Journal of Clinical Epidemiology 93 (2018) 103–111

Journal of Clinical Epidemiology

Predicted burden could replace predicted risk in preventive strategies for cardiovascular disease

Ghizelda R. Lagerweij[a,b,*], G. Ardine de Wit[a,c], Karel G.M. Moons[a], W.M. Monique Verschuren[a,c], Jolanda M.A. Boer[c], Hendrik Koffijberg[a,d]

[a]*Julius Center for Health Sciences and Primary Care, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands*
[b]*Netherlands Heart Institute, Holland Heart House, Moreelsepark 1, 3511 EP, Utrecht, The Netherlands*
[c]*Centre for Nutrition, Prevention, and Healthcare, National Institute for Public Health and the Environment (RIVM), Antonie van Leeuwenhoeklaan 9, 3721 MA, Bilthoven, The Netherlands*
[d]*Department of Health Technology & Services Research, MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente, Drienerlolaan 5, 7522 NB, Enschede, The Netherlands*

Accepted 15 September 2017; Published online 21 September 2017

Abstract

Objectives: The objective of this study was to explore the extent of the differences in definitions of composite end points and assess how these differences influence estimates of cardiovascular disease (CVD) burden.

Study Design and Settings: Data from a Dutch cohort study (n = 19,484) was used to calculate 10-year risks according to four CVD risk prediction models: Adult Treatment Panel (ATP) III, Framingham Global Risk Score (FRS), Pooled Cohort Equations (PCE), and SCORE. Health loss was estimated based on the impact of event types included in the corresponding composite end points. 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 end points varied widely across the four models. FRS predicted the highest CVD risks, and the composite end point 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 end points 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.

Keywords: Prediction model; Burden of disease; Cardiovascular disease; Composite end point

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 17.3 million in 2008 to 23.6 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, that is, 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 $\leq$20% 10-year CVD risk as low risk and individuals with a $>$20% 10-year CVD risk as high-risk, whereas the Pooled Cohort Equations (PCE) uses a 7.5% 10-year CVD risk threshold instead of 20% [3,6].

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.

* Corresponding author. Julius Center for Health Sciences and Primary Care, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands. Tel.: +31-88-75-681-81; fax: +31-88-75-680-99.

E-mail address: g.r.lagerweij@umcutrecht.nl (G.R. Lagerweij).

https://doi.org/10.1016/j.jclinepi.2017.09.014
0895-4356/

What is new?

Key findings

- There is a wide variation in composite end points used in cardiovascular disease (CVD) risk prediction models, which complicates their use for assessing burden of disease.
- Even for widely used CVD risk prediction models, the definition of composite end points is not always transparent.

What this adds to what was known?

- When the CVD burden is estimated using a risk prediction model, results will highly dependent on the prediction model used and should therefore be interpreted with caution.

What is the implication and what should change now?

- It is recommended that developers of CVD risk prediction models with a composite end point clearly describe the definition and incidence of that composite end point, and its individual components, in the development data.
- For assessing CVD burden and the potential impact of preventive interventions, it is recommended that CVD prediction models are used which cover a broad range of CVD events instead of predicting only a limited number of specific types of CVD events. For example, Framingham Global Risk Score and Pooled Cohort Equations are likely more useful in this particular context than Adult Treatment Panel and SCORE.

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 (QoL) 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 end points, including multiple and different types of CVD events. For example, one CVD risk prediction model may predict only (fatal or nonfatal) stroke, whereas another may predict only (fatal and nonfatal) myocardial infarction (MI). Often, even more complex composite end points are used, in which >10 different types of CVD events are combined. The use of composite end points may be favorable 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 end points makes it hard to estimate the health loss related to that end point, unless the included, separate CVD events are considered. When, in addition, different CVD risk prediction models use different composite end points, 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. First, we investigate the definition and constitution of the composite end points used in these CVD risk prediction models. Second, we estimate the CVD risk for all individuals in the cohort, and the health loss of the CVD events included in the composite end point, for each prediction model. Finally, we assess how the identified differences in composite end points in the prediction models considered influence the estimated CVD burden in this cohort.

