern Med, 2017. 166(10): p. 689-697.

20. Six, A.J., et al., *The HEART score for the assessment of patients with chest pain in the emergency department: a multinational validation study.* Crit Pathw Cardiol, 2013. 12(3): p. 121-6.

21. Backus, B.E., et al., *A prospective validation of the HEART score for chest pain patients at the emergency department.* Int J Cardiol, 2013. 168(3): p. 2153-8.

22. Backus, B.E., et al., *Chest pain in the emergency room: a multicenter validation of the HEART Score.* Crit Pathw Cardiol, 2010. 9(3): p. 164-9.

23. Streitz, M.J., et al., *A retrospective external validation study of the HEART score among patients presenting to the emergency department with chest pain.* Intern Emerg Med, 2018. 13(5): p. 727-748.

24. Oliver, J.J., et al., *An external validation of the HEART pathway among Emergency Department patients with chest pain.* Intern Emerg Med, 2018.

25. Kappen, T.H., et al., *Barriers and facilitators perceived by physicians when using prediction models in practice.* J Clin Epidemiol, 2016. 70: p. 136-45.

26. Moons, K.G., et al., *Risk prediction models: II. External validation, model updating, and impact assessment.* Heart, 2012. 98(9): p. 691-8.

27. Nieuwets, A., et al., *Medical consumption compared for TIMI and HEART score in chest pain patients at the emergency department: a retrospective cost analysis.* BMJ Open, 2016. 6(6): p. e010694.

28. Poldervaart, J.M., et al., *The impact of the HEART risk score in the early assessment of patients with acute chest pain: design of a stepped wedge, cluster randomised trial.* BMC Cardiovasc Disord, 2013. 13: p. 77.

29. Soekhlal, R.R., et al., *Treatment costs of acute myocardial infarction in the Netherlands.* Neth Heart J, 2013. 21(5): p. 230-5.

30. Bekelman, J.E., et al., *Comparison of Site of Death, Health Care Utilization, and Hospital Expenditures for Patients Dying With Cancer in 7 Developed Countries.* JAMA, 2016. 315(3): p. 272-83.

31. Mahoney, E.M., et al., *One-year costs in patients with a history of or at risk for atherothrombosis in the United States.* Circ Cardiovasc Qual Outcomes, 2008. 1(1): p. 38-45.

32. Osnabrugge, R.L., et al., *Cost-effectiveness of percutaneous coronary intervention versus bypass surgery from a Dutch perspective.* Heart, 2015. 101(24): p. 1980-8.

33. Baim, D. and W. Grossman, *Complications of cardiac catheterization.* Cardiac Catheterization, Angiography and Intervention. 1996, Baltimore Williams & Wilkins.

34. Johnson, L.W., et al., *Coronary arteriography 1984-1987: a report of the Registry of the Society for Cardiac Angiography and Interventions. I. Results and complications.* Cathet Cardiovasc Diagn, 1989. 17(1): p. 5-10.

35. Wyman, R.M., et al., *Current complications of diagnostic and therapeutic cardiac catheterization.* J Am Coll Cardiol, 1988. 12(6): p. 1400-6.

36. van der Zalm, P. *HEART-score biedt arts extra handvat*. 2016 31-10-2018]; Available from: https://www.zorgvisie.nl/heart-score-biedt-arts-extra-handvat/.

37. Lilford, R.J., et al., *Evaluating policy and service interventions: framework to guide selection and interpretation of study end points.* BMJ, 2010. 341: p. c4413.

38. Bader, C., et al., *A new approach for sample size calculation in cost-effectiveness studies based on value of information.* Bmc Medical Research Methodology, 2018. 18(1): p. 113.

39. Koffijberg, H., et al., *Value of Information Choices that Influence Estimates: A Systematic Review of Prevailing Considerations*. Med Decis Making, 2018. 38(7): p. 888-900.

40. Govers, T.M., et al., *Integrated prediction and decision models are valuable in informing personalized decision making.* J Clin Epidemiol, 2018.

41. Girling, A., et al., *Headroom Approach to Device Development: Current and Future Directions.* Int J Technol Assess Health Care, 2015. 31(5): p. 331-8.

42. Girling, A., et al., *Early-stage valuation of medical devices: the role of developmental uncertainty.* Value Health, 2010. 13(5): p. 585-91.

43. Van Nimwegen, K.J., *Feasibility of the Headroom Analysis in Early Economic Evaluation of Innovative Diagnostic Technologies With no Immediate Treatment Implications.* Value Health, 2014. 17(7): p. A550.

44. Cosh, E., et al., *Investing in new medical technologies: A decision framework.* Journal of commercial biotechnology, 2007. 13(4): p. 263-271.

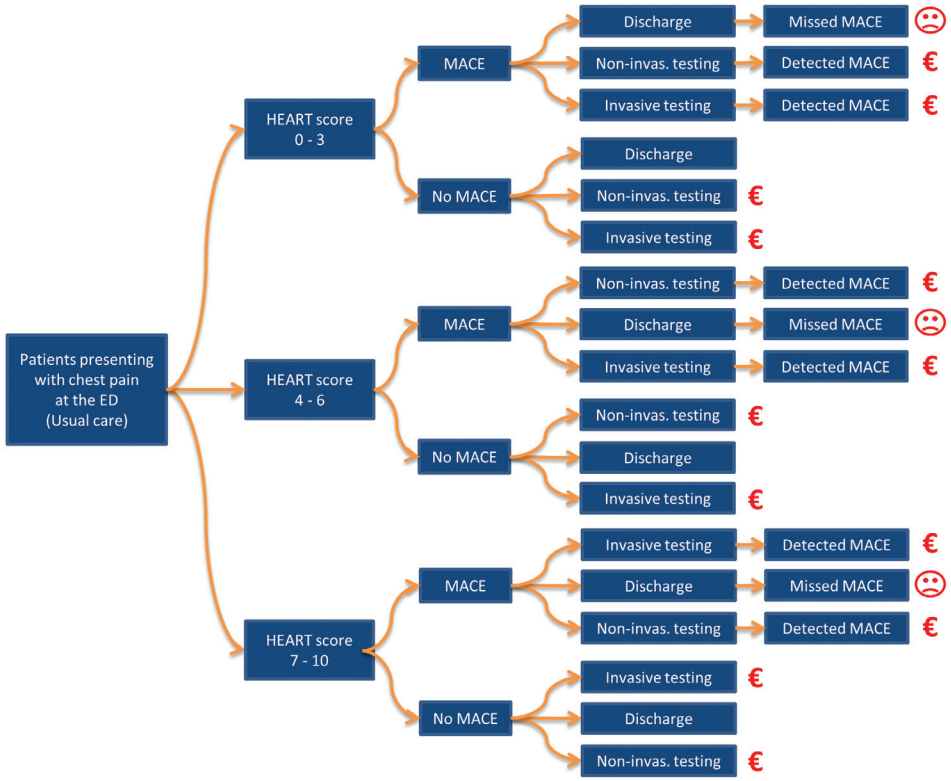
45. Miquel-Cases, A., et al., *Exploratory Cost-Effectiveness Analysis of Response-Guided Neoadjuvant Chemotherapy for Hormone Positive Breast Cancer Patients.* PLoS One, 2016. 11(4): p. e0154386.

46. Weinstein, M.C., et al., *Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies.* Value Health, 2003. 6(1): p. 9-17.

APPENDICES

A. DESCRIPTION OF USUAL CARE

Usual care strategy for the decision tree. Compliance and IDC have been removed compared to the HEART score strategy. HEART scores in the strategy can be calculated, but are not provided to clinicians, nor are the management recommendations attached to these scores.



Appendix A - Figure 1: Visual representation of the usual care strategy

B. MODEL PARAMETERS

Appendix B - Table 1: Overview of the used model parameters for the decision analytic model

	Average value	95% CI	Distribution	Source
Probabilities				
HEART score 0-3	0.400	0.363-0.438	Dirichlet	[27]
Probability at least one MACE < 6 weeks	0.020	0.006-0.041	Beta	[27]
• Compliance				
› Discharged	1.000		-	Consequential
› Average # of MACE events	1.000	0.905-1.095	Uniform	[27]
• Non-compliance				
› Non-invasive testing	0.750	0.236-0.997	Beta	[27]
› Average # of MACE events	1.333	1.206-1.460	Gamma	[27]
› Invasive testing	0.250	0.003-0.764	Beta	[27]
› Average # of MACE events	1.000	0.905-1.095	Uniform	[27]
Probability of no MACE < 6 weeks	0.980	0.959-0.994	Beta	[27]
• Compliance				
› Discharged	1.000	0.905-1.095	Uniform	Consequential
• Non-compliance				
› Non-invasive testing	0.948	0.898-0.981	Beta	[27]
› Invasive testing	0.052	0.019 -0.102	Beta	[27]
HEART score 4-6	0.444	0.309-0.490	Dirichlet	[27]
Probability at least one MACE < 6 weeks	0.236	0.188-0.287	Beta	[27]
• Compliance				
› Non-invasive testing	1.000		-	Consequential
› Average # of MACE events	1.194	1.081-1.308	Gamma	[27]
• Non-compliance				
› Discharged	0.226	0.099-0.388	Beta	[27]
› Average # of MACE events	1.000	0.905-1.095	Uniform	[27]
› Invasive testing	0.774	0.612-0.901	Beta	[27]
› Average # of MACE events	1.167	1.057-1.278	Gamma	[27]
Probability of no MACE < 6 weeks	0.764	0.713-0.812	Beta	[27]
• Compliance				
› Non-invasive testing	1.000		-	Consequential
• Non-compliance				
› Discharged	0.806	0.719-0.880	Beta	[27]
› Invasive testing	0.194	0.120-0.281	Beta	[27]

	Average value	95% CI	Distribution	Source
Probabilities				
HEART score 7-10	0.156	0.130-0.184	Dirichlet	[27]
Probability at least one MACE < 6 weeks	0.590	0.495-0.683	Beta	[27]
• Compliance				
› Non-invasive testing	1.000	0.905-1.095	Uniform	Consequential
› Average # of MACE events	1.333		Gamma	[27]
• Non-compliance				
› Discharged	0.077	0.009-0.205	Beta	[27]
› Average # of MACE events	1.000	0.905-1.095	Uniform	[27]
› Invasive testing	0.923	0.795-0.991	Beta	[27]
› Average # of MACE events	1.458	1.319-1.597	Gamma	[27]
Probability of no MACE < 6 weeks	0.410	0.317-0.505	Beta	[27]
• Compliance				
› Non-invasive testing	1.000		-	Consequential
• Non-compliance				
› Discharged	0.132	0.045-0.258	Beta	[27]
› Invasive testing	0.868	0.742-0.955	Beta	[27]
Costs				
HEART score 0-3				
• Non-invasive testing	€ 558	540-576	Gamma	[27]
• Invasive testing	€ 2,900	2843-2958	Gamma	[27]
HEART score 4-6				
• Non-invasive testing	€ 1,458	1,434-1,482	Gamma	[27]
• Invasive testing	€ 5,729	5,679-5,778	Gamma	[27]
HEART score 7-10				
• Non-invasive testing	€ 2,701	2,668-2,734	Gamma	[27]
• Invasive testing	€ 6,145	6,103-6,186	Gamma	[27]
MACE event	€ 5,484		Gamma	Weight. average

8.

