

# Prognostic Models With Competing Risks

## *Methods and Application to Coronary Risk Prediction*

Marcel Wolbers,<sup>a,b</sup> Michael T. Koller,<sup>a,b,c,d</sup> Jacqueline C. M. Witteman,<sup>d</sup> and Ewout W. Steyerberg<sup>c</sup>

**Abstract:** Clinical decision-making often relies on a subject's absolute risk of a disease event of interest. However, in a frail population, competing risk events may preclude the occurrence of the event of interest. We review competing-risk regression models with a view toward predictive modeling. We show how measures of prognostic performance (such as calibration and discrimination) can be adapted to the competing-risks setting. An example of coronary heart disease (CHD) prediction in women aged 55–90 years in the Rotterdam study is used to illustrate the proposed methods, and to compare the Fine and Gray regression model to 2 alternative approaches: (1) a standard Cox survival model, which ignores the competing risk of non-CHD death, and (2) a cause-specific hazards model, which combines proportional hazards models for the event of interest and the competing event. The Fine and Gray model and the cause-specific hazards model perform similarly. However, the standard Cox model substantially overestimates 10-year risk of CHD; it classifies 18% of the individuals as high risk (>20%), compared with only 8% according to the Fine and Gray model. We conclude that competing risks have to be considered explicitly in frail populations such as the elderly.

(*Epidemiology* 2009;20: 555–561)

Clinical decision-making and cost-effectiveness analyses often rely on a subject's absolute risk of a disease event of interest.<sup>1</sup> As an example, the National Cholesterol Education Program Adult Treatment Panel III<sup>2</sup> treatment recommendations for the prevention of coronary heart disease (CHD) are based on the predicted 10-year risk of CHD.<sup>2–4</sup> If

the 10-year risk exceeds 20%, patients are classified as high-risk and deserve aggressive treatment.<sup>2</sup> Hence, the implementation of preventive analytic strategies requires prognostic models that estimate the actual individual risk as accurately as possible.

In frail populations, such as elderly subjects, other causes of failure may occur prior to the occurrence of the disease event of interest. Because such competing risk events preclude the event of interest and thus the benefit of an intervention, prognostic models should take competing risk events into account.

We consider competing risk models and show how measures for the evaluation of calibration and discrimination of prognostic survival models can be adapted to the competing risks setting. An example of CHD risk prediction for women aged 55–90 years in the Rotterdam study illustrates how predictive models that properly account for competing risks can be developed and assessed. Prediction of CHD risk in the elderly becomes increasingly important because of the ageing of populations. There are a number of established risk scores,<sup>5–8</sup> but they were not specifically developed for the ageing population and do not account for the occurrence of competing non-CHD death. We compare 3 modeling approaches and show that substantial bias arises if competing risks are disregarded.

## REGRESSION MODELS FOR COMPETING RISKS DATA

We provide a short overview of the most popular regression models for competing risks with a view toward absolute risk prediction. A tutorial for competing risks analyses has recently been published.<sup>9</sup>

The observable data in competing risks models are represented by the (possibly censored) time to the event  $T$  and the cause of failure  $D$ , ie, either the event of interest ( $D = 1$ , eg, “CHD”) or the competing event ( $D = 2$ , eg, “non-CHD death”). A key quantity is the cumulative incidence function  $I_k(t)$  for the event of interest or the competing event, which describes the actual (absolute) risk of failing from cause  $k$  until time  $t$ :  $I_k(t) = P(T \leq t \text{ and } D = k)$ . The cumulative incidence functions for the 2 competing events are both increasing with time  $t$  and add up to 1 at time infinity, as ultimately either of the 2 events is bound to occur. A non-

Submitted 26 September 2008; accepted 20 November 2008; posted 13 April 2009.

<sup>a</sup>These authors contributed equally to the work.

From the <sup>b</sup>Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel, Basel, Switzerland; and Departments of <sup>c</sup>Public Health, and <sup>d</sup>Epidemiology and Biostatistics, Erasmus MC, Rotterdam, The Netherlands.

Supported by Santésuisse, and Gottfried and Julia Bangerter-Rhyner-Foundation (to M.W., M.T.K.).