2. Methods

2.1. Constitution of composite end points in MORGEN

Seven widely used CVD risk prediction models were initially selected for this study: Adult Treatment Panel (ATP) III, Framingham Global Risk Score (FRS), 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, for example, gender, age, and systolic blood pressure. Furthermore, all models were derived from general population cohorts. All prediction models except SCORE are Cox proportional hazards regression models, that is, semiparametric survival models where the form of the baseline hazard is not specified. The SCORE model is a Weibull model, that is, 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 end point, occurring within 10 years. The exact definition of the composite end point 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 end points of the seven CVD risk prediction models using a large population cohort (MORGEN) in the Netherlands. The MORGEN cohort includes men and women aged 20−74 years at baseline, recruited from the general population between 1993 and 1997 [17]. After a follow-up time of 10−15 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, 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 end points, 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 end point (Appendix B) [18]. As the different prediction models have different composite end points, 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 end point of each prediction model.

2.2. *Consequences of dissimilarities in composite end points*

Assessment of dissimilarities in the consequences of the composite end points 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) [19]. 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 [19]. 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 [19]. The discrimination of the original and recalibrated models was also assessed, using Harrell’s c-statistic [20]. The discrimination measure indicates the accuracy of the model by ordering individuals by their risk, that is, a subgroup with high-risk individuals should exhibit higher event rates than a low-risk subgroup [21].

The original CVD risk prediction models were developed with other data that 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%, and >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, that is, the individualized (weighted) impact, of a ‘‘composite end point,’’ Quality-Adjusted Life Years (QALYs) were used. The QALY is a measure combining the length of life and QoL of individuals [22]. 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 suboptimal health status, that is, 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 [23]. 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 5 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 end point. 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 end point as defined per model, this burden was also assessed using the most comprehensive end point used in the four models.

3. Results

3.1. Constitution of composite end points in MORGEN

Table 1 (columns 1 and 2) shows that composite end points 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 columns 1 and 2. Per prediction model, the type of individual components and observed number of individuals experiencing this component (event) are shown in Table 1 (columns 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 CVDs (Fig. 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 end points

Calibration and discrimination results for the original ATP, FRS, PCE, and SCORE models and the recalibrated models, based on the end points 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 end points, and type and number of individual components, directly led to large differences in predicted risks, as shown in Fig. 2. Incorporation of more individual components into the composite end point automatically lead to higher predicted risks and prediction models focusing only on more severe events, that is, 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.

Fig. 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 supralinear 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 end point (in terms of QALYs lost) according to prediction model SCORE were expected to be highest due to the severity of the incorporated individual components, that is, only fatal CVD events. For the other models, the consequences of the composite end points were much lower and in the same order of magnitude. For all models, the risk and consequences of the composite end point were 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 nonfatal events increases with age, due to decreasing life expectancy (see Appendix F). SCORE showed the most rapid decrease in consequences of the composite end point (Fig. 3B). Fig. 3C illustrates the results for the predicted individualized CVD burden per individual, that is, 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 end points. Fig. 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 end points. Of the four models considered, the Framingham model used the most comprehensive end point (Table 1). Using the Framingham composite end point 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 (Fig. 4B).

Table 1. Individual components and structure of composite end points in cohort

		ATP		FRS		PCE		SCORE		QRISK		PROCAM		Reynolds	
Individual components	**ICD-10 code**		#		#		#		#		#		#		#
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		
Hemorrhagic 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		1,060		374		0		1,060		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		
Hemorrhagic 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 end points (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		1,148		436		141		1,148		362		436

Abbreviations: ATP, Adult Treatment Panel; FRS, Framingham Global Risk Score; PCE, Pooled Cohort Equations; SCORE, Systematic COronary Risk Evaluation risk score; PROCAM, Prospective Cardiovascular Munster study; QRISK, cardiovascular disease risk algorithm for UK.

4. Discussion

In this study, the definitions and constitution of composite end points 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 end point, that is, they include different sets of CVD event types (individual components). This variation in individual components induces large differences in predicted risk, that is, 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 end point 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 end points in studies may be more relevant to patients and clinicians as they cover more aspects and outcomes of the disease [24]. The usefulness of composite end points in the context of randomized trials, however, is still debated, due to the ensuing difficulty of interpreting differences in "sets of outcomes" [25–31]. Moreover, even commonly used prediction models, such as the four models considered here, often have hard to find, or unclear, definitions of the composite end point in terms of ICD codes included. This affects a direct comparison of CVD risk prediction models, as each different composite end point has to be unraveled into its individual components, and each component has to be linked to a unique disease code. This process complicates the statistical analysis, for example, evaluation, comparison, and external validation of prediction models. Still, a transparent description of the composite end point 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 [32–34]. For example, assuming that preventive statin treatment reduces the risk of a composite CVD end point by a certain percentage will result in estimated health benefits which are highly