General discussion

ABSTRACT

Risk prediction models are useful tools to estimate the risk of the presence (diagnosis) or future occurrence (prognosis) of a particular outcome using a combination of an individual's characteristics.

The process of prediction model development, validation and additional updating or tailoring to the target population is described extensively. Unfortunately, the prediction model evaluation process typically stops after these steps. Accurate predictive performance of a model is often assumed to be sufficient for implementation of the model in guidelines or even daily clinical practice. However, good predictive performance does not imply that implementing the model will impact treatment decisions, and consequently improve the health outcomes of the targeted individuals. Quantification of this impact of using a prediction model compared to not using the model is just as necessary as its development and validation.

To assess the impact of prediction models, one may perform a study in a randomised comparative design. Such studies are often time consuming, costly and sometimes infeasible if the predicted outcome occurs over a long time period (e.g. 30 years). An alternative is a decision analytic modelling approach, where evidence of the predictive accuracy of the model and effects of and adherence to subsequent treatments is linked. Decision analytic models are less time consuming and costly, may evaluate more than two model-treatment strategies at once, address the differential effects of treatment, address costs and side effects of considered treatments, address long term outcomes, and are ideal in situations with low number of outcome events. There are also disadvantages of using a decision analytic model, for example, they may require evidence which is lacking, limited, or available only in poor quality, such as information on actual prediction model use and actual treatment adherence.

This paper gives an overview of the advantages and disadvantages of a model-based approach to assess the impact of prediction models, without performing a prospective comparative study first. Two examples of impact studies of a prediction model with a model-based approach are described in detail, and guidance on how to perform a model-based impact assessment is discussed.

The content included in this thesis aimed to explore the challenges and solutions for decision analytical impact assessments of prediction models. This thesis ends with a general discussion and overview of these challenges and solutions.

Prediction models may address diagnosis and prognosis of individuals. *Diagnostic prediction models* calculate the probability that an individual suspected to have a certain disorder actually *has* that disorder, for example to guide referral to more advanced testing. *Prognostic prediction models* calculate the probability of the *future occurrence* of a particular outcome in individuals with a certain health state (e.g. with a certain diagnosis) to guide, for example, patient counselling or therapeutic management decisions. Numerous guidelines and books have proposed some kind of a stepwise process in prediction modelling before the model is implemented in daily practice or recommended in clinical guidelines [1-7]. The first step is *model development*, i.e. to identify important predictors and estimate the weights per predictor, typically using some form of multivariable analysis. The model development is followed by *model validation*, i.e. to evaluate the predictive performance of the developed model in patient-level data that was not used to develop the model. If indicated by the results of the validation study, the model may need to be further tailored or updated to the validation setting at hand. Finally, one needs to quantify to what extent the actual use of a successfully validated prediction model, as compared to not using the model, indeed impacts decision making and subsequently the health outcomes of the targeted individuals, the costs and the cost-effectiveness of care [2, 5, 8].

Alike every diagnostic or prognostic test or device, prediction models are 'information generating tools' that only indirectly impact health and healthcare outcomes [2, 4, 8-10]. They are not effective or therapeutic by themselves. Prediction models typically provide a probability of some outcome occurrence, that is supposed to guide decision making and subsequent preventive or therapeutic actions, which in turn – if properly indicated and effective - impact health outcomes and cost-effectiveness of care. Examples of such actions are healthcare providers administering drugs or other type of interventions or encouraging patients to improve their lifestyle to lower their prognostic outcome risk [11]. The use of prediction models themselves also comes with burden and costs, as they incorporate test, device or biomarker results as predictors, whereas the interpretation of predicted risks by care professionals and subsequent counselling requires time and thus also is not free of costs.

Prediction model impact studies

As is common for any health intervention, to study the effectiveness or impact of the use of a prediction model, with acceptable performance in (several) validation studies, typically requires a comparative design, ideally a randomised comparative design. In the index group, the prediction model is used and individuals are managed based on their predicted risks (e.g. treat individuals above a certain risk threshold and not treating anyone below that risk threshold), and in the control group the prediction model is not used and individuals are managed by prevailing care [2, 4, 8]. The observed patient (health) outcomes in both groups are then compared. Ideally, these impact studies follow a parallel cluster design, as randomisation at the individual participant level may lead to learning effects in the healthcare worker applying the prediction model [2, 4, 8]. Randomised stepped wedge designs may be particularly attractive given the typically limited de-implementation costs in case the index prediction model appears to have no favourable impact and needs to be de-implemented in the clusters that were randomised to the index group [12]. Also, less cumbersome but also more susceptible to bias, non-randomised longitudinal comparative designs may be used, such as before-after studies (comparing the outcomes observed in a period *before* introducing a prediction model into a particular healthcare practice with outcomes observed *after* its introduction) or geographical comparison studies (compare different centres that have or have not introduced the prediction model) [2, 4, 8].

To different extents, all these longitudinal impact studies are time consuming, costly, and quickly become almost infeasible, especially for prognostic models that predict outcomes over a longer period of time. An additional complicating factor for proper evaluation is that massive numbers of new prediction models have been published over the past two decades. For example, there are over 100 models to predict outcome after brain trauma [13], over 60 models for breast cancer prognosis [14], over 350 models predicting CVD risk in the general population [15, 16], over 45 models for cardiovascular events after being diagnosed with diabetes [17] and over 40 models for predicting prevalent and incident type 2 diabetes [18]. Hence there is a large amount of prediction models among which numerous for the same outcome or target population. Moreover, with the abundance of models for some indications, it becomes more urgent to compare models among each other, to select the model with the highest impact on health outcomes or the most favourable cost-effectiveness.

Decision analytic modelling approaches

An efficient alternative to empirical longitudinal impact studies is to use decision analytic modelling approaches, based on so-called linked evidence approaches or evidence synthesis methods [19-21]. In general, such evidence synthesis methods link the evidence from different sources or study types. One source contains the evidence that the prediction model shows good predictive performance across various validation studies. A second source includes the estimated effects and effectiveness of the interventions that are supposed to be administered by the predicted risks of the model – ideally obtained from randomised studies on these interventions. A third source includes evidence of the undesired events and (monetary) costs of using these interventions and prediction model [4, 19, 22-24]. For example, data on well-developed and validated models predicting the 10-year risk of CVD in middle aged males with mild hypercholesterolemia, and the estimated effects of interventions for treating middle aged males with this condition (from randomised trials) can be linked in a decision analytic model to estimate the impact of introducing such models with subsequent risk-based treatment in daily care [25].

In contrast to longitudinal (randomised or non-randomised) prospective studies, such decision analytical modelling approaches may more easily compare more than one prediction model, compare different risk thresholds for treatment selection, include the effects of different treatments and estimate the effects across relevant subgroups. Of course, such decision analytic model approaches also come with challenges and disadvantages. For example, there may yet be limited empirical evidence on the predictive accuracy of a model from multiple validation studies; on which therapeutic/preventive action is indicated at which risk thresholds; on the (long term) effectiveness of alternative therapeutic/preventive strategies from randomised studies; adverse events and costs of these strategies; the different patient relevant outcomes or the adherence to each intervention.

Indeed, decision analytic approaches cannot fully replace well designed, long term, comprehensive cluster randomised trials of prediction models where all necessary data can be prospectively collected, and potential biases can be prevented by design and analysis. But given the numerous existing prediction models and the attractive extra possibilities of decision analytic approaches, this method is suitable for at least separating the chaff from the wheat, indicating which validated models are most promising in a certain field and might undergo further evaluation of their impact in empirical comparative studies. However, despite their advantages, model based impact evaluations of prediction models are still rarely performed [26].

The aim of this paper is not to provide all details on how to perform decision analytical modelling – for this we will refer to the literature [27-30]. We rather provide a gentle overview of the strengths and weaknesses of such approaches when applied to assess and appraise the impact and cost-effectiveness of prediction models. This paper is not only suitable and useful for prediction model researchers, but is also informative for health policy makers and guideline developers to help understanding to what extent decision analytical approaches can be directive for deciding which prediction model or risk based approach might be advocated into guidelines or practice.

Advantages and disadvantages of the decision analytic approach

As partially discussed in the previous section, a decision analytic modelling approach to assess the impact of prediction models has advantages over a trial-based approach, but also has disadvantages. Both approaches address very different aspects or dimensions that are not easily weighed and compared. For example, decision analytic modelling approach in general has the advantage of being less costly than a trial-based approach, but also generates lower level evidence, as defined in evidence hierarchies [31, 32]. A comprehensive overview of the main advantages and disadvantages of a decision analytic modelling approach for impact assessment of prediction models is provided in Box 1. Box 2 shows different scenarios in which using a decision analytic modelling approach may be worthwhile for the impact assessment of prediction models.

Guidance on the impact assessment of a prediction model

Guidance on how to assess the impact of a prediction model without performing a large scale prospective comparative study is described in three steps below. Table 1 summarizes the different steps, with corresponding references to the literature.

Step 1: Development, validation and updating of the prediction model

a. The prediction model

The prediction model needs to be developed and validated for, or otherwise tailored to the targeted population and setting, as acknowledged by many papers and books (we cite a few) [4, 6, 7, 33-35]. Although a properly validated and updated model does not automatically lead to improved health outcomes and cost savings, the predictive performance of the model needs to be accurate since the predicted probabilities need to be correct (i.e. good model calibration). Poor predictive performance, i.e. poor calibration, leads to poor risk estimates and thus misclassification of individuals according to their (poorly) predicted probabilities, which in turn may result in unnecessarily treated (i.e. over-treated) individuals and incorrectly not-treated individuals.

b. Uncertainty in predicted risk estimates

The individuals' predicted risk estimates of the considered model are the input values for the decision analytic model. A perfectly calibrated and discriminating prediction model is utopic since there is always uncertainty in the predictor-outcome associations (i.e. the predictor weights), which in turn leads to uncertainty in the risk predictions by the model. The uncertainty in the risk predictions needs to be included in the decision analytic model, as it impacts the effectiveness of the use of the model on decision making and thus on individuals' health outcomes. Hence, it is preferred that some statistical distribution around the model's predicted risk estimates is defined based on individual patient data from development and/or validation study, or that individual patient data is available.

Step 2: Development of the decision analytic model

a. Selecting an appropriate structure

The choice of an appropriate structure for the decision analytic model depends on multiple considerations. A few important ones are explained below, for the other consideration we conclude with references to literature.

First, the structure of the decision analytic model depends on whether individuals are represented as a group, i.e. a cohort, or as individuals. The main advantage of an individual

based model is the possibility to include individuals in the decision analytic model based on their characteristics. For example, individuals can be selected on a particular predicted risk level (e.g. all with a predicted probability of 10% or higher), or on a combination of risk level and predictor values (e.g. a predicted risk above 10% or a systolic blood pressure of 180 mmHg or higher) [36]. In a cohort based model, a whole (sub) group is selected or not and moreover, all individuals get the same treatment (with the same treatment effectiveness) regardless of their individual characteristics. In cohort models, heterogeneity beyond age and gender, in terms of individuals' characteristics and (medical) history, may not be fully captured. Decision tree approaches and Markov-chain models [37-39] are examples of a cohort-based model, whereas a so-called micro-simulation model represents an individual based approach [40, 41].