**e** Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article ([www.epidem.com](http://www.epidem.com)).

Correspondence: Marcel Wolbers, Oxford University Clinical Research Unit, Hospital for Tropical Diseases, 190 Ben Ham Tu, District 5, Ho Chi Minh City, Vietnam. E-mail: [mwolbers@oucr.org](mailto:mwolbers@oucr.org).

Copyright © 2009 by Lippincott Williams & Wilkins

ISSN: 1044-3983/09/2004-0555

DOI: 10.1097/EDE.0b013e3181a39056

parametric estimator of the cumulative incidence function,<sup>9–11</sup> similar to Kaplan-Meier estimates in survival analysis, provides a useful summary of competing risks data.

## Standard Cox Regression Models

In many applications, competing risks have been ignored (ie, patients experiencing competing events were censored at the time of these events) and standard Cox regression was applied.<sup>5–8</sup> Predicted risks for the event of interest were then derived by combining hazard ratio estimates with the estimated baseline hazard function.<sup>10</sup> This approach is adequate when competing risks are rare. However, in the presence of strong competing risks, as with frail or elderly populations,<sup>12</sup> standard survival predictions may substantially overestimate the absolute risk of the event of interest because subjects with a competing (and thus censored) event are treated as if they could experience the event of interest in the future.<sup>9,13,14</sup>

Predictions from a standard survival analysis in the presence of competing risks have been said to refer to the risk of failing from the event of interest in a virtual world where the competing risk is absent, ie, to the marginal failure time distribution of the event of interest.<sup>9</sup> This is true only if censoring due to competing events is independent of the occurrence of the event of interest,<sup>9</sup> an assumption that is often clinically implausible and cannot be empirically tested.<sup>10,15</sup> Moreover, for clinical decision-making in the real world, where competing risks do occur, actual rather than virtual absolute risks are often more relevant.<sup>13,14</sup>

## Cause-specific Hazards Models

The cause-specific hazard function for failure cause  $k$  is the instantaneous failure rate of failing at time  $t$  of cause  $k$ .<sup>9,10</sup> We denote the cause-specific hazard functions for the event of interest and the competing event by  $\lambda_1(t)$  and  $\lambda_2(t)$ , respectively. The cumulative incidence function of the event of interest can be shown to depend on both the cause-specific hazard of the event of interest and the competing event according to the formula  $I_1(t) = \int_0^t \lambda_1(s) \exp(-\int_0^s (\lambda_1(u) + \lambda_2(u)) du) ds$ .<sup>10</sup> The cause-specific hazard modeling approach to absolute risk prediction therefore corresponds to first developing proportional cause-specific hazards models for both the event of interest and the competing event, and then combining them according to this formula.

Cause-specific hazards models can be estimated by censoring patients with the respective competing event and then fitting standard Cox regression models.<sup>9</sup> However, the 2 approaches differ in the way absolute risk predictions are calculated; while standard survival predictions depend only on the cause-specific hazard of the event of interest (and thus overestimate absolute risks in the presence of competing events), proper predictions from cause-specific hazards models are based on the formula for  $I_1(t)$  above.

Cause-specific hazard models yield correct absolute risk estimates but are not without problems. First, they require modeling not only the event of interest but also the competing event to obtain valid risk predictions. Second, covariate effects on the cause-specific hazard of the event of interest cannot be directly interpreted in terms of the cumulative incidence function, which depends on both cause-specific hazards.<sup>9,16</sup> Third, the formula for  $I_1(t)$  combining cause-specific hazards is essentially a black box, ie, it is not possible to write down a simple formula for risk predictions given covariate profiles. This complicates the communication of prediction rules based on cause-specific hazards models.

