Chapter 23 Case study on survival analysis: prediction of secondary cardiovascular events

Background
Survival is an important long-term outcome in prognostic research, including medical areas such as cardiovascular disease and oncology. We consider a model for the occurrence of vascular events in patients with symptomatic cardiovascular disease. Patient data were from the Second Manifestations of ARTerial disease (SMART) study. We go through the 7 steps of the checklist for developing valid prediction models, as presented in Part II. The final model looks promising, but needs external validation to proof its actual value. The data set and R code is made available at the book's website.

23.1 Prognosis in the SMART study

The SMART study is an ongoing prospective cohort study at the University Medical Center Utrecht, the Netherlands, initiated and led by Prof Van der Graaf and colleagues. The study was designed to

a) establish the prevalence of concomitant arterial diseases and risk factors for cardiovascular disease in a high-risk population;
b) identify predictors of future cardiovascular events in patients with symptomatic cardiovascular disease [388].

Currently available prediction models include the Framingham risk score, PROCAM, and SCORE [487 17 441]. There were all developed with data from subjects without clinically manifest atherosclerosis and cannot reasonably be used for patients with clinically manifest cardiovascular disease. These models may be able to rank patients with clinically manifest disease according to risk, but would be expected to underestimate absolute risk [113].

Assessment of absolute risk is important for secondary prevention. According to the current guidelines all patients who experienced a symptomatic cardiovascular event should be considered as at high risk (more than 20% absolute risk on a future event in the next 10 years). No further categorization is available.

Relevant outcomes in patients with cardiovascular disease (coronary artery disease, cerebral artery disease, peripheral arterial disease and abdominal aortic aneurysm) include stroke, myocardial infarction or cardiovascular death (Table 23.1). Other endpoints can be considered depending on the research question, e.g. including cardiovascular interventions. Hard outcomes are generally preferred because they lead to better comparability between studies and hence a better generalizability. The aim in the current study was to develop a prediction rule for patients with cardiovascular disease. We estimate the 1, 3 and 5 year risks on the occurrence of vascular events (stroke, myocardial infarction or cardiovascular death).

Table 23.1 Definitions of fatal and non-fatal vascular events in the SMART study

Event	Definition
Ischemic stroke	Definite: relevant clinical features that have caused an increase in impairment of at least one grade on the modified Rankin scale, accompanied by a fresh ischemic infarction on a repeat brain-scan Probable: clinical features that have caused an increase in impairment of at least one grade on the modified Rankin scale; without a fresh ischemic infarction on a repeat brain-scan
Myocardial infarction	Fatal or non-fatal myocardial infarction: at least two of the following criteria 1. chest pain for at least 20 minutes, not disappearing after administration of nitrates 2. ST-elevation > 1 mm in two following leads or a left bundle branch block on the ECG* 3. CK elevation of at least two times the normal value of CK* and a MB*-fraction > 5% of the total CK
Vascular death	Sudden death: unexpected cardiac death occurring within 1 hour after onset of symptoms, or within 24 hours given convincing circumstantial evidence Death from ischemic stroke Death from intracerebral hemorrhage (hemorrhage on CT-scan) Death from congestive heart failure Death from myocardial infarction Death from rupture of abdominal aortic aneurysm Vascular death from other cause, such as sepsis following stent placement

23.1.1 Patients in SMART

We consider 3873 patients who were enrolled in the study in the period September 1996 and March 2006. Patient had a clinical manifestation of atherosclerosis (transient ischemic attack, ischemic stroke, peripheral arterial disease, abdominal aortic aneurysm or coronary heart disease). After written informed consent, they underwent a standardized vascular screening including a health questionnaire for clinical information, laboratory assessment and anthropometric measurements at enrolment. During follow-up, patients were biannually asked to fill in a questionnaire on hospitalizations and outpatient clinic visits. When a possible event was reported by a participant, correspondence and relevant data were collected (discharge letters, laboratory and radiology results). Based on all obtained information, every event was audited by 3 physicians from different departments. The endpoints of interest for the present study were (acute) vascular death, (non-)fatal ischemic stroke or (non-)fatal myocardial infarction and the composite end point of any of these vascular events (Table 23.1). If a patient had multiple events, the first recorded event was used for analysis. Data were available on 14,530 person-years collected during a mean follow-up of 3.8 years (range 0-9 years). A total of 460 events occurred, corresponding to 1, 3, and 5 year cumulative incidences of 4.0%, 8.4%, and 14.1% respectively.