Second, the prediction horizon plays a role in selecting the right structure of the model. If there is a short time horizon, a decision tree approach is useful whereas for a longer time horizon, a state-transition modelling is more appropriate, e.g. a Markov chain model or a micro-simulation model [28]. A decision tree approach is a simple model and is mainly used for acute conditions or diagnostic models where the probability of occurrence of an event is constant over time, and typically only a limited number of (subsequent or recursive) events are included. A state-transition model is a recursive model where outcome events can occur repeatedly and outcome probabilities can change over time, e.g. due to ageing or undergoing treatments or changing lifestyle. This type of model is more appropriate for chronic diseases, e.g. heart disease, cancer, and diabetes.

Third, the definition of the health states (i.e. outcome) plays a role. The defined health states in the decision analytic model need to correspond to the underlying disease process or to the health status, and need to reflect the clinical classification of the disease. Different health states are for example, the stages of cancer or different types of CVD events (stroke, heart failure).

For more details on decision model characteristics and guidance on selecting the most appropriate structure for the decision analytic model, we refer to the literature [27-30].

b. Collect values for model parameter estimates

After selecting an appropriate design, it is important to obtain the necessary evidence from the literature to use as input values for the model. The following input estimates for the model parameters need to be considered.

I. Probability estimates

› *Transition probabilities*

These parameters represent the probability to transit into each possible outcome state.

A very important transition is the probability to transit from a "healthy" state to an "outcome" state. This probability is represented by the predicted outcome risks of the prediction model. Preferably, individual patient data from the prediction model development study are available to get detailed information on the predicted risks distribution of the targeted patients. Other transition probabilities are estimates of the disease progression without treatment, taking into account survival data to estimate the probability to die from natural causes.

› *Health state probabilities*

Other important probabilities for the decision analytical model are probability estimates within a state, e.g. probabilities for the considered treatments' intended effects, adverse effects, and treatment adherence. Evidence on these probabilities can be obtained from randomised treatment studies, observational (long term) cohorts, (inter)national guidelines, or by expert elicitation.

II. Impact estimates health state

Besides the occurrence of the above defined outcome events in the pre-defined health states, also the (average) health effects and monetary costs within the health states are important elements of long term decision analytic models.

› *Health effects*

The (average) quality of life (QoL) of the considered individuals, the decrease of QoL due to (stages of) a disease, outcomes and/or complications of the treatment, and health effects of treatment are important health effects to incorporate in the

decision analytic model. Evidence on these health effects can also be obtained from the literature or from individual patient data.

› *Costs*
Costs of different treatments, of complications due to an intervention, and of tests or biomarkers that are incorporated as predictor in the prediction model are examples of monetary costs. Preferably, all parameters considering costs (if possible and available) are taken into account: the number of hospital admission days, subsequent (laboratory) tests, medication, visits to therapists, or rehabilitation. It is even possible to consider costs from outside the hospital, e.g. visits to an acupuncturist or 'over-the-counter' medication.

Evidence on costs can be obtained from financial hospital records or guidelines on costs, measured in RCTs via cost-diaries, or calculated from count data (electronic health records or medical consumption data) and unit prices.

c. Uncertainty of model parameter estimates

Since all above described parameter estimates are empirical research based, they inherently have imprecision. It is important to explore the robustness of the estimated impact of the use of a prediction model by addressing the uncertainty of the decision analytical model parameter estimates. Uncertainty analyses can be performed in a deterministic way (i.e. a deterministic sensitivity analysis (DSA)) where model parameters are varied manually to test the sensitivity of the model impact estimates to specific parameters or sets of parameters. Alternatively, a probabilistic sensitivity analysis (PSA) can be performed where parameters are sampled from pre-defined statistical distributions around these model parameters and varied simultaneously within a large number of simulations. The outcomes of a PSA are confidence intervals of the impact estimates or cost-effectiveness estimates [42], and can be used to perform a so-called value of information (VOI) analysis. A VOI analysis can be used to investigate the value of collecting additional data to reduce the uncertainty in impact outcomes and thereby potentially aid decision making regarding the implementation of the prediction model in practice [43-45].

Step 3: Compare different strategies

Assessing the impact of the use of the prediction model as compared to not using the model is in fact a comparison between an index and a control (care as usual) strategy [46]. Decision modelling approaches have the advantages that multiple strategies can be compared. One may compare different index strategies, i.e. using different treatment probability thresholds, or different potential treatment strategies that are administered at different probability thresholds [24, 47]. For example, different screening strategies due to using different risk thresholds or a combination of different thresholds for osteoporosis in postmenopausal women [48]. Furthermore, it is possible to determine the 'optimal' impact (e.g. minimize complications or maximize health outcomes) of a prediction model by gradually selecting more individuals for treatment (i.e. gradually lowering risk thresholds) [24].

Table 1: Guidance on how to perform a model-based impact assessment of a prediction model

Step	Description	Sources
1.	**Development, validation and updating of the prediction model**	
	a. The prediction model › Prediction model is developed and validated for, or otherwise tailored to targeted population. › Prediction model has a high predictive performance, i.e. good calibration and discrimination.	[4, 6, 7, 33-35]
	b. Uncertainty in predicted risk estimates › Defined distribution around model's predicted risk based on individual patient data from the development and/or validation study, or individual patient data set is available.	
2.	**Development of the decision analytic model**	[27]
	a. Select an appropriate structure for the model The structure depends on different consideration; › Representation of individuals; Level of details required for the predictors of the prediction model indirectly influence choice of representation of patients; small low level of details (i.e. decision rule or homogeneous patient group) versus high (i.e. blood pressure or a heterogeneous patient group) results in cohort versus individual based decision analytic model. › Time horizon of the modelled outcomes; Diagnostic models often have a short time horizon whereas prognostic models have a long horizon; this may impact the choice for a decision tree approach versus a state-transition approach (i.e. Markov chain model) › Definition of the health states The defined states need to correspond to the underlying disease process, health status, or a combination of both, e.g. the detected stages of cancer.	[27, 28]
	b. Collect values for model parameter estimates Obtain necessary evidence from literature. The following estimates need to be considered: I. Probability estimates › *Transition probabilities* Probability to transit between health states, e.g. to transit from 'healthy' state to 'outcome' state. › *Health state probabilities* Probabilities within a health state, e.g. to experience treatment effects, adherence to treatment, or adverse side-effects of the treatment. II. Impact estimates health state › *Health effects* Change in quality of life due to the disease, adverse side-effects or health effects of the treatment. › *Costs* All possible and available cost estimates need to be incorporated e.g. cost of a new predictor in the prediction model, cost concerning treatment, or costs due to adverse side-effects of the treatment.	
	c. Uncertainty in model parameter estimates Uncertainty analysis can be performed in a deterministic or a probabilistic way; › Deterministic sensitivity analyses, i.e. varying explicit model parameters manually (e.g. sensitivity of diagnostic test) lead to evaluation of different cases of the outcome estimates related to the value of the model parameter estimate. › Probabilistic sensitivity analyses, i.e. allowing variation in all model parameters (i.e. parameters with a statistical distribution), lead to confidence intervals of all outcome estimates (e.g. impact and cost-effectiveness).	[42] [49] [49, 50]
3.	**Compare strategies** › Compare use of the prediction model to care as usual › Compare different strategies, for example different starting moments for screening, different screening intervals, or different risk thresholds for treatment selection. › Potentially optimize risk thresholds to determine the optimal impact of a prediction model instead of the average impact.	[24, 46, 47]

Empirical Examples

Below we illustrate the possibilities and the interpretation of decision analytic approaches for estimating the impact of prediction models, using two empirical examples, one for a prognostic and one for a diagnostic model. The first example concerns the impact of using a risk prediction model for coronary heart disease (CHD). To determine the indication for preventive drug treatment in a healthy general population, information from carotid ultrasound is added to the conventionally used Framingham CHD risk prediction model (without this new test added) [39, 51]. The second example concerns the impact of a diagnostic strategy including D-dimer testing to rule out deep venous thrombosis (DVT) in patients suspected of this disorder, as compared to care as usual, i.e. all patients are referred for ultrasound [52-54].

Case study 1: Impact of a prognostic prediction model

The impact of an updated CHD risk prediction model (PM), incorporating conventional predictors plus results from ultrasound on the carotid intima-media thickness (CIMT-PM; index group), was estimated using a decision analytical technique, as compared with usual care in which the conventional Framingham CHD risk prediction model was used (FRS-PM; control group) [39]. In the usual care strategy, individuals were classified conform prevailing guidelines into very low (<5%), low (5-10%), intermediate (10-20%), and high (>20%) 10-year CHD risk according to FRS-PM. Individuals at *intermediate* and *high* risk were considered eligible for preventive pharmacological treatment (e.g. lipid lowering or blood pressure lowering

drugs) to reduce CHD risk. Individuals in the *very low* and *low* risk group did not receive any treatment. In the index group, it was considered that the individuals with an estimated intermediate and high risk (according to FRS-PM) were further reclassified according to their estimated 10-year CHD probabilities after using the CIMT-PM.

Evidence on the incidence and CHD event occurrences in each of the four risk categories, and the reclassification table of CIMT-PM and FRS-PM, was then used as input for the observed health outcomes in each strategy, and included in a so-called Markov model (See appendix A).

In the used Markov model, the following different outcomes (health states) were defined: *Healthy, First Myocardial Infarction (MI), Second MI, Ischemic Stroke, Hemorrhagic Stroke, Gastrointestinal Bleeding,* and *Death*. Accordingly, both the intended and unintended effects (i.e. bleedings) of the interventions were addressed, as well as the consequences of a CHD.

Costs due to treatment and occurrence of all potential events were included as well as the cost of a single CIMT-measurement. Additionally, non-adherence to medication was taken into account by shrinking the actual treatment effect with 50%. A probabilistic sensitivity analysis was applied to investigate the uncertainty in model outcomes.

A hypothetical cohort of 100,000 individuals (age 50 – 59 years), i.e. men or women, was then simulated, where individuals had their risk estimated according to FRS-PM, or FRS-PM followed by CIMT-PM, and were then treated if their final 10-year CHD risk estimated exceeded 10%. Individuals entered the model (in the *Healthy* state) and every year after start of the simulation could move to other health states.

The impact of adding the CIMT-PH prediction model to the FRS-PM prediction model, along with the consequences of changes in the prescription of pharmacological treatment for reducing CHD risk was evaluated, over a 10, 20, and 30 year time horizon.

Ultimately, the two prediction models were compared on their: 1. fractions of treated individuals; 2. Frequency of CHD events; 3. Frequency of adverse events from treatment; 4. Average costs; 5. Average health outcomes expressed in quality-adjusted-life-years (QALYs). All estimates were derived separately for men and women.

The fraction of treated men in both scenarios was similar whereas the fraction of treated women was slightly higher in the CIMT-PM scenario compared to the FRS-PH scenario. One in two men and one in eleven women had an intermediate or high risk, and were reclassified by the new index strategy. Reclassification of these individuals with CIMT-PM would result in a different preventive treatment decisions in 15.8% of these men, and in 16.0% of these women.

Using a 10 year time horizon, health outcomes in terms of QALYs were similar when using FRS-PM or CIMT-PM, whereas CIMT-PM resulted in +0.02 QALYs on average in men, and +0.05 QALYs on average in women, when a 30 year time horizon was used. Using CIMT-PM resulted in additional costs of $100 for men, and cost saving of $200–300 for women, compared to using only FRS-PM. The strategy using CIMT-PM had a 25% probability of being cost-effective within 10 years for men, and 87% probability for women. For a time horizon of 30 years, the probabilities increased to 93% and 98% for men and women respectively.