## Fine and Gray Model

Several direct regression models for the cumulative incidence have been proposed.<sup>17–19</sup> We focus on the Fine and Gray<sup>17</sup> model, which is most widely used and allows for a proportional hazards interpretation. It is a proportional hazards model for the subdistribution hazard of the event of interest, defined as  $\bar{\lambda}_1(t) := -\frac{d \log(1 - I_1(t))}{dt}$ . Given covariates  $Z$ , the model is of the form  $\bar{\lambda}_1(t|Z) = \bar{\lambda}_{1,0}(t) \exp(\beta'Z)$  where  $\bar{\lambda}_{1,0}(t)$  is the baseline subdistribution hazard for the event of interest. Solving for the cumulative incidence function gives the formula:

$$I_1(t|Z) = 1 - \exp\left(-\exp(\beta'Z) \cdot \int_0^t \bar{\lambda}_{1,0}(s) ds\right)$$

where  $\int_0^t \bar{\lambda}_{1,0}(s) ds$  is the cumulative subdistribution baseline hazard. From this last formula it is straightforward to calculate predicted risks at a specific time point based on the cumulative subdistribution baseline hazard and the estimates of the regression coefficients from the Fine and Gray model. The formula also illustrates the fact that covariate effects can be interpreted directly in terms of the cumulative incidence function. If the regression coefficient for a covariate is positive (ie, a subdistribution hazards ratio that is greater than 1.0), higher values of a covariate imply a constant relative increase of the subdistribution hazard, and hence a higher predicted cumulative incidence at every time point. The Fine and Gray model has been called an interpretation-friendly alternative to cause-specific hazard models.<sup>16</sup> We consider it well suited for predictive modeling in the competing-risk setting. Predictive models based on the Fine and Gray model can be developed with essentially the same modeling strategies as for other regression models.<sup>20,21</sup>

## QUANTIFYING PREDICTIVE ACCURACY OF COMPETING RISKS MODELS

### Calibration

Calibration refers to whether the predicted risks from a prognostic model agree with the observed risks. This is

particularly important for external validation of a prognostic model.<sup>22,23</sup> To assess calibration for competing risks models, the analyst may choose one or several time points and then plot the actual observed risk, ie, the cumulative incidence function estimate,<sup>9–11</sup> computed within percentiles of predicted risk, against the average predicted risk within the same percentiles for the event of interest. In the case of CHD prediction, one would use the 10-year risk estimates since these are the basis of current guidelines.<sup>2</sup>

It was shown in the previous section that the standard Cox approach overestimates the actual risk (cumulative incidence). If severe, this overestimation is evident when calibrated against the cumulative incidence function estimates. Ironically, if the observed risk is calculated with the standard Kaplan-Meier estimator, calibration may falsely appear to be satisfactory because, just as the standard Cox approach, the Kaplan-Meier estimator overestimates absolute risk in the presence of competing risks.<sup>9,11,13</sup>

## Discrimination and Reclassification

Intuitively, a prediction model for the event of interest discriminates well if it assigns high risks to individuals experiencing the event of interest early, lower risks to individuals experiencing the event of interest later, and negligible risk to those never experiencing the event of interest (ie, individuals with competing events) and to those without any event during follow-up (censored observations). This intuition is quantified by our adapted definition of the *c* index below.

A *c* (for concordance) index is a widely used measure for assessing predictive discrimination for continuous, binary, and survival-type outcomes.<sup>22,24</sup> It can be defined as the proportion of all evaluable ordered patient pairs for which predictions and outcomes are concordant. In the context of competing risks, we propose to define evaluable and concordant patient pairs as follows.

An ordered patient pair is defined as evaluable if the first patient experiences the event of interest at a time point when the second patient is still at risk; all other ordered pairs are nonevaluable. The risk set here is defined as follows. Both patients who experience the event of interest and patients who are censored are at risk until the event or the censoring time, respectively. Individuals who fail from the competing risk event remain in the risk set and are at risk at any time. The rationale for this unorthodox definition is that patients experiencing competing events are definitely known to never experience the event of interest, whereas censored patients are only known not to experience the event of interest until censoring. The same definition is inherently used in the Fine and Gray<sup>17</sup> approach.