Table 23.2 Checklist for developing a valid prediction model in the SMART study

Step	Specific issues	SMART model
General considerations		
Research question	Aim: predictors / prediction?	Emphasis on prediction
Intended application	Clinical practice / research?	Clinical practice
Outcome	Clinically relevant?	Hard cardiovascular events
Predictors	Reliable measurement? Comprehensiveness	Detailed work-up; comprehensive set of candidate predictors
Study design	Retrospective/prospective? Cohort; case-control	Prospective cohort
Statistical model	Appropriate for research question and type of outcome?	Cox regression
Sample size	Sufficient for aim?	3873 patients, 460 events: very good
7 modeling steps		
1. Preliminary	Inspection of data Missing values	Table 23.3 Multiple and single imputation
2. Coding of predictors	Continuous predictors Combining categorical predictors Combining predictors with similar effects	Truncation and spline transformations for continuous predictors; sum scores for cardiovascular history
3. Model specification	Appropriate selection of main effects? Assessment of assumptions (distributional, linearity and additivity)?	Stepwise selection with high p-value and Lasso Additivity checked with interaction terms, one included Proportional hazards checked
4. Model estimation	Shrinkage included? External information used?	Penalized estimation with Lasso No
5. Model performance	Appropriate measures used?	Focus on discrimination
6. Model validation	Internal validation including model specification and estimation? External validation?	Bootstrap No external validation
7. Model presentation	Format appropriate for audience	Nomogram
Validity		
Internal: overfitting	Sufficient attempts to limit and correct for overfitting?	Large sample size, predictors from literature, Lasso for selection and shrinkage
External: generalizability	Predictions valid for plausibly related populations?	Large set of predictors, representing important domains; not assessed in this study

23.2 General considerations in SMART

23.2.1 Research question and intended application

The aim was to develop a prediction model for long term outcome. Given the available follow-up, 1, 3 and 5 year risks could be assessed. Achieving adequate predictions was more prominent than insight in the predictor effects per se (Table 23.2). The intended application was in patient counseling; a high absolute risk might motivate patients to change inappropriate lifestyles and to comply with their medication regimens.

23.2.2 Outcome and predictors

The primary outcome was any cardiovascular event, comprising cardiovascular death, non-fatal stroke and non-fatal myocardial infarction. Combining different events is a common approach in vascular research to increase statistical power. A cardiovascular event occurred in 460 patients during follow-up.

The selection of predictors was motivated by characteristics included in Framingham and SCORE models. The relation with future events has also been established for several traditional risk factors, including hyperhomocysteinemia, intima media thickness, and creatinin [98 165]. Other candidate predictors were demographics (sex and age), and risk factors for vascular events in the general population (smoking, alcohol use, body mass index (BMI), diastolic and systolic blood pressure, lipids, and diabetes). It is well conceivable that indicators of the extent of atherosclerosis are very relevant to predict events in patients with symptomatic atherosclerosis. Such indicators are the location of symptomatic vascular disease (cerebral, coronary, peripheral arterial disease or AAA), and markers of the extent of atherosclerosis (homocysteine, creatinin, albumin, intima media thickness (IMT) and presence of a carotid artery stenosis, Table 23.3). In sum, a relatively limited set of well-defined predictors was studied.