The authors concluded that applying CIMT measurements resulted in a small health benefit for men and women Further analysis with more optimistic (e.g. relative risk of 0.55) or pessimistic (e.g. relative risk of 0.80) estimates of the effectiveness of the pharmacological treatments, indicated that these effectiveness estimates had only very limited influence on results and did not alter the conclusions.

Interpretation:

This model-based study showed that measurements of CIMT to stratify patients at potential high risk for cardiovascular disease is promising in terms of health benefits but it takes time for the health benefits to outweigh the costs of CIMT measurements. Therefore, implementing CIMT measurements in combination with a prediction model in clinical practice is not favourable currently. Nevertheless, when the costs of a single CIMT measurement would decrease, or the treatment effectiveness would increase, a strategy with CIMT measurements to select patients for preventive CVD treatment may become very promising. Investigating the influence of a more opportunistic or conservative effect of treatment within a RCT would have taken many years, since CVD events occur over a long-time period.

Case study 2: Impact of a diagnostic prediction model

Recently, rapid point-of-care D-dimer assays were introduced which, in combination with a diagnostic prediction model, made it possible to realize an efficient diagnostic work-up for patients suspected of having deep venous thrombosis (DVT) in primary care. The impact of a diagnostic prediction model (AMUSE model) combined with a point-of-care D-dimer test, was assessed compared to care as usual (UC) to stratify patient suspected of DVT [53].

Usual care was defined as referral to the hospital, i.e. including an ultrasound, for all primary care patients suspected of DVT *(hospital strategy)*. Two different diagnostic strategies were compared with this hospital strategy. The first strategy was applying the AMUSE model in primary care, which was defined as stratifying primary care patients suspected of DVT with an absolute threshold of 4, i.e. patients were classified as high risk when their AMUSE score was ≥ 4 *(AMUSE strategy)*. These high risk individuals were referred to the hospital for ultrasound and received anticoagulant treatment. The second diagnostic strategy was referral to the hospital for all primary care patients suspected of DVT with application of the AMUSE model at the emergency department (instead of the primary care practice). Subsequently, D-dimer test was performed in patients with an AMUSE score of 2 or higher and an additional ultrasound was evaluated if the D-dimer test was elevated *(hospital AMUSE strategy)*.

A Markov model was developed to simulate the course of events of patients suspected of DVT (see appendix B). The model included different health states; no previous DVT, post venous thrombosis (post VTE), post thrombotic syndrome (PTS), central nervous system bleeding (CNS), and death.

The events that could occur were; DVT, pulmonary embolism (PE), major (gastrointestinal) bleeding and CNS bleed (see also appendix B). Evidence on the probabilities of the transition between health states and on the consequences of the treatment, was based on the literature [54].

The costs of the three strategies included costs of medical care (i.e. events, potential ER visit and/or ultrasound), travel costs to the GP and/or hospital, and costs to perform a D-dimer test (i.e. cost of the test, GP time. Adherence to medication was not taken into account in the analyses. Uncertainty of the parameters was assessed with probabilistic sensitivity analyses.

A hypothetical cohort of 1,002 patients, i.e. the number of patients from a previous empirical longitudinal diagnostic study in the same field, in which the safety and efficiency of the AMUSE strategy in primary care was evaluated, was simulated in which individuals could move between the health states every 6 months. The impact of these three different strategies on health effects, costs, and cost-effectiveness was assessed over a 5-year time horizon.

The *AMUSE strategy* in primary care had both slightly lower health effects and costs compared to the other strategies, i.e. 3.853 QALYs and €3,589. The *hospital AMUSE strategy* has similar health effects, i.e. 3.586 QALYs, and lower costs, i.e. €3,727 versus €3,768, compared with the *hospital strategy*. The *AMUSE strategy* compared with the *hospital AMUSE strategy* results in an average health loss of -0.003 QALY and cost saving of € 138, and an estimated ICER of 55,753 €/QALY.

The incremental difference in health benefits and costs of the *hospital AMUSE strategy* compared with the *hospital strategy* is a health benefit of 0.0005 QALYs and cost saving of €41, i.e. ICER of 89, 956 €/QALY. Assuming a willingness-to-pay threshold of 80,000 €/QALY, the *AMUSE strategy* is preferred above the other strategies. In other words, referral to ultrasound for patients, suspected of DVT, with a AMUSE score ≥ 4 is preferred above referral of all patients to ultrasound (*hospital strategy*) or referral of patients with an AMUSE score ≥ 2 and an elevated D-timer test.

Authors concluded that a diagnostic management based on the clinical decision rule and D-dimer test is likely to be cost-effective compared with the other strategies [53]. Further analysis with lower and higher costs for the incorporated health states did not have an influence on the results. However, analyses with a more conservative sensitivity of the D-dimer test (i.e. 0.9032 versus 0.9265) showed a slightly decrease in health effects and costs.

Interpretation:

This model-based study showed a comparison of three strategies to diagnose DVT in clinical practice and the influence of different parameters on the results within just one investigation. Comparing these three strategies in terms of (cost-)effectiveness and investigating the influence of uncertain parameters on (cost-)effectiveness in a RCT would have required substantial time, funds, and inclusion of many patients.

Concluding remarks

Assessments of the impact or effectiveness of the use of risk guided management using prediction models are scarce. This is simply due to the fact that the typical required comparative longitudinal, randomised study design is often considered too time consuming, too cumbersome, and also limited in the number of strategies that can be compared. Moreover, there are simply too many prediction models to assess their impact with a randomised study.

This paper aims to increase the awareness that a decision analytic approach can be a valid and useful alternative to performing randomized trials for the assessment of the effectiveness and value of prediction models to guide clinical decision making [19, 24]. Decision analytic approaches are relatively limited in costs and time, although it does require a multidisciplinary team with clinical, epidemiological and mathematical expertise [55]. Impact assessment of prediction models by decision analytic models will become even more relevant with the advancement of artificial intelligence techniques that are increasingly used for the development of prediction models.

Still, challenges do remain when decision analytic approaches are applied to assess the impact of prediction models, as summarised in Box 1 [26]. Furthermore, comprehensive validation of decision analytic model is recommended to ensure the validity of the model itself and the model outcomes [56, 57]. Nevertheless, we believe that before jumping into the design and conduct of an empirical long term, large scale randomised comparative study in order to quantify the impact of the use of prediction models on actual health outcomes and cost-effectiveness of care, a linked evidence approach using some type of decision analytical model is useful. Such an approach can indicate which models should never be empirically studied on their impact or cost-effectiveness, despite good performance in several validation studies. For example, because care-as-usual is already very effective, the long term effects of subsequent treatment are too low or because costs and adverse effects of the use of the prediction model outweighs its benefits. Furthermore, when an empirical study is planned based on the results of a preceding modelling approach, the design and analysis of such an empirical study can be fed by the results of the model study. For example, the modelling study may indicate which subgroups benefit more from a risk based management approach or what the minimal adherence to the predicted risks and subsequent treatments should be.

We conclude therefore that when the aim is to assess the impact or effectiveness of the use of a validated and promising prediction model to guide clinical decision making it is better to model first and to trial later.

Box 1: Advantages and disadvantages of decision analytical modelling techniques to quantify the impact and cost-effectiveness of the use of prediction models to guide clinical management.

Advantages

1. Impact assessment of the prediction model can be performed quickly and at low cost.
   - For example, when the number of (low risk) individuals that needs to be included to observe actual events in that subgroup is high, an empirical study would be very large and thus expensive.
2. Decision modelling techniques can provide estimates of long term effects of the use of the model on individual health outcomes and costs-effectiveness of care that would not be observable in empirical studies.
   - For example, for prognostic models predicting 10 year or even life time outcome risks, any empirical study on the effectiveness of the use of such model on the occurrence of these outcomes, would be very cumbersome, expensive and indeed perhaps unfeasible.
3. Updating a model-based impact assessment is straightforward and requires relatively little time.
   - For example, when a prediction model is updated with a new predictor, or a new prediction model is developed and validated, an empirical study would be expensive whereas a decision analytic model can be re-used and updated.
4. Different strategies can be compared (incrementally) regarding the actual use of a prediction model.
   - For example, one can compare the impact of the use of different risk thresholds for treatment administration or abstention; one can explore the impact of many different treatments, at many different treatment adherence rates and in many different patient subgroups.
5. Different strategies can be compared (incrementally) regarding one patient group.
   - For example, one can compare the impact of the use of different prediction models in one patient group; one can explore the use of different prediction models in different patient subgroups.

Disadvantages

1. Evidence is required on which predicted risks actually should be used to guide treatment management decisions in practice.
   - For example, there are 350 prediction models for CVD in primary care, each with different predictors in the risk equation and different CVD event types in the (composite) endpoints, resulting in a wide variation of predicted risks.
2. Uncertainty in model parameters is often unknown, hence assumptions need to be made, which may lead to incorrect impact outcome estimates.
   - For example, uncertainty in the risk estimates, i.e. output of the prediction model, may lead to incorrect risk classification, treatment decisions and therefore health loss or extra costs.
3. Estimating health outcomes requires evidence on the effects of the possible treatments, preferably as function of the baseline risk of individuals.
   - For example, overall estimates of the effects of a treatment in an entire patient population, may not apply to individuals in every risk category.
4. Decision analytic modelling requires assumptions on behavioural changes that individuals would make following application of a prediction model.
   - For example, effect of treatment is influenced by the adherence to the treatment. However, (long-term) evidence on adherence to medication or lifestyle changes is lacking.
5. When the targeted outcome is a composite outcome, this composite outcome needs to be unravelled into all its separate components as each event typically has a different impact on health outcomes and costs.
   - For example, a CVD event comprising stroke, myocardial infarction, and heart failure may require different treatments, with different effects, and each event may have different consequences in terms of quality of life and health care costs.

Box 2: Context in which decision analytic impact assessment of the use of prediction models in daily practice is worthwhile.

- Decisions regarding the risk categories used for classification and subsequent clinical management have not yet been made, for example, the impact of using different risk thresholds for treatment indication or withholding treatment, or the impact of using a rather different risk threshold than commonly considered.
- High quality evidence on how predicted risk estimates influence clinical management decisions is available from clinical studies or guidelines.
- High quality evidence on how clinical management or treatment effect individual health outcomes is readily available from (ideally randomised) comparative intervention studies.
- To guide the actual design of future randomised prediction model impact studies, by selecting those strategies (e.g. risk thresholds) that are most promising, identifying which subgroups may benefit most from which risk based treatments, or identifying which adherence to predicted risks and subsequent treatments is minimally required.