An evaluable ordered patient pair is defined as concordant if the first patient (ie, the patient experiencing the event of interest at the time the second patient is still at risk) has the higher risk prediction than the second. When predicted risks are identical, 0.5 rather than 1 is added to the count of

concordant pairs.<sup>22</sup> If the predictive model is a Fine and Gray model containing only baseline covariates, the risk ordering at any time point is the same and given by the linear predictor of the model. However, for risk predictions based on cause-specific hazards models or Fine and Gray models that include covariate-time interactions, the risks have to be compared at a specific time point, ie, at the time the first patient experiences the event of interest.<sup>25</sup>

For the Fine and Gray model containing only baseline covariates, this definition reduces to the standard definition of the *c* index for survival data, except that patients experiencing the competing event are treated as being censored at infinity to indicate that they will never experience the event of interest. Without censored data, the *c* index indicates perfect concordance (*c* index = 1) if all patients experiencing the competing event have lower risk predictions than patients experiencing the event of interest, and event times of patients experiencing the event of interest are perfectly ordered according to the predicted risks. If the prediction is essentially random without any discriminative ability, the *c* index is around 0.5.

An alternative measure of prognostic separation in survival data is measure D, proposed by Royston and Sauerbrei,<sup>26</sup> which can be adapted to the competing risks setting by replacing Cox regression in their derivation by Fine and Gray regression. Reclassification methods for comparing new models to established ones are increasingly used<sup>27</sup> and can also be applied in the presence of competing risks.

## ILLUSTRATION: PREDICTION OF CORONARY HEART DISEASE IN WOMEN AGED 55–90 YEARS OF THE ROTTERDAM STUDY

The Rotterdam study is a prospective, population-based study among subjects 55 years and older living in a suburb area of Rotterdam, the Netherlands.<sup>28,29</sup> For this analysis, we selected from the Rotterdam cohort all women aged 55–90 years who were free of CHD or cerebrovascular disease at baseline, and we included follow-up information until January 2006.

The end point of interest was the time from inclusion in the cohort until first CHD (event of interest) or death other than CHD (non-CHD death, competing event). We used a definition of “hard” CHD consisting of nonfatal myocardial infarction, coronary interventions, and objectively determined fatal CHD, including sudden cardiac death, death from chronic ischemic heart disease, and death due to heart failure other than hypertensive or nonrheumatic valve disorders.<sup>29</sup>

Three different regression models were fitted and compared: a Fine and Gray model, a cause-specific hazards model, and a standard Cox model. Because model selection was not the aim of this work, we predefined a “traditional” model similar to existing popular Framingham models<sup>5</sup> using the following covariates: age, treatment for high blood pressure (yes versus no), systolic blood pressure (separate slopes

depending on whether the patient was on blood pressure treatment or not), diabetes mellitus, log-transformed total cholesterol to HDL cholesterol ratio, and smoking status (current versus never or former smoker) at baseline.

## Regression Models

A total of 4144 women were included with a median follow-up time of 12.8 years (quartiles, 12.0–13.5 years; completeness of follow-up,<sup>30</sup> 98%). Median age at baseline was 69 years (quartiles, 62–77 years); the total number of first hard CHD events was 465 (243 of them fatal); and 1263 women experienced competing non-CHD death.

Results from the Fine and Gray regression model for the outcome CHD are displayed in Table 1. The baseline cumulative subdistribution hazard at 10 years, provided in the footnote of the table, can be used to make individual risk predictions. All traditional risk factors (with the exception of blood pressure in patients on treatment) were strongly associated with CHD.

Results of the cause-specific proportional hazards models for the event of interest and the competing event are displayed in Table 2. Hazard ratios for the event of interest were very similar to those of the Fine and Gray model for covariates that did not affect non-CHD death, ie, blood pressure lowering medication and systolic blood pressure.<sup>16</sup> In contrast, age, diabetes, and smoking status were also strong predictors for non-CHD death, and cause-specific hazard ratios for CHD were larger than in the Fine and Gray model. The cholesterol-to-HDL-cholesterol ratio was inversely related to non-CHD death.

## Calibration, Reclassification, and Discrimination

Calibration plots are displayed in the Figure. Calibration of the Fine and Gray model and the cause-specific

hazards model were good, and the results of the 2 models were similar to each other. The risk overestimation of the standard Cox analysis is apparent.