23.2.3 Study design and analysis

The SMART study is designed as an ongoing, prospective dynamic cohort study. Patients are enrolled when presenting at the hospital, with follow-up starting from study inclusion. We used the Cox regression model, which is the default statistical model for survival outcomes. This model is appropriate for prediction of an outcome at relatively short-term such as 5 year cumulative incidence of cardiovascular events. For long-term predictions (e.g. 10 year incidences), a parametric model might be preferable such as a Weibull model. A Weibull model provides more stable estimates at the end of the follow-up [312 65].

With respect to sample size, the balance of 460 events and approximately 25 candidate predictors is reasonable (Table 23.3). At least 10 to 20 events per candidate predictor have been proposed in previous guidelines for sensible development of a prediction model [175 326 410].

Table 23.3 Potential predictors in the SMART study data set (n=3873).

Characteristics	
Demographics	
Female sex ('SEX', n, 0 missing)	975 (25%)
Age ('AGE', in years, 0 missing)	60 [52 – 68]
Classical risk factors	
Smoking ('SMOKING', n (%), 25 missing)	
Never	693 (18%)
Former	2711 (70%)
Current	444 (12%)
Packyears ('PACKYRS', in years, 21 missing)	20 [6 – 34]
Alcohol ('ALCOHOL', n (%), 25 missing)	
Never	751 (20%)
Former	408 (11%)
Current	2689 (69%)
Body mass index ('BMI', in kg/m^2, 3 missing)	26.7 (24 – 29)
Diabetes ('DIABETES', n (%), 40 missing)	846 (22%)
Blood pressure	
Systolic, by hand ('SYSTH', in mm Hg, 1498 missing)	140 (126 – 155)
Systolic, automatic ('SYSTBP', in mm Hg, 1223 missing)	139 (127 – 154)
Diastolic, by hand ('DIASTH', in mm Hg, 1499 missing)	82 (75 – 90)
Diastolic, automatic ('DIASTBP', in mm Hg, 1221 missing)	79 (73 – 86)
Lipid levels	
Total cholesterol ('CHOL', in mmol/L, 18 missing)	5.1 [4.4 – 5.9]
High-density lipoprotein cholesterol ('HDL', mmol/L, 30 missing)	1.17 [0.96 – 1.42]
Low-density lipoprotein cholesterol ('LDL', mmol/L, 216 missing)	3.06 [2.39 – 3.83]
Triglycerides ('TRIG', mmol/L, 28 missing)	1.54 [1.12 – 2.23]
Previous symptomatic atherosclerosis	
Cerebral ('CEREBRAL', n (%), 0 missing)	1147 (30%)
Coronary ('CARDIAC, n (%), 0 missing)	2160 (56%)
Peripheral ('PERIPH, n (%), 0 missing)	940 (24%)
Abdominal aortic aneurysm ('AAA', n (%), 0 missing)	416 (11%)
Markers of atherosclerosis	
Homocysteine ('HOMOC', µmol/L, 463 missing)	12.8 [10.3 – 15.7]
Glutamine ('GLUT', µmol/L, 19 missing)	5.7 [5.3 – 6.5]
Creatinine clearance ('CREAT', mL/min, 17 missing)	89 [78 – 101]
Albumin ('ALBUMIN', n (%), 207 missing)	
No	2897 (79%)
Micro	655 (18%)
Macro	114 (3%)
Intima media thickness ('IMT', mm, 98 missing)	0.88 [0.75 – 1.07]
Carotid artery stenosis >50% ('STENOSIS', n (%), 93 missing)	722 (19%)