REFERENCES

1. Justice, A.C., K.E. Covinsky, and J.A. Berlin, *Assessing the generalizability of prognostic information.* Ann Intern Med, 1999. 130(6): p. 515-24.
2. Reilly, B.M. and A.T. Evans, *Translating clinical research into clinical practice: impact of using prediction rules to make decisions.* Ann Intern Med, 2006. 144(3): p. 201-9.
3. Moons, K.G., et al., *Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker.* Heart, 2012. 98(9): p. 683-90.
4. Moons, K.G., et al., *Risk prediction models: II. External validation, model updating, and impact assessment.* Heart, 2012. 98(9): p. 691-8.
5. Steyerberg, E.W., et al., *Prognosis Research Strategy (PROGRESS) 3: prognostic model research.* PLoS Med, 2013. 10(2): p. e1001381.
6. Steyerberg, E.W., *Clinical Prediction Models*. Statistics for Biology and Health. New York: Springer-Verlag.
7. Harrell, F., *Regression Modeling Strategies*. 2001, New York: Springer.
8. Moons, K.G., et al., *Prognosis and prognostic research: application and impact of prognostic models in clinical practice.* BMJ, 2009. 338: p. b606.
9. Ferrante di Ruffano, L., et al., *Assessing the value of diagnostic tests: a framework for designing and evaluating trials.* BMJ, 2012. 344: p. e686.
10. Bossuyt, P.M., et al., *Beyond diagnostic accuracy: the clinical utility of diagnostic tests.* Clin Chem, 2012. 58(12): p. 1636-43.
11. Moons, K.G., et al., *Prognosis and prognostic research: what, why, and how?* BMJ, 2009. 338: p. b375.
12. de Hoop, E., et al., *The need to balance merits and limitations from different disciplines when considering the stepped wedge cluster randomized trial design.* Bmc Medical Research Methodology, 2015. 15.
13. Perel, P., et al., *Systematic review of prognostic models in traumatic brain injury.* BMC Med Inform Decis Mak, 2006. 6: p. 38.
14. Altman, D.G., *Prognostic models: a methodological framework and review of models for breast cancer.* Cancer Invest, 2009. 27(3): p. 235-43.
15. Beswick, A.D., et al., in *A Systematic Review of Risk Scoring Methods and Clinical Decision Aids Used in the Primary Prevention of Coronary Heart Disease (Supplement)*. 2008: London.
16. Damen, J.A., et al., *Prediction models for cardiovascular disease risk in the general population: systematic review.* BMJ, 2016. 353: p. i2416.
17. van Dieren, S., et al., *Prediction models for the risk of cardiovascular disease in patients with type 2 diabetes: a systematic review.* Heart, 2012. 98(5): p. 360-9.
18. Collins, G.S., et al., *Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting.* BMC Med, 2011. 9: p. 103.

19. Schaafsma, J.D., et al., *Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness.* J Clin Epidemiol, 2009. 62(12): p. 1248-52.

20. Hunink, M.G. and G.P. Krestin, *Study design for concurrent development, assessment, and implementation of new diagnostic imaging technology.* Radiology, 2002. 222(3): p. 604-14.

21. Koffijberg, H., B. van Zaane, and K.G. Moons, *From accuracy to patient outcome and cost-effectiveness evaluations of diagnostic tests and biomarkers: an exemplary modelling study.* Bmc Medical Research Methodology, 2013. 13: p. 12.

22. Merlin, T., *The use of the 'linked evidence approach' to guide policy on the reimbursement of personalized medicines.* Personalized Medicine, 2014. 11(4): p. 435-448.

23. Merlin, T., et al., *The "linked evidence approach" to assess medical tests: a critical analysis.* Int J Technol Assess Health Care, 2013. 29(3): p. 343-50.

24. van Giessen, A., et al., *An alternative approach identified optimal risk thresholds for treatment indication: an illustration in coronary heart disease.* J Clin Epidemiol, 2018. 94: p. 122-131.

25. Greving, J.P., et al., *Statin treatment for primary prevention of vascular disease: whom to treat? Cost-effectiveness analysis.* BMJ, 2011. 342: p. d1672.

26. van Giessen, A., et al., *Systematic Review of Health Economic Impact Evaluations of Risk Prediction Models: Stop Developing, Start Evaluating.* Value Health, 2017. 20(4): p. 718-726.

27. ISPOR. 2018 [cited 2018 10/09]; Synthesizing Health Care Evidence - Modelling/Decision Analysis Methods]. Available from: https://www.ispor.org/heor-resources/more-heor-resources/outcomes-research-guidelines-index.

28. Roberts, M., et al., *Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2.* Med Decis Making, 2012. 32(5): p. 678-89.

29. Weinstein, M.C., et al., *Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies.* Value Health, 2003. 6(1): p. 9-17.

30. Philips, Z., et al., *Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment.* Pharmacoeconomics, 2006. 24(4): p. 355-71.

31. Evans, D., *Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions.* J Clin Nurs, 2003. 12(1): p. 77-84.

32. OCEBM, *The Oxford 2011 levels of evidence* L.o.E.W. Group, Editor 2011, Oxford Centre for Evidence-Based Medicine.

33. Hingorani, A.D., et al., *Prognosis research strategy (PROGRESS) 4: stratified medicine research.* BMJ, 2013. 346: p. e5793.

34. Kappen, T.H., et al., *Barriers and facilitators perceived by physicians when using prediction models in practice.* J Clin Epidemiol, 2016. 70: p. 136-45.

35. Janssen, K.J., et al., *Updating methods improved the performance of a clinical prediction model in new patients.* J Clin Epidemiol, 2008. 61(1): p. 76-86.

36. Nederlands Huisartsen Genootschap. *Cardiovascular risk management*. 2012; Available from: https://www.nhg.org/standaarden/volledig/cardiovasculair-risicomanagement.

37. Blanken, M.O., et al., *Cost-effectiveness of rule-based immunoprophylaxis against respiratory syncytial virus infections in preterm infants.* Eur J Pediatr, 2018. 177(1): p. 133-144.

38. van Giessen, A., et al., *Cost-effectiveness of screening strategies to detect heart failure in patients with type 2 diabetes.* Cardiovasc Diabetol, 2016. 15: p. 48.

39. den Ruijter, H.M., et al., *Long-term health benefits and costs of measurement of carotid intima-media thickness in prevention of coronary heart disease.* J Hypertens, 2013. 31(4): p. 782-90.

40. Degeling, K., et al., *Matching the model with the evidence: comparing discrete event simulation and state-transition modeling for time-to-event predictions in a cost-effectiveness analysis of treatment in metastatic colorectal cancer patients.* Cancer Epidemiol, 2018. 57: p. 60-67.

41. Kip, M.M.A., et al., *The cost-utility of point-of-care troponin testing to diagnose acute coronary syndrome in primary care.* BMC Cardiovasc Disord, 2017. 17(1): p. 213.

42. Briggs, A.H., et al., *Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6.* Med Decis Making, 2012. 32(5): p. 722-32.

43. Claxton, K., et al., *How to estimate the health benefits of additional research and changing clinical practice.* BMJ, 2015. 351: p. h5987.

44. Koffijberg, H., et al., *Value of Information Choices that Influence Estimates: A Systematic Review of Prevailing Considerations.* Med Decis Making, 2018. 38(7): p. 888-900.

45. Tuffaha, H.W., L.G. Gordon, and P.A. Scuffham, *Value of information analysis in healthcare: a review of principles and applications.* J Med Econ, 2014. 17(6): p. 377-83.

46. Buxton, M.J., et al., *Modelling in economic evaluation: an unavoidable fact of life.* Health Econ, 1997. 6(3): p. 217-27.

47. Ferrante di Ruffano, L., et al., *A capture-recapture analysis demonstrated that randomized controlled trials evaluating the impact of diagnostic tests on patient outcomes are rare.* J Clin Epidemiol, 2012. 65(3): p. 282-7.

48. Nayak, S., M.S. Roberts, and S.L. Greenspan, *Cost-effectiveness of different screening strategies for osteoporosis in postmenopausal women.* Ann Intern Med, 2011. 155(11): p. 751-61.

49. *Guideline for economic evaluations in healthcare* 2016, Health Care Insurance Board: The Netherlands.

50. Drummond, M.S., M.J.; Claxton, K.; Stoddart, G.L.; Torrance, G.W., *Methods for the economic Evaluation of Health Care Programmes*. 4th ed. 2015, Oxford, United Kingdom: Oxford University Press.

51. Nambi, V., et al., *Carotid intima-media thickness and presence or absence of plaque improves prediction of coronary heart disease risk: the ARIC (Atherosclerosis Risk In Communities) study.* J Am Coll Cardiol, 2010. 55(15): p. 1600-7.

52. Oudega, R., K.G. Moons, and A.W. Hoes, *Ruling out deep venous thrombosis in primary care. A simple diagnostic algorithm including D-dimer testing.* Thromb Haemost, 2005. 94(1): p. 200-5.

53. Ten Cate-Hoek, A.J., et al., *Cost-effectiveness of ruling out deep venous thrombosis in primary care versus care as usual.* J Thromb Haemost, 2009. 7(12): p. 2042-9.

54. Buller, H.R., et al., *Safely ruling out deep venous thrombosis in primary care.* Ann Intern Med, 2009. 150(4): p. 229-35.

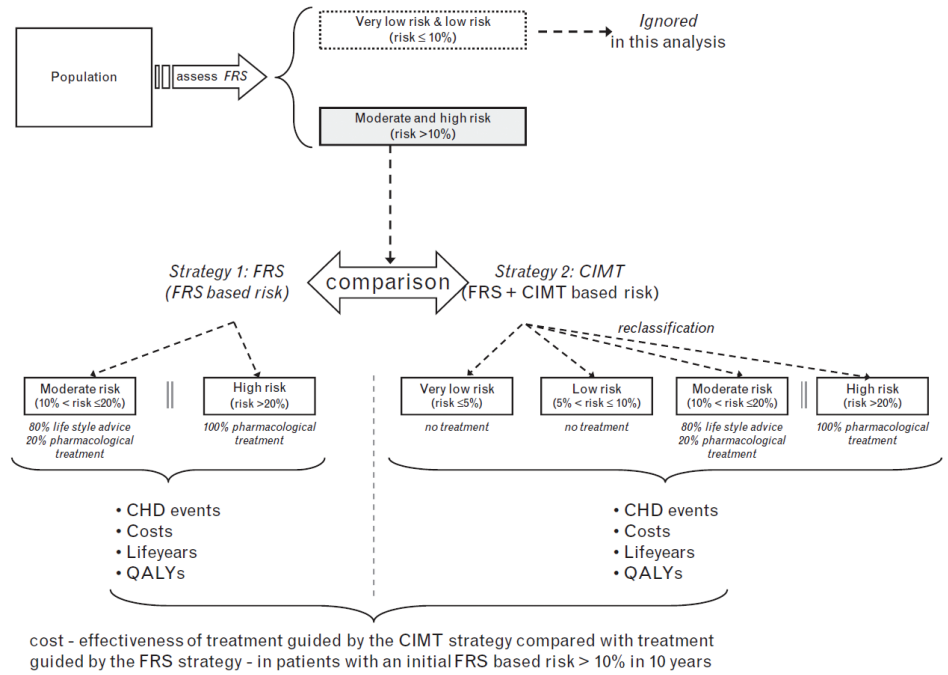
55. Sadique, Z., et al., *An integrated approach to evaluating alternative risk prediction strategies: a case study comparing alternative approaches for preventing invasive fungal disease.* Value Health, 2013. 16(8): p. 1111-22.

56. Vemer, P., et al., *AdViSHE: A Validation-Assessment Tool of Health-Economic Models for Decision Makers and Model Users.* Pharmacoeconomics, 2016. 34(4): p. 349-61.