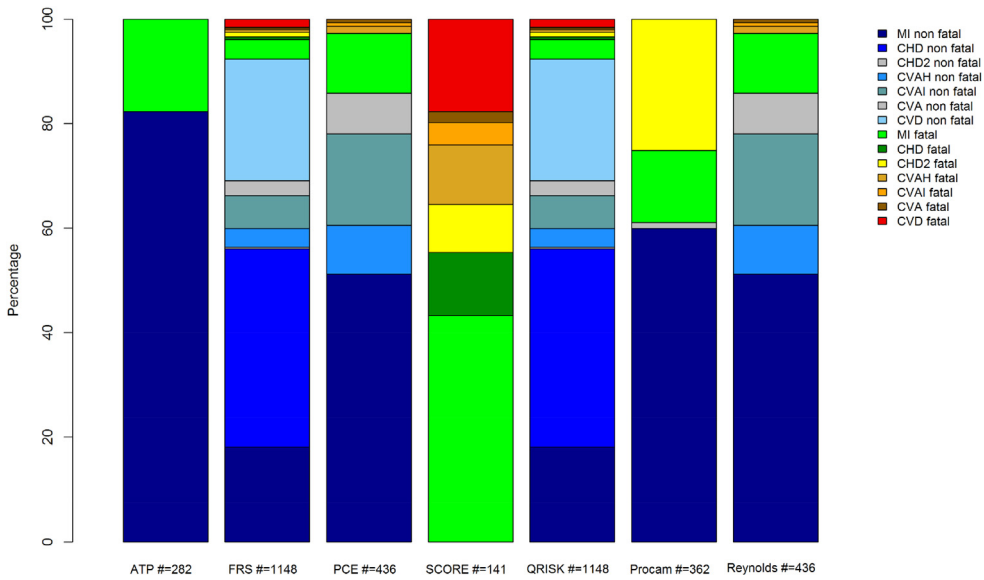

Fig. 1. Overall distribution of included individual components per CVD risk prediction model. MI, myocardial infarction; CHD, coronary heart disease; CVAH, hemorrhagic stroke; CVAI, ischemic stroke; CVA, cerebrovascular accident; CVD, cardiovascular disease; ATP, Adult Treatment Panel; FRS, Framingham Global Risk Score; PCE, Pooled Cohort Equations; SCORE, Systematic COronary Risk Evaluation risk score; PROCAM, Prospective Cardiovascular Munster study; QRISK, cardiovascular disease risk algorithm for UK.

dependent on the prediction model used [35]. Appropriate impact analysis of risk-based preventive interventions requires evidence of (1) the initial risk of different types of CVD events, (2) their consequences, and (3) 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.

4.1. Strengths

Four widely used CVD risk prediction models are compared regarding their composite end points, their risk estimates, and the associated burden of disease. Furthermore, this study unambiguously links all CVD end points of interest to ICD-10 codes, thereby improving clarity and ensuring replicability of the analyses in other cohorts. In addition, the size of the data set 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 end points.