A comparison of the standard Cox model and the Fine and Gray model with a reclassification table also illustrates the miscalibration of the standard Cox model (eTable 1, <http://links.lww.com/A939>): 12% of the low-risk patients and 37% of the intermediate-risk patients, as classified according to the Fine and Gray model, were incorrectly reclassified as intermediate risk and high risk, respectively, using the standard Cox model. In contrast, the Fine and Gray model and the

**TABLE 2.** Cause-specific Cox Models For CHD (Event of Interest) and Competing Non-CHD Death (Competing Risk)

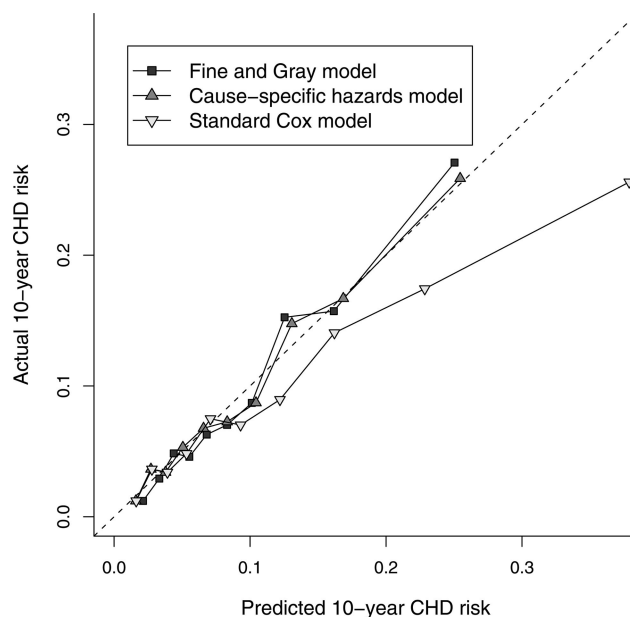
Parameter	Event of Interest: CHD HR (95% CI)	Competing Event: Non-CHD Death HR (95% CI)
Age (by +10 years)	2.23 (1.99–2.51)	3.63 (3.37–3.92)
On blood pressure lowering medication (offset at a systolic blood pressure of 120 mm Hg)	1.73 (1.28–2.33)	1.17 (0.98–1.41)
Systolic blood pressure (by +10 mm Hg)—for patients without medication	1.14 (1.07–1.20)	1.03 (1.00–1.07)
Systolic blood pressure (by +10 mm Hg)—for patients with medication	1.05 (0.99–1.12)	1.03 (0.99–1.07)
Diabetes mellitus	1.52 (1.19–1.94)	1.29 (1.10–1.52)
log(total cholesterol/HDL)	2.51 (1.85–3.42)	0.64 (0.53–0.78)
Current smoker	1.96 (1.56–2.48)	1.91 (1.65–2.22)

**TABLE 1.** Fine and Gray<sup>17</sup> Regression Model for Predicting CHD Based on 4144 Women in the Rotterdam Study Aged 55–90 Years and Free of CHD at Baseline

Parameter	sHR (95% CI)
Age (by +10 years)	1.64 (1.47–1.82)
On blood pressure lowering medication (offset at a systolic blood pressure of 120 mm Hg)	1.69 (1.26–2.25)
Systolic blood pressure (by +10 mm Hg)—for patients without medication	1.13 (1.07–1.20)
Systolic blood pressure (by +10 mm Hg)—for patients with medication	1.05 (0.99–1.12)
Diabetes mellitus	1.38 (1.07–1.79)
log(total cholesterol/HDL)	2.72 (2.01–3.70)
Current smoker	1.71 (1.36–2.15)

Baseline cumulative subdistribution hazard at 10 years is 0.0302. Baseline refers to a woman of age 60 with a systolic blood pressure of 120, a log(Cholesterol/HDL) of 1.61 (= log(5)) who is nonsmoking, without blood-pressure lowering medication, and without diabetes.

sHR indicates subdistribution hazard ratio; CI, confidence interval.



**FIGURE.** Calibration plot of the 3 regression models for the prediction of CHD in 4144 women in the Rotterdam Study aged 55–90 years and free of CHD at baseline.