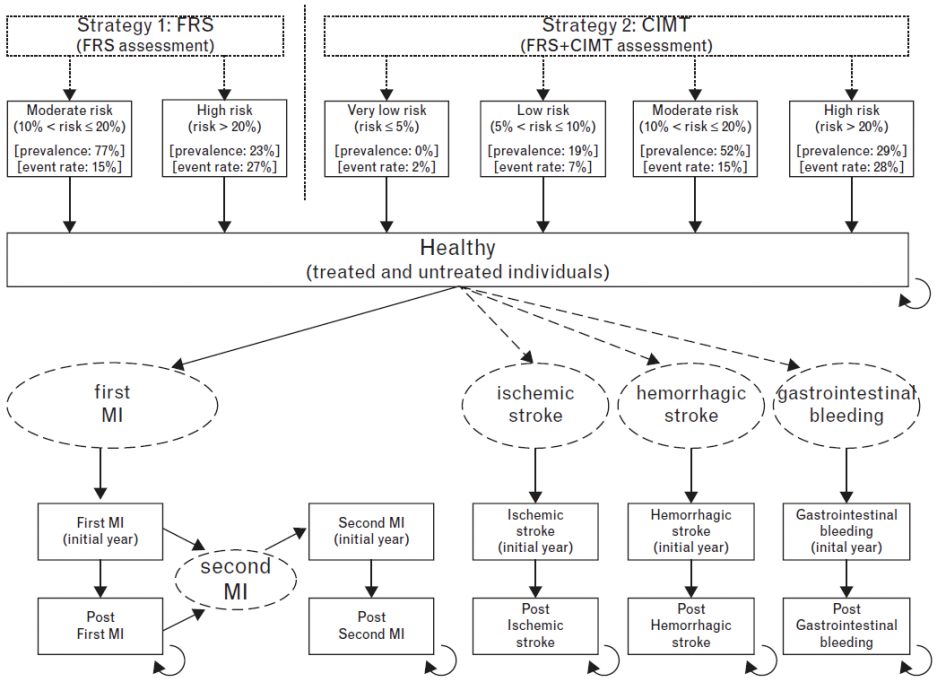
57. Eddy, D.M., et al., *Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7.* Med Decis Making, 2012. 32(5): p. 733-43.

APPENDICES

A. EMPIRICAL CASE STUDY 1

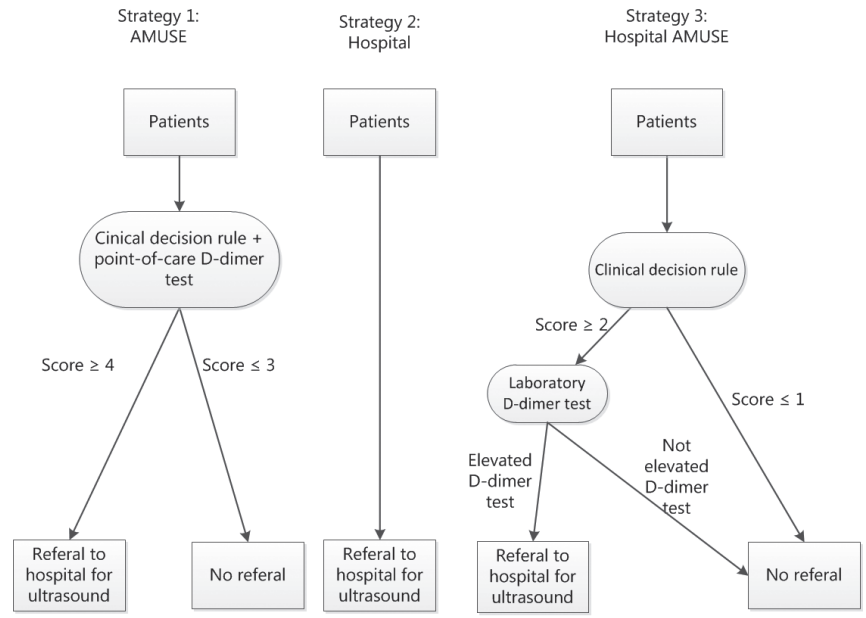


Appendix A - Figure 1: Flowchart of the analytic procedure [39].
FRS = Framingham Risk SCORE; CIMT = carotid intima-media thickness; CHD = coronary heart disease; QALY = Quality-Adjusted-Life-Year

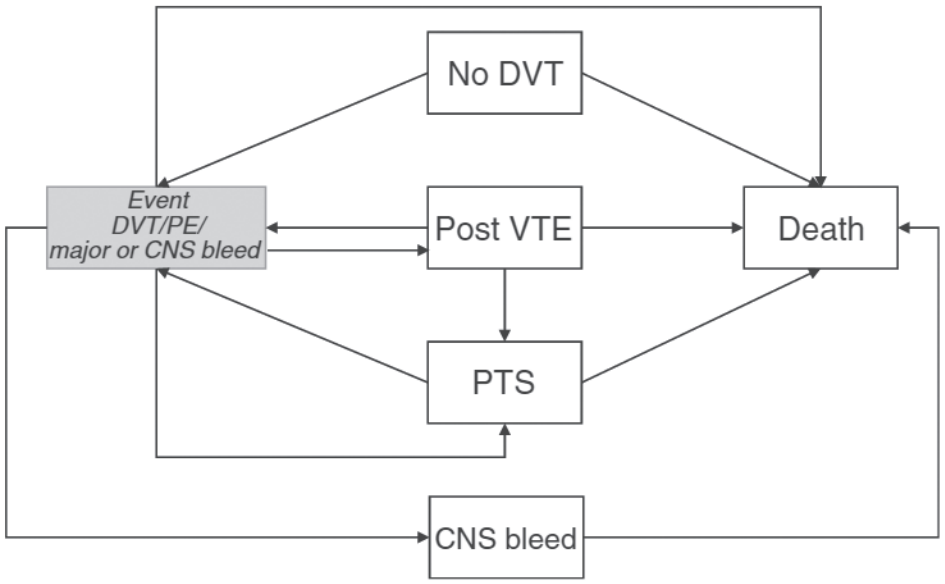


Appendix A - Figure 2: Representation of the Markov model [39].
FRS = Framingham Risk SCORE; MI; myocardial infarction.
Health states are presented by the rectangular boxes and event by the ellipses. High-risk individuals (i.e. risk estimate>20%) received statin treatment and could experience side-effects, i.e. mild or severe complications including death.

B. EMPIRICAL CASE STUDY 2



Appendix B - Figure 1: Flowchart of the process of referral in the different strategies.



Appendix B - Figure 2: Structure of the Markov model [53].

DVT = deep venous thrombosis; post VTE = post venous thrombosis; PTS = post thrombotic syndrome; CNS; central nervous system bleeding.

The events that could occur were; DVT, pulmonary embolism (PE), major (gastrointestinal) bleeding and CNS bleed. Evidence on the probabilities of the transition between health states and on the consequences of the treatment, was based on the literature [54].

A.

Appendices

SUMMARY

Impact assessment of prediction models is complex and rarely performed. The included content in this thesis aims to explore the challenges and solutions for the impact assessment of prediction models. This thesis is separated in two parts; the challenges of different used endpoints in competing prediction models and the challenges in the design of impact assessment of prediction models.

Different prediction models may predict multiple endpoints, using a so-called composite endpoint. The predicted risks are thus related to the combination of endpoints, which each in turn have different associated impact, i.e. health outcomes and costs, and may require different decision making and management. In **Chapter 2,** we explored the extent of the differences in definitions of composite endpoints and assessed how these differences influence estimates of cardiovascular disease (CVD) risk and burden. Data from a Dutch cohort study was used to calculate 10-year risks and health burden according to four CVD prediction models; ATP-III, Framingham (FRS), Pooled Cohort Equations (PCE) and SCORE. The investigated CVD risk prediction models showed huge variation in definition of composite endpoints, 10-year CVD risks and associated health burden. Therefore, health consequences related to predicted risks cannot be readily compared across prediction models, and estimates of burden of disease depend crucially on the prediction model used.

CVD prevention is commonly focused on providing individuals at high predicted CVD risk, i.e. often elderly individuals, with preventive medication. However, the lifelong (preventable) consequences of CVD events may be larger in younger individuals. In **Chapter 3**, we investigated if the health benefits from preventive treatment for CVD might increase when the selection strategy was burden rather than risk based. Data from three Dutch cohorts was combined and used to calculate 10-year CVD risk and burden estimates according to the Framingham Global Risk Score (FRS). When individuals were selected based on their expected CVD burden, rather than their expected CVD risk, the additional health benefits estimates from preventive treatment were larger. Selecting individuals for preventive treatment based on their expected CVD burden will provide more younger and less older individuals with treatment, and will reduce the overall CVD burden. Improvement of the selection approach of individuals can help to further reduce the CVD burden.

Prediction models are often used to recommend treatment for individuals with a risk estimate above a certain risk threshold, and no treatment otherwise. However, when there are competing prediction models for the same targeted population, the same individuals can be classified differently according to their predicted risk by different models. In order to investigate if the use of different prediction models may actually lead to variation of selection of high risk individuals, four CVD prediction models and a Dutch cohort (similar as in **Chapter 2**) are used to investigate this hypothesis in **Chapter 4**. Besides the variation of the predicted risks according to the four models, the recommended treatment thresholds are also different. Given the variation in predicted risks and treatment thresholds, the percentage of selected individuals for preventive treatment varies between 0.2% and 14.9%. Widely used CVD prediction models vary substantially regarding their outcomes and associated absolute risk estimates. Consequently, absolute predicted 10-year risks from different prediction models cannot be compared directly. In addition, treatment decisions often depend on which prediction model is applied and its recommended risk threshold, introducing unwanted practice variation into risk-based preventive strategies for CVD.

To assess the impact of prediction models, using a long-term randomized study may not always be feasible. For example, manifestation of CVD at old(er) age can sometimes be traced back to non-traditional risk factor levels at young(er) age. To reduce the burden of CVD, it may be beneficial to intervene at young(er) age. However long-term impact assessment of an early preventive intervention is almost infeasible in a randomized trial due to the long-term period. Conversely a modelling approach may be more feasible. As evidence on the impact of prevention at young age is very limited, in **Chapter 5**, we studied which parameters and assumptions influence the expected long-term benefits of preventive strategies at young age most. A micro-simulation model with a lifetime horizon is developed to explore the influence of parameters on the long-term impact assessment of preventive strategies for CVD. Women with hypertensive pregnancy disorder (HPD) are used as a case study and screening at 30, 40 or 50 years is investigated as preventive strategy. Additionally, a Value of Information analysis and net benefit regression are performed to identify parameters that warrant future further research and that have a large influence on the health benefits. Results show that in the case of prevention of CVD in women with HPD, it would be crucial to further study risk reduction of preventive treatment and the long-term adherence rate of medication in women.

In **Chapter 6**, we investigated whether early preventive CVD risk screening combined with risk-based lifestyle intervention in women with previous preeclampsia are beneficial and cost-effective. A micro-simulation model (similar as in **Chapter 5**), was applied to two Dutch datasets measuring CVD risk parameters in women at different time intervals after pre-eclampsia. Screening was started at the age of 30 or 40 and repeated every 5 years. 10-year CVD risk estimates were calculated according to FRS and two absolute risk thresholds (2% and 5%) were evaluated for treatment selection, i.e. lifestyle interventions (including smoking cessation, weight reduction and increasing physical activity). Results show that early CVD risk screening followed by risk-based lifestyle interventions can improve long-term health outcomes in women with a history of preeclampsia. However, the cost-effectiveness of establishing a lifelong cardiovascular prevention program for women starting early after experiencing preeclampsia by risk-based lifestyle advice alone is relatively unfavourable. A combination of risk-based lifestyle advice plus medical therapy may be more beneficial.