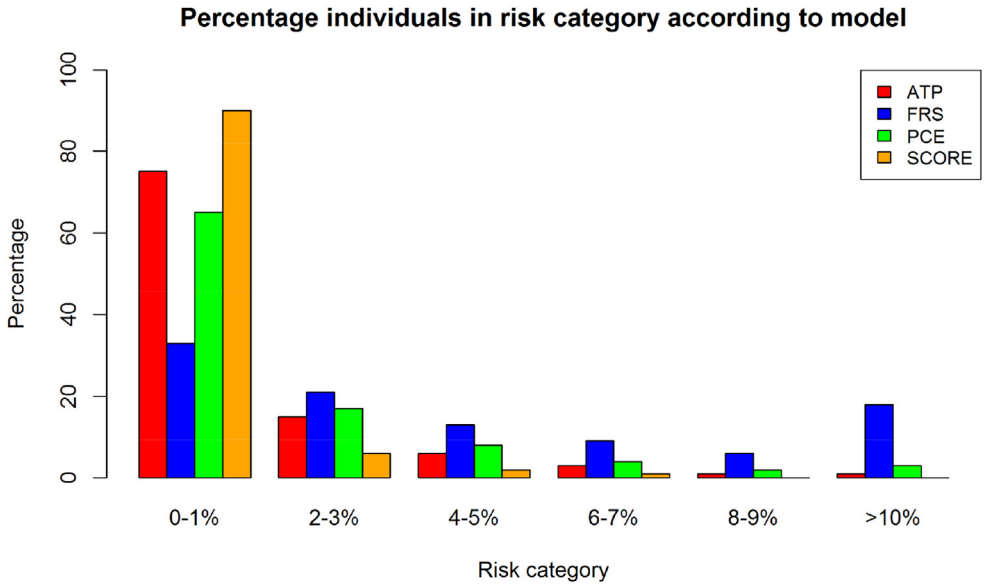

Fig. 2. Distribution of individuals per risk category and CVD risk prediction model. ATP, Adult Treatment Panel; FRS, Framingham Global Risk Score; PCE, Pooled Cohort Equations; SCORE, Systematic COronary Risk Evaluation risk score; CVD, cardiovascular disease.

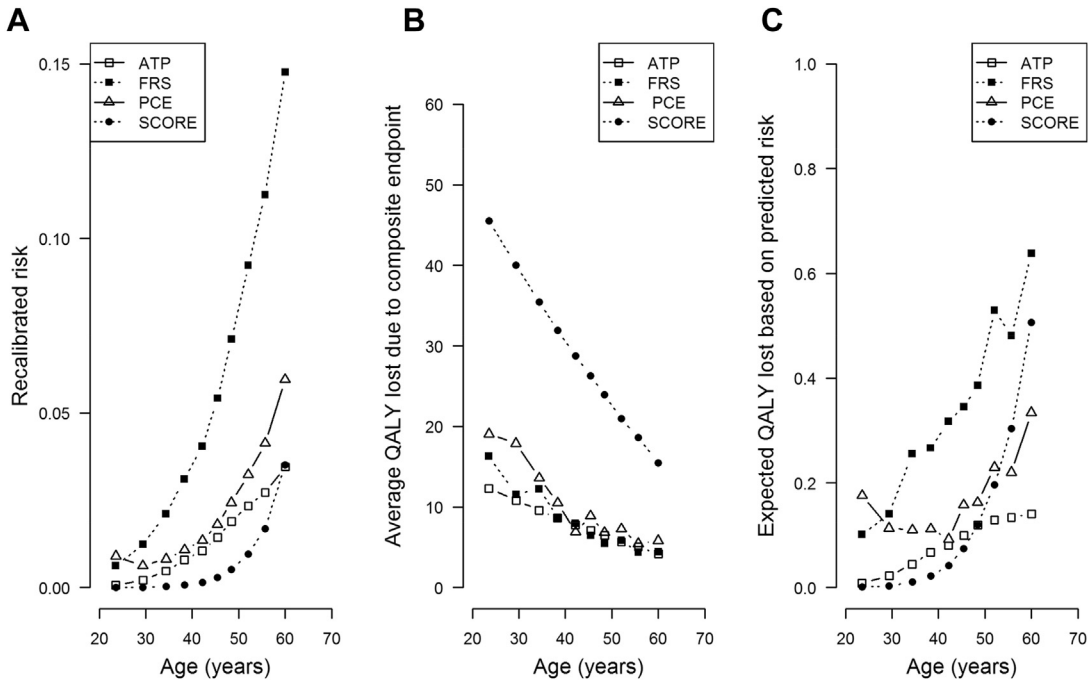

Fig. 3. Three figures as function of age, with plot (A) the 10-year CVD predicted risks, plot (B) the expected (lifetime) consequence of a composite end points per individual, and plot (C) the expected (potentially preventable) CVD burden per individual. Distribution of individual components was evaluated per age category, except for ATP where this distribution was assessed in the entire population due to limited number of included end points. ATP, Adult Treatment Panel; FRS, Framingham Global Risk Score; PCE, Pooled Cohort Equations; SCORE, Systematic COronary Risk Evaluation risk score; CVD, cardiovascular disease; QALY, Quality-Adjusted Life Year.