**TABLE 3.** Discriminatory Power of the 3 CHD Prediction Models: Adapted c Index and Royston and Sauerbrei's D for Competing Risk Data With 95% CI

	Fine and Gray Model Index (95% CI)	Standard Cox Model Index (95% CI)	Cause-specific Hazards Model Index (95% CI)
c index <sup>a</sup>	0.70 (0.68–0.73)	0.70 (0.68–0.73)	0.70 (0.68–0.73)
Royston and Sauerbrei's D <sup>a</sup>	1.17 (1.02–1.32)	1.15 (1.00–1.30)	1.17 <sup>b</sup> (1.02–1.32)

<sup>a</sup>Adapted versions for competing risk data.<sup>b</sup>Approximate version based on predicted 10-year CHD risk. Note that, unlike for Fine and Gray and standard Cox models, the risk ordering of patients may differ at different follow-up time points for cause-specific hazards models.

CIs for the c index are basic bootstrap confidence intervals.

cause-specific model are more similar (with some advantage for the Fine and Gray model) and classify 94% of women into the same risk strata (eTable 2, <http://links.lww.com/A939>). Discrimination, as measured by the adapted c index and Royston and Sauerbrei's D, is virtually identical for all 3 models (Table 3).

## DISCUSSION

In this comparison of competing risks regression models, we find advantages of the Fine and Gray<sup>17</sup> model, which provides direct estimates of absolute risks for developing predictive models. We show how established measures for assessing calibration and discrimination, such as the c index can be adapted to the context of competing risks. The proposed methods have been applied to the problem of risk prediction of CHD in older women based on data from the Rotterdam study.

### Importance of Absolute Risks

In line with others,<sup>13,14,31</sup> we regard the absolute risk of the event of interest as crucial for medical decision-making in the competing risks setting. For example, in a subpopulation with a 10-year absolute CHD risk of 20%, 1 in 5 individuals may profit within 10 years from aggressive treatment that lowers the patients' risk of CHD but does not affect the competing risk.

In some instances, considerations of the absolute risk of competing events may also be important. As an example, consider 2 women with the same 10-year CHD risk of 20% and a risk of competing non-CHD death of 20% or 60%, respectively. Although both patients have the same chance of benefiting from CHD-specific aggressive treatment, a clinician may be somewhat less inclined to initiate a costly treatment in the second patient with a much higher morbidity and mortality for the competing event (eg, advanced stage chronic obstruction pulmonary disease or known cancer). She would likely gain fewer years of life even if CHD could be prevented.

### Comparison of Different Competing Risks Regression Models

Our example illustrates the overestimation of actual CHD risk if the competing risk is ignored, ie, if Cox regres-

sion analysis is applied in a naive way. It is well-known that Kaplan-Meier curves overestimate absolute risk in the presence of competing risks (eg,<sup>9,13</sup>) but it appears to be less well-known that the same holds for Cox regression.

In contrast, both the cause-specific hazards model and the Fine and Gray model performed similarly well in our example. Other comparisons between Fine and Gray and cause-specific models have investigated differences in parameter estimates if both are applied to the same data and their interpretation.<sup>16,32</sup>

### Measures of Accuracy

We adapted calibration plots, the c index, Royston and Sauerbrei's measure D, and reclassification methods to the competing risks context. The c index was discussed because it is a simple and widely used measure of discrimination that is unaffected by systematic calibration problems. Our adapted c index accounts for the fact that competing events prevent the occurrence of the event of interest, and in situations with strong competing risks this can have a substantial impact: In our example, a naive calculation of the traditional c index would falsely suggest better predictability of the event of interest (naive c = 0.75 versus adapted c = 0.70 for the standard Cox model). However, controversial issues have been raised that relate to the c index in general and merit further discussion.

First, by conditioning on the events being observed, the c index depends on the censoring distribution.<sup>33</sup> This is a minor issue in our case study, in which all patients have a similar follow-up duration and loss to follow-up is minimal. However, it could be more problematic in other situations, especially when the censoring distribution depends heavily on covariates. Second, the c index is relatively insensitive to small but potentially clinically-relevant changes in predictive accuracy.<sup>34,35</sup> This could be the reason why the c index, but not calibration, is virtually identical for all 3 models in our case study. An alternative that directly compares the clinical impact of different models are reclassification methods.<sup>27,34,35</sup> Third, for actual clinical decision making, approaches that take application-specific loss functions into account may be more relevant than the c index. Gail and Pfeiffer<sup>31</sup> have written a useful review of decision-theoretic

approaches for models of absolute risks. There is no universally accepted and perfect summary measure that covers all aspects of model performance. Therefore, several measures should be calculated in applications. We encourage the adaptation of additional measures for survival data<sup>26,36,37</sup> to competing risks.