The impact of a prediction model on health outcomes of patients is not only determined by the performance of the model, but also by care providers' compliance with management recommendations based on the predictions. In **Chapter 7**, we assessed the impact of compliance with management recommendations from a prediction model, on the number of missed major adverse cardiac (MACE) events in an emergency setting. A decision tree model was developed to compare the application of the prediction model (i.e. HEART score) to usual care. Impact on patient outcomes (missed MACE events) and costs was assessed for scenarios in which the degree of compliance with HEART score management recommendations, and informed deviation (ID) from these recommendations, were varied. Results show that the impact of using the HEART score in a clinical setting is influenced by the interplay of compliance and informed deviation. Scenario analysis showed that 100% compliance (and 0% ID) reduces the number of missed MACE events compared to usual care. When ID gets influence, to at least 25% (with any compliance above 50%), missed events are reduced.

Decision analytic modelling is a useful approach to assess the potential influence of certain factors on the impact of risk prediction model, in case there is limited data available on key factors such as compliance. This approach could provide evidence for deciding whether or not to conduct a subsequent clinical impact study, or to inform the design and conduct of such studies.

Assessing the impact of prediction models in a randomised comparative design are often time consuming, costly and sometimes infeasible if the predicted outcome occurs over a long time period (e.g. 30 years). An alternative is a decision analytic modelling approach, where evidence of the predictive accuracy of the model and effects of and adherence to subsequent treatments is linked. There are many advantages of using a decision analytic model for the impact assessment of prediction models, there are also disadvantages of using a decision analytic model. An overview of the advantages and disadvantages of a model-based approach to assess the impact of prediction models, without performing a prospective comparative study first, is presented in **Chapter 8.** Two examples of impact studies of a prediction model with a model-based approach are described in detail, and guidance on how to perform a model-based impact assessment is discussed.

SAMENVATTING

Het bepalen van de impact van predictiemodellen is complex en wordt helaas nog weinig gedaan. De inhoud van dit proefschrift beschrijft de uitdagingen en mogelijke oplossingen bij het bepalen van de impact van predictiemodellen. Het proefschrift is opgesplitst in twee delen; de uitdagingen voortkomend uit het gebruik van verschillende eindpunten in concurrerende predictiemodellen en de uitdagingen voortkomend uit de verschillende manieren voor het bepalen van de impact van predictiemodellen.

Verschillende predictiemodellen kunnen verschillende en meerdere eindpunten voorspellen, zogenaamde 'samengestelde eindpunten'. De voorspelde risico's van een predictiemodel zijn dus gerelateerd aan de samenstelling van de gebruikte eindpunten, ieder met mogelijk verschillende consequenties, zowel in termen van gezondheidsuitkomsten als kosten, en mogelijk verschillende besluitvorming. In **Hoofdstuk 2** is onderzocht hoe het verschil in de definities van samengestelde eindpunten en hoe deze verschillen de risicoschatting voor, en de ziektelast van hart- en vaatziekten (HVZ) beïnvloeden. Data van een Nederlands cohort is gebruikt om 10-jaar risico's en bijbehorende ziektelast van HVZ te schatten op basis van vier verschillende predictiemodellen; ATP-III, Framhingham (FRS), Pooled Cohort Equations (PCE) en SCORE. De definities van de (samengestelde) eindpunten van deze vier predictiemodellen verschilden flink wat leidde tot zeer uiteenlopende 10-jaar risico's op, en levenslange ziektelasten van HVZ. Hieruit valt te concluderen dat het vergelijken van voorspelde risico's en hun gerelateerde gezondheidseffecten over verschillende predictiemodellen niet evident is. Daarnaast is het schatten van ziektelast op basis van risico's bijna volledig afhankelijk van het gebruikte predictiemodel.

Preventie van HVZ focust meestal op het selecteren en preventief behandelen van mensen met een hoog risico op HVZ. In bijna alle gevallen zijn mensen met een hoog risico mensen met een hoge leeftijd. Echter, de levenslange (en mogelijk te voorkomen) consequenties van HVZ zijn hoger bij jonge mensen. In **Hoofdstuk 3** is onderzocht of de gezondheidswinst van een preventieve behandeling voor HVZ mogelijk hoger zou zijn als de selectie van mensen gebaseerd was op ziektelast in plaats van op risico. Data van drie verschillende Nederlandse cohorten zijn gebruikt om met het FRS predictiemodel de 10-jaar risico's op en levenslange ziektelast van HVZ te schatten,. De gezondheidswinst van een preventieve behandeling was (inderdaad) hoger wanneer mensen geselecteerd werden op ziektelast in plaats van op risico. Het selecteren op basis van ziektelast leidde ertoe dat er meer jonge mensen en minder oudere mensen behandeld werden en dat de totale ziektelast daalde. Het verbeteren van de selectiestrategie kan een hulpmiddel zijn in het verminderen van de HVZ ziektelast .

Voor het schatten van een risico worden vaak predictiemodellen gebruikt, waarbij het risico vervolgens wordt gebruikt voor het voorschrijven van een preventieve behandeling voor mensen van wie het risico boven een bepaalde drempelwaarde ligt. Echter, er zijn verschillende (concurrerende) predictiemodellen voor dezelfde populatie. Dit kan ertoe leiden dat dezelfde mensen anders gegroepeerd worden op basis van hun geschatte risico, wanneer verschillende predictiemodellen gebruikt worden. Vier predictiemodellen (dezelfde als in **Hoofdstuk 2**) en een Nederlandse dataset (dezelfde als in **Hoofdstuk 2**) zijn gebruikt om in **Hoofdstuk 4** te onderzoeken of het gebruik van verschillende predictiemodellen leidt tot variatie in de groep mensen met een hoog risico op HVZ. Naast een grote spreiding in geschatte risico's zijn de aanbevolen drempelwaardes voor behandeling volgens de vier modellen ook verschillend. Vanwege de verschillen in zowel risico's als drempelwaardes varieert het percentage mensen met een hoog risico op HVZ tussen de 0.2% en 14.9% voor de vier gebruikte predictiemodellen. Predictiemodellen voor HVZ worden gebruikt over de hele wereld maar verschillen onderling aanzienlijk met betrekking tot de samengestelde eindpunten en de daaraan gerelateerde risico's. Dit heeft tot gevolg dat de 10-jaar risico's van verschillende predictiemodellen niet onderling vergeleken kunnen worden. Hierdoor is de beslissing om een individu wel/niet te behandelen sterk afhankelijk van het gebruikte predictiemodel en de bijbehorende drempelwaarde. Oftewel, de groep mensen met een hoog risico op HVZ dat preventieve behandeling krijgt, kan mogelijk variëren per zorginstelling.

Voor het bepalen van de klinische impact van predictiemodellen is het niet altijd mogelijk om een gerandomiseerde studie uit te voeren. Het krijgen van HVZ bij oude(re) mensen bijvoorbeeld, kan soms herleid worden tot bepaalde eigenschappen die sinds jonge leeftijd aanwezig zijn. Ingrijpen en mogelijk aanpassen van deze eigenschappen op jonge leeftijd kan de ziektelast van HVZ op oudere leeftijd verlagen. Echter, het bepalen van de klinische impact van deze vroegtijdige preventieve interventie is vervolgens een lastige opgave en bijna onmogelijk in een gerandomiseerde studie vanwege de lange doorlooptijd. Een aanpak op basis van (wiskundige) modellering is daarentegen een goed alternatief.

Bewijsstukken omtrent de klinische impact van preventieve interventies op jonge leeftijd zijn vaak beperkt. In **Hoofstuk 5** is onderzocht welke parameters en variabelen de verwachte (lange-termijn) effecten van een preventieve behandeling het meest beïnvloeden. Hiervoor is een model ontwikkeld om de (levenslange) effecten van een preventieve behandeling voor HVZ te simuleren. Vrouwen, die tijdens een zwangerschap een hoge bloeddruk hadden, zijn gebruikt als voorbeeld en zij werden gescreend (en behandeld) op 30-, 40-, en 50-jarige leeftijd. Er is een tweetal analyses uitgevoerd om de parameters met de grootste invloed op de gezondheidsuitkomsten te identificeren. In het geval van vrouwen. Die een hoge bloeddruk tijdens de zwangerschap doormaakten, lieten de analyses zien dat twee parameters belangrijk zijn om verder te onderzoeken: a) het verlagen van het risico op HVZ door behandeling, oftewel de effectiviteit van de behandeling, en b) het percentage vrouwen dat op lange termijn nog steeds trouw is aan de behandeling.

Het ondervinden van zwangerschapsvergiftiging is een niet-traditionele risico eigenschap die op latere leeftijd kan bijdragen tot het krijgen van HVZ. In **Hoofdstuk 6** is onderzocht of vroege preventieve screening op een hoog risico op HVZ gecombineerd met een bijbehorende levensstijl aanpassing gunstig en kosteneffectief is voor vrouwen met een geschiedenis van zwangerschapsvergiftiging. Hiervoor is een simulatiemodel gebruikt (dezelfde als in Hoofdstuk 5) in combinatie met twee Nederlandse datasets. Deze datasets bevatten de metingen van risico eigenschappen voor HVZ (gedaan op meerdere tijdsmomenten) van vrouwen met een geschiedenis van zwangerschapsvergiftiging. Het screenen van deze vrouwen startte op de leeftijd van 30 of 40 jaar, en werd iedere 5 jaar herhaald. De bijbehorende risico's van deze vrouwen was berekend met het FRS-predictiemodel gecombineerd met twee absolute drempelwaardes (i.e. 2% en 5%) voor het selecteren van "hoog risico" vrouwen. De preventieve behandeling was een levensstijl aanpassing met verschillende facetten, zoals stoppen met roken, afvallen en meer lichamelijke beweging stimuleren. Resultaten van de studie lieten zien dat het vroeg screenen op risico gevolgd door (op risico gebaseerde) levensstijl aanpassingen op de lange termijn gezondheidsuitkomsten kan verbeteren voor vrouwen met een geschiedenis van zwangerschapsvergiftiging. Echter, de kosteneffectiviteit van het levenslang screenen van vrouwen, direct startend na een zwangerschap met zwangerschapsvergiftiging, gevolgd door een levensstijl aanpassing is ongunstig. Een combinatie van levensstijl aanpassing en medicatie zou een beter alternatief zijn, vanuit een kosteneffectiviteitsoogpunt.

De impact van een predictiemodel op klinische uitkomsten van patiënten wordt bepaald door de statistische prestatie van het model, maar ook door de meegaandheid van zorgverleners op de voorgeschreven aanbevelingen met betrekking tot de geschatte risico's. In **Hoofdstuk 7** is onderzocht wat de invloed is van de meegaandheid van zorgverleners op de voorgeschreven aanbevelingen met betrekking tot een predictiemodel (HEART-score model). Dit predictiemodel voorspelt de kans op een grote cardiale voorval (MACE-event) op de spoedeisende hulp. Met behulp van een zelf ontwikkeld beslismodel is het gebruik van de HEART-score model een vergeleken met de gebruikelijke zorg. Voor verschillende scenario's waarin de meegaandheid van zorgverleners en hun geïnformeerde afwijking op de meegaandheid zijn gevarieerd, is de invloed op het aantal gemiste MACE voorvallen en de bijbehorende kosten berekend. De resultaten lieten zien dat de consequenties van het gebruik van het predictiemodel beïnvloed worden door een samenspel van meegaandheid en de afwijking op de meegaandheid van zorgverleners. De analyses van de verschillende scenario's lieten zien dat als de meegaandheid 100% (met 0% afwijking) was, dat het aantal gemiste MACE voorvallen daalde vergeleken met de gebruikelijke zorg. Op het moment dat de afwijking (> 25%) een rol ging spelen (met meegaandheid > 50%), daalde ook het aantal gemiste MACE voorvallen. Het modelleren van besluitvorming is een zeer waardevolle methodiek om de potentiële invloed van bepaalde factoren op de klinische impact van een predictiemodel te bepalen. Deze methodiek biedt ook mogelijkheden als er maar weinig data beschikbaar is van belangrijke parameters, zoals de meegaandheid van zorgverleners. Daarbij kan deze methodiek handvatten bieden bij het besluit om wel/niet een opeenvolgend klinisch onderzoek te gaan uitvoeren, of bij het ontwerp van zo'n klinisch onderzoek.