4.2. Limitations

The actual results from this study are dependent on the data set used, that is, the observed differences between CVD risk prediction models may be different in other data sets 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

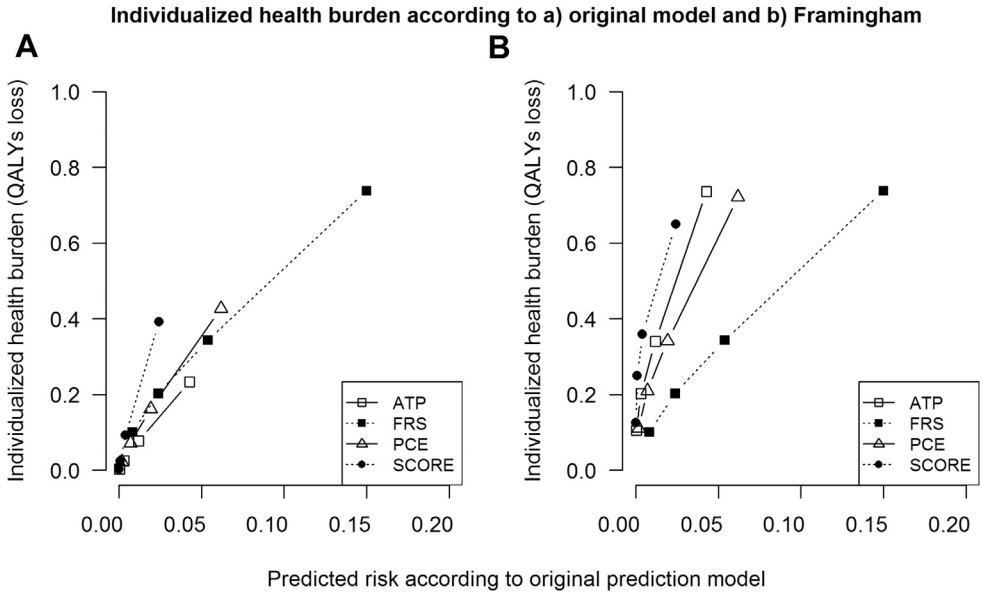

Fig. 4. Individualized CVD burden (the CVD burden was estimated for categories based on risk quartiles) (QALYs loss) according to plot (A) the original end point and plot (B) comprehensive end point (Framingham model). ATP, Adult Treatment Panel; FRS, Framingham Global Risk Score; PCE, Pooled Cohort Equations; SCORE, Systematic COronary Risk Evaluation risk score; CVD, cardiovascular disease; QALYs, Quality-Adjusted Life Years.

which composite end points are common such as for example the C-WATCH risk score for upper gastrointestinal bleeding [36]. Further analyses in other disease areas require large individual patient data sets 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 end points into ICD-10 codes, certain assumptions are required due to unclear definitions of the composite end points in the original publications. For PCE and ATP, the defined end points ''nonfatal MI and CHD death'' are translated in ''nonfatal 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 end points ''hard CHD'' as developing an 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 need 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 [37].

4.3. Recommendations

First, it is recommended that developers of CVD risk prediction models with a composite end point clearly describe the definition of that composite end point, as well as all its individual components, and their incidence in the development cohort. Second, studies comparing (the performance of) different prediction models should clearly describe the data set(s) used and the link defined between the composite end points 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 end point.

5. Conclusions

Our results suggest that the number of different composite end points 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 end point 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 end points are used to reflect risks and side effects from such treatments. Whereas a broad common set of end points 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 FRS and QRISK, cover more manifestations of CVD and might therefore yield more meaningful estimates regarding the (preventable) burden of CVD.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2017.09.014.

References

[1] Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB, et al. Executive summary: heart disease and stroke statistics−2013 update: a report from the American Heart Association. Circulation 2013;127:143−52.
[2] Cardiovascular disease 2015. Available at http://www.who.int/cardiovascular_diseases/en/, Accessed October 19, 2017.
[3] Goff DC Jr, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, 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:S49−73.
[4] Hippisley-Cox J, Coupland C, Robson J, Brindle P. Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. BMJ 2010;341:c6624.
[5] Pencina MJ, D'Agostino RB Sr, Larson MG, Massaro JM, Vasan RS. Predicting the 30-year risk of cardiovascular disease: the framingham heart study. Circulation 2009;119:3078−84.
[6] Grundy SM, Cleeman JI, Merz CN, Brewer HB Jr, Clark LT, Hunninghake DB, et al. Implications of recent clinical trials for the National Cholesterol Education Program Adult treatment Panel III guidelines. J Am Coll Cardiol 2004;44:720−32.
[7] Hennekens CH, Buring JE. Epidemiology in medicine. 1st ed. Lippincott Williams & Wilkins; 1987.
[8] Bailey L, Vardulaki K, Langham J, Chandramohan D. Introduction to epidemiology. 1st ed. Open University Press; 2005.
[9] Lim E, Brown A, Helmy A, Mussa S, Altman DG. Composite outcomes in cardiovascular research: a survey of randomized trials. Ann Intern Med 2008;149:612−7.
[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;285:2486−97.