## Clinical Implications for CHD Risk Prediction

As the population ages, CHD risk prediction in older subjects is a challenge, and our risk model for older women has clinical implications. First, established risk models<sup>5–8</sup> may extrapolate poorly to older women and substantially overestimate actual risks. Second, CHD risk prediction becomes more difficult with the growing incidence of competing events.

In our example, overestimation of the standard Cox model was particularly severe in high-risk strata. As age is a strong predictor of CHD risk, these high-risk strata contain a large fraction of relatively old patients. At the same time, the mortality due to reasons other than CHD increases even more dramatically with increasing age, thus explaining this finding. Studying calibration according to age strata revealed that the standard Cox model calibrates reasonably well up to an age of about 75 years, ie, the age range for which most well-known CHD risk prediction models, which ignored competing risks, were developed.<sup>5–8</sup>

The discriminative power of our model (adapted c index = 0.70) was substantially lower compared with CHD risk models in younger women among whom it has been reported<sup>5</sup> as 0.77, based on the same covariate information and a similar follow-up duration. The c index in younger women<sup>5</sup> was not adapted for competing risks, but this should have minimal impact because this population is less frail and the incidence of competing events should be much lower. Assuming comparability of the 2 cohorts otherwise, the difference suggests that CHD risk prediction is more difficult in our elderly population. This conclusion is further supported by the observation that several of the risk factors for CHD (such as age and smoking) are equally strong or even stronger risk factors for competing non-CHD death. Even though smoking is a causal risk factor of CHD, it may be a less useful prognostic predictor for identifying future CHD cases in older women because CHD is also more often precluded by competing non-CHD death in smokers (eg, due to lung cancer or chronic obstructive pulmonary disease) compared with nonsmokers. For better case identification in the presence of competing risks, strong CHD-specific predictors are required, as exemplified by the cholesterol to HDL cholesterol ratio. A high cholesterol to HDL cholesterol ratio was inversely associated with the cause-specific hazard for non-CHD death that may be explained, in part, by a poor nutritional status in patients with a low ratio.<sup>38</sup> The prognostic value of the cholesterol to HDL cholesterol ratio is therefore of increasing importance in older women.

As the focus of our study was methodologic, we restricted the application to women and basic risk factors. The methods, however, are equally valid for men or for any prognostic model in the competing risks setting where the focus is on one event of interest. In other settings, all competing events may be of similar importance, and in the future models for all competing events jointly and corresponding accuracy criteria might expand into medical decision-making.

We encourage using competing risks models as a standard tool for developing predictive models in frail populations where a relevant proportion of patients do not experience the event of interest because they previously fail from a competing event. Our study illustrates how such models may be developed and their prognostic accuracy assessed. In the presence of competing risks, Kaplan-Meier estimates and naive applications of standard Cox regression overestimate the actual incidence of the event of interest and may lead to inappropriate risk stratification.

## REFERENCES

1. Jackson R, Lawes CM, Bennett DA, Milne RJ, Rodgers A. Treatment with drugs to lower blood pressure and blood cholesterol based on an individual's absolute cardiovascular risk. *Lancet*. 2005;365:434–441.
2. 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) final report. *Circulation*. 2002;106:3143–3421.
3. De Backer G, Ambrosioni E, Borch-Johnsen K, et al. European guidelines on cardiovascular disease prevention in clinical practice. Third Joint Task Force of European and other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of eight societies and by invited experts). *Atherosclerosis*. 2004;173:381–391.
4. Chobanian AV, Bakris GL, Black HR, et al. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA*. 2003;289:2560–2572.
5. Wilson PW, D'Agostino RB, Levy D, et al. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97:1837–1847.
6. 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;105:310–315.
7. Conroy RM, Pyörälä K, Fitzgerald AP, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J*. 2003;24:987–1003.
8. D'Agostino RB Sr, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation*. 2008;117:743–753.
9. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med*. 2007;26:2389–2430.
10. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. Hoboken, NJ: John Wiley and Sons; 2002.
11. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Stat Med*. 1999;18:695–706.
12. Koller MT, Schaer B, Wolbers M, Sticherling C, Bucher HC, Osswald S. Death without prior appropriate implantable cardioverter-defibrillator therapy: a competing risk study. *Circulation*. 2008;117:1918–1926.
13. Grunkemeier GL, Jin R, Eijkemans MJ, Takkenberg JJ. Actual and actuarial probabilities of competing risks: apples and lemons. *Ann Thorac Surg*. 2007;83:1586–1592.
14. Grunkemeier GL, Anderson RP, Miller DC, Starr A. Time-related analysis of nonfatal heart valve complications: cumulative incidence (actual) versus Kaplan-Meier (actuarial). *Circulation*. 1997;96(suppl 9):II-4.