Het bepalen van de klinische impact van een predictiemodel in een gerandomiseerd ontwerp kost vaak veel tijd en geld en is soms onmogelijk als de voorspelde uitkomst pas over vele jaren plaatsvindt. Een goed alternatief is het uitvoeren van een analytisch beslismodel. In zo'n model worden bewijsstukken uit verschillende gebieden met elkaar gecombineerd, zoals de statistische prestatie van het predictiemodel met de effecten van en trouwheid van patiënten aan een bepaalde behandeling. Naast de vele voordelen van het gebruik van zulke analytisch beslismodellen zijn er zeker ook nadelen. Een overzicht van de belangrijkste voor- en nadelen voor het bepalen van de klinische impact van een predictiemodel met een analytisch beslismodel (zonder het uitvoeren van een klinische studie) zijn weergegeven in **Hoofdstuk 8**. Twee voorbeelden van studies naar de klinische impact van een predictiemodel met behulp van een analytisch beslismodel zijn beschreven in detail. Daarnaast wordt er in een begeleidende tabel weergegeven hoe je een modelmatige impact studie kunt uitvoeren.

ACKNOWLEDGEMENTS

The consortium members consist of (in alphabetical order):

Y. Appelman[1], S.Baart[2,3], L. Benschop[2,3], E.Boersma[2], L. Brouwers[3,4], R.P.J. Budde[2], S. C. Cannegieter[5], V. Dam[3,6], M.J.C. Eijkemans[6], B.C.J.M. Fauser[4], M.D. Ferrari[5], A. Franx[4], C.J.M. de Groot[1], M.N. Gunning[3,4], A. Hoek[7], H. Koffijberg[6,8], M.P.H. Koster[2], M.C. Kruit[5], G.R. Lagerweij[3,6], C.B. Lambalk[1], J.S.E. Laven[2], K.M. Linstra[2,3,5], A. van der Lugt[2], A.H.E.M. Maas[9], A. Maassen van den Brink[2], C. Meun[2,3], S. Middeldorp[10], K.G.M. Moons[6], B.B. van Rijn[2,4], J.E. Roeters van Lennep[2], J.W. Roos-Hesselink[2], L.J.J. Scheres[3,10], Y.T. van der Schouw[6], E.A.P. Steegers[2], R.P.M. Steegers-Theunissen[2], G.M. Terwindt[5], B.K. Velthuis[4], M.J.H. Wermer[5], B. Zick[2,5], G.A. Zoet[3,4]

1 Vrije Universiteit 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, Utrecht, 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

DANKWOORD

Lang lang geleden riep ik “promoveren is niks voor mij, daar begin ik niet aan”. Het is maar goed dat ik toen geen glazen bol had, want ik was direct weg gerend. Met trots in mijn stem kan ik nu zeggen, ik ben blij dat ik toen geen glazen bol had want ik had het promoveren niet willen missen. Het was een leerzaam en mooi traject, wat ik niet had kunnen voltooien zonder hulp van de volgende mensen:

Geachte Prof Moons, beste Carl, het enthousiasme en motivatie die jij overdraagt aan anderen is enorm. Dank je wel voor het leerzame traject, alle gedane moeite om bepaalde overgangen naar werkgevers soepel te laten verlopen. Daarnaast mijn dank voor alle adviezen en het vertrouwen in mij en het promotietraject. Uiteindelijk ging ik er zelf ook in geloven en kreeg ik er nog meer lol in.

Geachte Dr. Koffijberg, beste Erik, ik kan me nog goed herinneren hoe ik voor de tweede keer in Utrecht, bij de HTA-afdeling, en bij jou op sollicitatiegesprek kwam voor een promotieplek. Deze keer kreeg ik de functie en ik was super blij. Die blijdschap was (even) iets minder toen ik halverwege 2014 hoorde dat je in Enschede een nieuwe baan had gekregen. Ondanks de lange afstand spraken we elkaar voldoende, volgens jou zelfs vaker dan met de PhD’s in Enschede. Dank je wel voor alle gezellige, leuke, vrolijke, technische en lange telefoongesprekken. Mijn wangen waren soms rood van de inspanning, luisteren, en het opschrijven van alle ideeën en gedachtes die er omhoog borrelden. Daarnaast zagen we elkaar ook dikwijls in Enschede. Ik werd warm opgenomen door de mensen in Twente en het was altijd erg leuk om tijdens de lunches goede verhalen over jou te horen. Gedurende de treinreizen vanuit Twente naar Amersfoort was ik vaak helemaal vol met ideeën en to-do dingen, de tijd vloog voorbij dan.

Geachte Dr. De Wit, beste Ardine, het was fijn om naast de twee mannelijke begeleiders ook een vrouwelijke begeleider te hebben. Je zorgde ervoor dat als de gesprekken te ver afdwaalden dat we weer teruggingen naar de basis. Ik heb ook enorm veel van je geleerd omtrent HTA en HTA-onderzoek doen, waardoor ik mezelf ook echt als HTA-onderzoeker ging zien. Dank je wel voor alle wijze lessen, puntjes op de i, praktische en toegepaste vragen waardoor ik weer even met beide benen op de grond stond. Daarnaast ook bedankt voor alle ‘bijklets-momentjes’, met ervaringen uit jouw leven en onderzoek. Ik had deze gesprekken soms even nodig.

Geachte leden van de beoordelingscommissie, Prof Joore, Prof Smit, Prof Verkooijen, Prof Bots, en Prof Visseren, hartelijk dank dat jullie deel wilden nemen in de beoordelingscommissie en dat jullie de tijd hebben genomen om dit proefschrift te beoordelen.

Leden van CREW, hartelijk dank voor alle inbreng, adviezen en ervaringen vanuit de kliniek. Beste mede-CREW promovendi, wat een heerlijk bij elkaar geraapt groepje waren we. Naast inhoudelijke vragen over onderzoek werd er ook verhalen uit de kliniek en 'miscommunicatie' ervaringen met een bepaald instituut met elkaar. Dank je wel voor de leuke CRapp jaren met elkaar.

Laura, onze samenwerking was een succes. Ik heb genoten van je directheid maar ook van je doortastbaarheid en doorzettingsvermogen. Ik vind ons paper een toppertje en wellicht schrijven we er nog een vervolg op, om eindelijk van het puntje "hoe zit het met medicatie" af te zijn. Veel succes en plezier met je opleiding en laten we proosten als het artikel geaccepteerd is. Kevin, onze samenwerking was geslaagd. Onze meetings waren vaak erg verhelderend en lang, maar het resultaat mag er zeker zijn. Ik hoop dat we binnenkort een drankje kunnen doen als het artikel geaccepteerd is.

Beste Dr. Kip en Dr. Weernink, beste Marieke en Michele, dank je wel voor alle goede zorgen als ik weer in Enschede was. Ik vond het erg gezellig en leuk op bij jullie op de kamer te mogen "flexen". Ik zorgde altijd dat ik bijtijds in Twente was zodat er voldoende tijd was voor bijkletsen en thee halen. Dankzij jullie waren het SMDM in Londen en de verschillende LolaHesg congressen een groot succes.

Beste HTA en methodologie collega's, alle mensen met ik een kamer heb gedeeld, dank je wel voor alle gezellige, vrolijke en top jaren met elkaar. Daarnaast wil ik Timo en Romin bedanken dat ze de 'mannenkamer' in het van Geuns wilden delen met mij. Het was fijn om vragen te kunnen stellen over allerlei technische dingen, maar ook het delen van frustraties omdat papers (nog) niet gereviewed waren. Katrien, wat heb ik genoten van je aanwezigheid in het van Geuns. Je toverde de kamer om in een plantenparadijs, zorgde voor dagelijkse hilarische momenten door samen met mij te gymastieken in de kamer, en door de week fantastisch af te sluiten met een taart-moment. Dank je wel voor het opvrolijken van de (soms stressvolle) laatste maanden van mijn promotie. Marian, thanks for all the chocolate you brought from your trips, the great talks (e.g. picking a scale as Christmas present) and looking after you by giving you warm coffee and lots of water.

Anneke, Femke, en Linda, jullie zie ik niet meer als kamergenootjes maar meer als kamer-vriendinnen. Ik heb een fantastische tijd gehad met z'n viertjes in onze oude kamer, maar ook de jaren erna in het van Geuns, of tijdens onze goede koffie meetings, diners of andere bijeenkomsten. Ondertussen is meer dan de helft van ons gepromoveerd en/of moeder geworden, allemaal super gave gebeurtenissen waar ik allemaal deelgenoot van mocht zijn. Ik hoop dat we nog lang deelgenoot mogen zijn van elkaars leven.

Familieleden en vrienden wil ik graag bedanken voor hun luisterend oor, leuke activiteiten als afleiding en gezellige koffie/thee en diner momenten.

Lief broertje Almer, super dat jij vandaag naast mij willen staan als mijn paranimf en poortwachter. Ik beloof je dat ik geen vragen aan je zal doorgeven. Mijn supertoffe zusje, lieve Mileen, dank je wel dat jij mijn paranimf (en paranimf-logistiek) wilde zijn zowel vandaag als tijdens de afgelopen maanden. Ik vond het super fijn dat je mijn gekrabbel wilde doorlezen, voorzien van commentaar, je mening wilde geven over de opmaak, en mij wilde helpen bij de organisatie van deze mooie dag (en de heren in het gareel houden). Lieve ouders, mede dankzij jullie sta ik hier nu vandaag. Wie had dat ruim 19 jaar geleden gedacht toen ik van de lagere school af ging met een mavo/havo advies? Dank jullie wel dat het ouderlijk huis altijd open staat voor een kopje thee met een knappertje. Maarn zal altijd voor mij een "thuis" zijn.

Bijna acht jaar geleden begon ik met waarschuwen; "het gaat een zware tijd worden, stage en afstuderen". Je bent gebleven, zelfs toen ik ging promoveren. Lieve lieve Almar, dank je wel voor alles!

BIOGRAPHY

Giske was born on the 21st of March 1987 in Amersfoort, the Netherlands. In 2005, she finished secondary school at the Revius Lyceum in Doorn. In the same year, she started her study Applied Mathematics at the University of Twente. During her study she was active in the Study Association W.S.G. Abacus and assisted many mathematical courses for different studies. She has always had an interest in combining her technical background with a medical subject, therefore she did her internship and graduation project on unravelling noise and potential tumor tissue on PET scans at Medical Spectrum Twente. After obtaining her master's degree in 2012, she started working at a software company where she helped developing applications for care providers to submit declarations to insurance companies. In 2014, she started working on the research described in this thesis at the Julius Center for Health Science and the Dutch Heart Foundation. She combined her research with the postgraduate master in Epidemiology at Utrecht University. In 2017, she finished this master with a specialization in medical statistics.