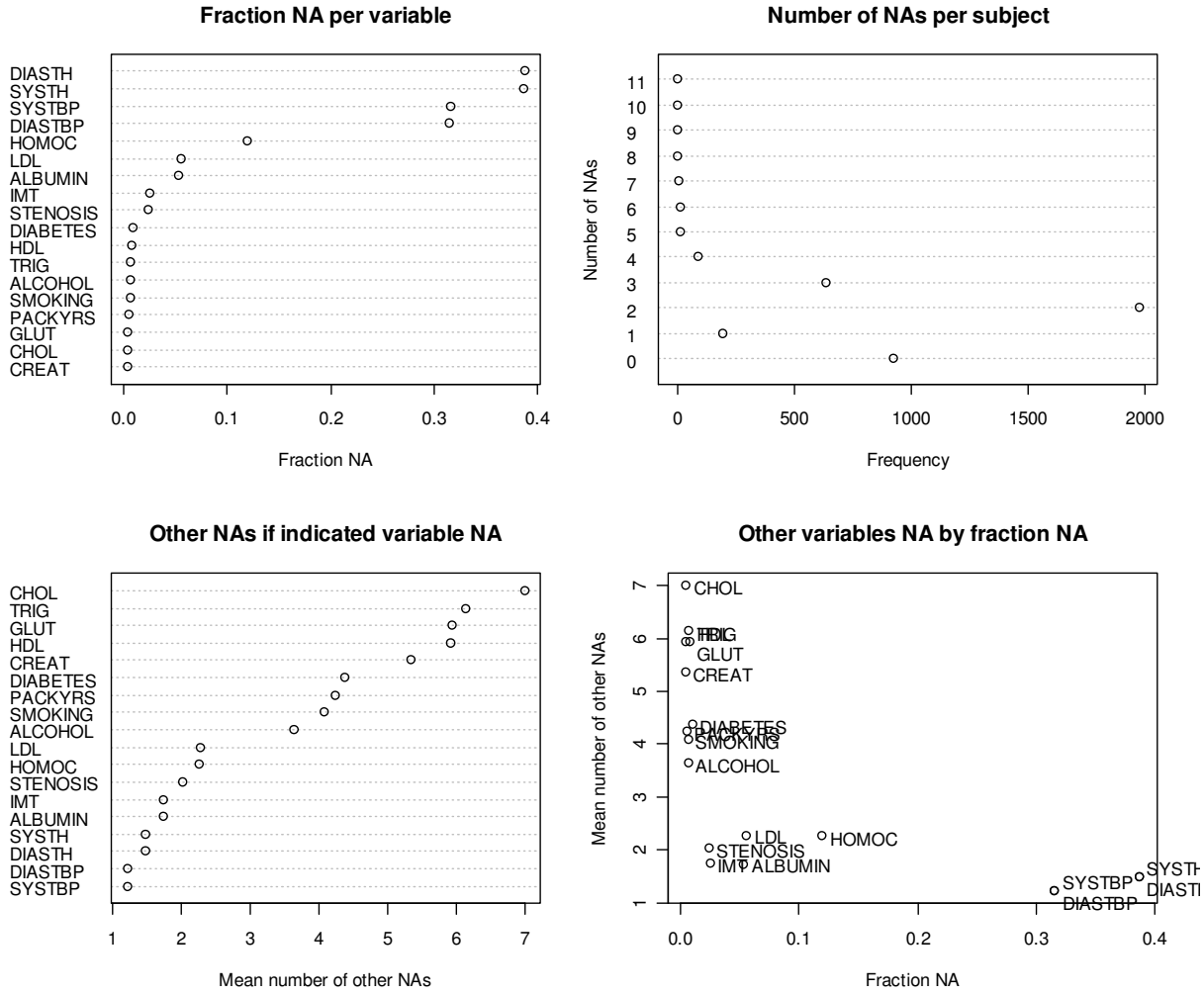

Fig 23.1 Patterns of missing data in the SMART study (n=3873, `na.plot2` function). NA: not available ('missing')

23.3 Preliminary modeling steps in the SMART cohort

It appeared that the number of missing values was rather limited for most of the 18 potential predictors (<1%, Table 23.3 and Fig 23.1). Many missings were however noted among 4 variables that relate to blood pressure measurements (2 for diastolic and 2 for systolic pressure). In the first years of the study, blood pressure was measured combined with measurement of the distensibility of the carotid artery wall ('SYSTBP' and 'DIASTBP' variables). Four years after the start of the study it was decided to measure blood pressure with the conventional sphygmomanometry as well. This measurement is considered in most current guidelines. Hence, conventional diastolic and systolic measurements are obvious candidate predictors for our model rather than the automated measurements. Nearly all patients had at least one type of blood pressure measurement, but many had missing values for conventional sphygmomanometry (n=1498, 'SYSTH' and 'DIASTH' variables). Pearson correlation coefficients were 0.69 and 0.59 for systolic and diastolic blood pressure measurements in 1155 and 1156 patients with conventional as well as automatic measurements available, respectively.