[11] D'Agostino RB Sr, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 2008;117:743−53.
[12] Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. JAMA 2007;297:611−9.
[13] Conroy RM, Pyorala K, Fitzgerald AP, Sans S, Menotti A, De Backer G, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J 2003;24:987−1003.
[14] Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ 2007;335:136.
[15] Assmann G, Cullen P, Schulte H. 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;105:310−5.
[16] Ridker PM, Paynter NP, Rifai N, Gaziano JM, Cook NR. C-reactive protein and parental history improve global cardiovascular risk prediction: the Reynolds Risk Score for men. Circulation 2008;118:2243−51. 4p following 2251.
[17] Beulens JW, Monninkhof EM, Verschuren WM, van der Schouw YT, Smit J, Ocke MC, et al. Cohort profile: the EPIC-NL study. Int J Epidemiol 2010;39:1170−8.
[18] Struijk EA, May AM, Wezenbeek NL, Fransen HP, Soedamah-Muthu SS, Geelen A, et al. Adherence to dietary guidelines and cardiovascular disease risk in the EPIC-NL cohort. Int J Cardiol 2014;176:354−9.
[19] Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21:128−38.
[20] FE H. Regression modeling strategies. New York: Springer; 2001.
[21] Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol 2013;13:33.
[22] Philipp CS. Health economics. 2nd ed. 2009. Available at http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/QALY.pdf. Accessed October 19, 2017.
[23] LondonIn lipid modification: Cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease. London: NICE; 2014.
[24] Ferreira-Gonzalez I, Alonso-Coello P, Sola I, Pacheco-Huergo V, Domingo-Salvany A, Alonso J, et al. [Composite endpoints in clinical trials]. Rev Esp Cardiol 2008;61(3):283−90.
[25] Montori VM, Permanyer-Miralda G, Ferreira-Gonzalez I, Busse JW, Pacheco-Huergo V, Bryant D, et al. Validity of composite end points in clinical trials. BMJ 2005;330:594−6.
[26] Ross S. Composite outcomes in randomized clinical trials: arguments for and against. Am J Obstet Gynecol 2007;196:119.e1−6.
[27] Chi GY. Some issues with composite endpoints in clinical trials. Fundam Clin Pharmacol 2005;19(6):609−19.
[28] Kip KE, Hollabaugh K, Marroquin OC, Williams DO. 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;51:701−7.
[29] Carneiro AV. Composite outcomes in clinical trials: uses and problems. Rev Port Cardiol 2003;22(10):1253−63.
[30] Huque MF, Alosh M, Bhore R. Addressing multiplicity issues of a composite endpoint and its components in clinical trials. J Biopharm Stat 2011;21(4):610−34.
[31] Cordoba G, Schwartz L, Woloshin S, Bae H, Gotzsche PC. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. BMJ 2010;341:c3920.
[32] Gondrie MJ, Janssen KJ, Moons KG, van der Graaf Y. A simple adaptation method improved the interpretability of prediction models for composite end points. J Clin Epidemiol 2012;65:946−53.
[33] Glynn RJ, Rosner B. Methods to evaluate risks for composite end points and their individual components. J Clin Epidemiol 2004;57:113−22.
[34] Ferreira-Gonzalez I, Permanyer-Miralda G, Busse JW, Bryant DM, Montori VM, Alonso-Coello P, et al. Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. J Clin Epidemiol 2007;60:651−7. discussion 658−662.
[35] Heeg BM, van Hout BA. Assessing uncertainties surrounding combined endpoints for use in economic models. Med Decis Making 2014;34:300−10.
[36] Hoffmann V, Neubauer H, Heinzler J, Smarczyk A, Hellmich M, Bowe A, et al. A novel easy-to-use prediction scheme for upper gastrointestinal bleeding: Cologne-WATCH (C-WATCH) risk score. Medicine (Baltimore) 2015;94(38):e1614.
[37] Wattanakit K, Folsom AR, Chambless LE, Nieto FJ. Risk factors for cardiovascular event recurrence in the Atherosclerosis Risk in Communities (ARIC) study. Am Heart J 2005;149:606−12.