15. Tsiatis A. A nonidentifiability aspect of the problem of competing risks. *Proc Natl Acad Sci USA*. 1975;72:20–22.
16. Beyersmann J, Dettenkofer M, Bertz H, Schumacher M. A competing risks analysis of bloodstream infection after stem-cell transplantation using subdistribution hazards and cause-specific hazards. *Stat Med*. 2007;26:5360–5369.
17. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94:496–509.
18. Fine JP. Regression modeling of competing crude failure probabilities. *Biostatistics*. 2001;2:85–97.
19. Scheike TH, Zhang M, Gerds TA. Predicting cumulative incidence probability by direct binomial regression. *Biometrika*. 2008;95:205–220.
20. Harrell FE. *Regression Modeling Strategies With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer; 2001.
21. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York: Springer; 2008.
22. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15:361–387.
23. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med*. 1999;130:515–524.
24. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982;247:2543–2546.
25. Antolini L, Boracchi P, Biganzoli E. A time-dependent discrimination index for survival data. *Stat Med*. 2005;24:3927–3944.
26. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. *Stat Med*. 2004;23:723–748.
27. 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;297:611–619.
28. Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol*. 1991;7:403–422.
29. Koller MT, Steyerberg EW, Wolbers M, et al. Validity of the Framingham point scores in the elderly: results from the Rotterdam study. *Am Heart J*. 2007;154:87–93.
30. Clark TG, Altman DG, De Stavola BL. Quantification of the completeness of follow-up. *Lancet*. 2002;359:1309–1310.
31. Gail MH, Pfeiffer RM. On criteria for evaluating models of absolute risk. *Biostatistics*. 2005;6:227–239.
32. Latouche A, Boisson V, Chevret S, Porcher R. Misspecified regression model for the subdistribution hazard of a competing risk. *Stat Med*. 2007;26:965–974.
33. Gonen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*. 2005;92:965–970.
34. Pencina MJ, D'Agostino RBSr, D'Agostino RBJr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;27:157–172.
35. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin Chem*. 2008;54:17–23.
36. Schemper M, Stare J. Explained variation in survival analysis. *Stat Med*. 1996;15:1999–2012.
37. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med*. 1999;18:2529–2545.
38. Liu Y, Coresh J, Eustace JA, et al. Association between cholesterol level and mortality in dialysis patients: role of inflammation and malnutrition. *JAMA*. 2004;291:451–459.
39. R Development Core Team. R: A Language and Environment for Statistical Computing (v 2.5.1). R Foundation for Statistical Computing. Available at: <http://www.R-project.org>. Accessed November 12, 2008.
40. Gray B. Subdistribution Analysis of Competing Risks. The cmprsk library (v 2.1–7). Available at: <http://cran.r-project.org>. Accessed November 12, 2008.

## APPENDIX

The statistical software R<sup>39</sup> was used for all analyses. For cause-specific hazards models, hazards were estimated using the survival library and the cumulative incidence function was calculated with the R function CumInc, which is part of the supportive code for the tutorial<sup>9</sup> and available at: [www.msbi.nl/multistate](http://www.msbi.nl/multistate). Fine and Gray regression models were fitted with the contributed R package cmprsk.<sup>40</sup> The first author of the present publication has written a formula interface for Fine and Gray regression as well as several other utility functions, including a function to calculate the c index and corresponding bootstrap confidence intervals. These functions are available on request.