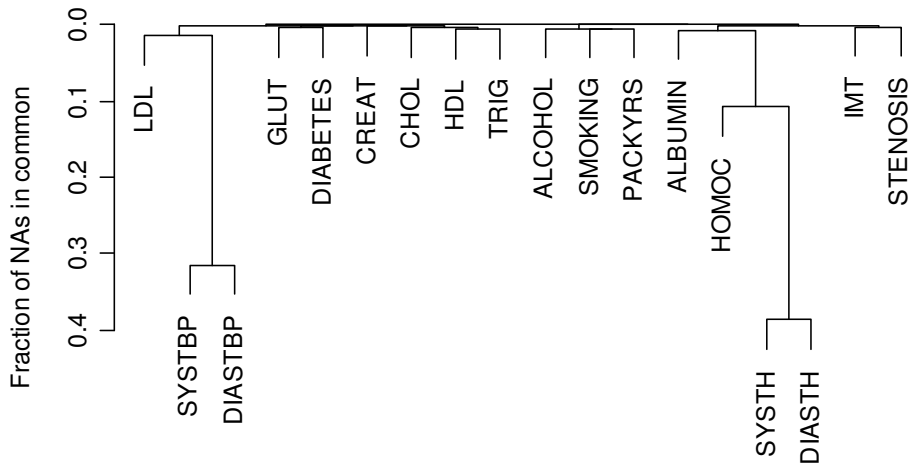

Fig 23.2 Cluster analysis of patterns of missingness in the SMART study (n=3873, `naclus` function).

The variable homocysteine ('HOMOC') had 463 missings (12%, Table 23.3, Fig 23.1, upper left panel). This was related to the fact that homocysteine was not routinely measured in the first years of the study. This is a typical 'missing completely at random' (MCAR) situation. Also for the other variables we assume that missingness was more related to logistic reasons, because all patients underwent the same screening protocol. The decision to measure variables was not obviously dependent on other observations (MAR mechanism), the values of the characteristic itself, or characteristics not available in our dataset (MNAR mechanisms).

A total of 925 patients had no missings values among the 18 potential predictors, and 1975 had 2 missings values (mostly: 1 type of blood pressure measurement not performed). A few patients had many missings (18 with 7 or more missings, Fig 23.1, upper right panel). If one type of blood pressure measurement was missing, few other variables had missings values. If cholesterol or triglycerides were missing (which was very rare), many other predictors were also missing (Fig 23.1, lower left and lower right panel). Further details on the combinations of missing values are shown in Fig 23.2. Again we note that the diastolic and systolic blood pressure measurements are always jointly missing. In the early years of the study, both homocysteine ('HOMOC') and conventional sphygmomanometry blood pressure measurements ('SYSTH' and 'DIASTH' variables) were not performed, leading to some correlation of missingness between these variables.

Missing data per predictor would lead to a substantial loss of information if only complete cases were used in the multivariable model. We therefore used multiple imputation techniques (`aregImpute` function) to replace the missing values (Fig 23.3). The set of first imputations was used for further analyses ('single imputation'). Although multiple imputation is preferable from a theoretical view point, single imputation was considered more practical and sufficient to obtain reasonable predictions (Chapter 7). Final models were also constructed with multiple imputed data sets to check for any relevant differences in point estimates, and widening of confidence intervals.

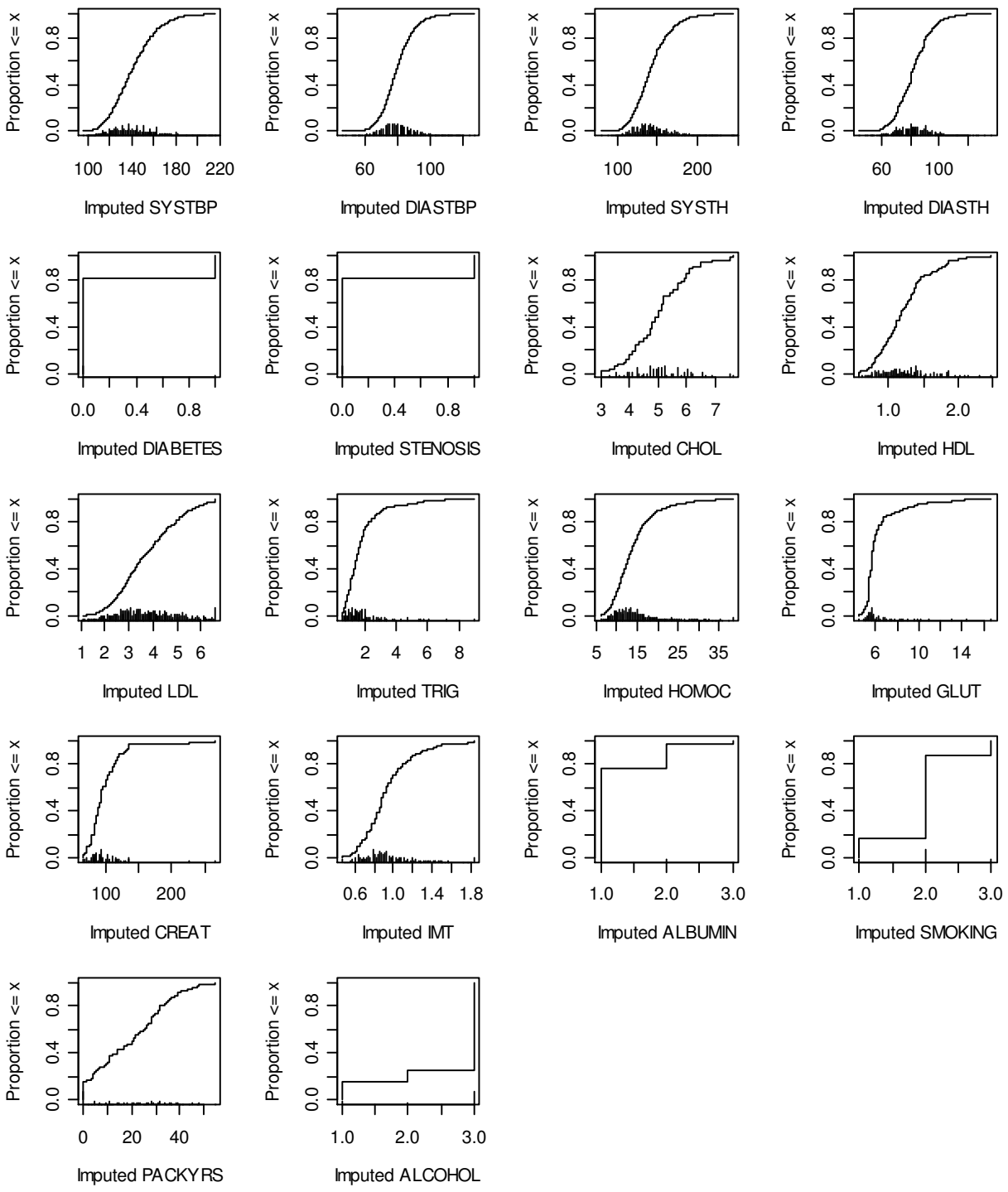

Fig 23.3 Distribution of imputed values for the 18 most relevant predictors which had missing values in the SMART study

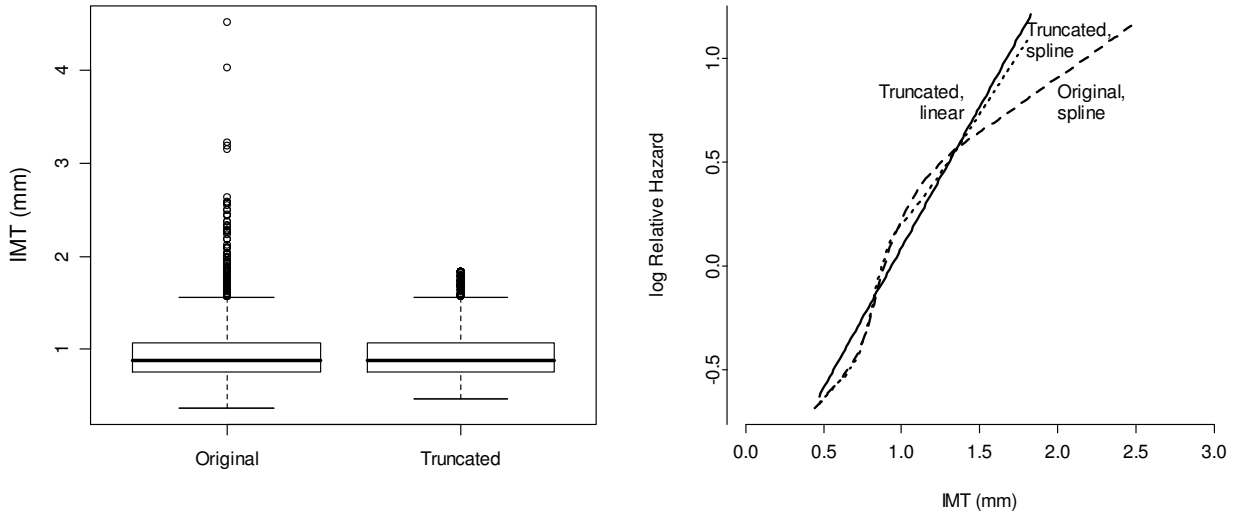

Fig 23.4 Boxplot of intima media thickness (IMT, in mm, left panel) before and after truncation, and a plot of the effect of IMT on cardiovascular events in a univariate Cox regression model (right panel). The original IMT values are sometimes extremely high, leading to a spline with flattens off with high IMT values. The truncated IMT values have a smaller range and lead to a quite linear relationship (solid line, linear term; dotted line, spline).

23.4 Coding of predictors

23.4.1 Extreme values

Before any modeling started, the distributions of all potential predictors were carefully examined for extreme values. Preferably data are checked with the source documents but sometimes such decisions have to be made on common sense. Biologically implausible values were set to missings values, and remaining extreme values were truncated by shifting the values below the 1th centile and above the 99th centile to 'truncation points' (Chapter 9). Such truncation may prevent distortion of the relationship between predictor and outcome due to high leverage of the extreme values, which is not desirable [356].

We truncated extreme values for intima media thickness (IMT, Fig 23.4). The mean IMT was 0.94 mm, but some patients had measurements as high as 4 mm. These high values are the result of plaque formation in the carotid artery, and may have an unduly large influence on estimates of cardiovascular event risk. A total of 51 values higher than 1.83 were shifted to 1.83 (the upper truncation point), and 13 values below 0.47 were shifted to 0.47 (the lower truncation point). We note a substantial effect of truncation on the relationship between IMT and outcome (Fig 23.4, right panel). A restricted cubic spline based on the original IMT values flattens off with high IMT (>1.5 mm), while a restricted cubic spline based on the truncated IMT values is very close to a straight line. This finding illustrates that truncation may obviate the need for a non-linear transformation. Before truncation the Cox regression coefficient for a linear IMT variable was 0.91, while it was 1.36 after truncation. The univariate model χ^2 improved from 61 before to 75 (1 df) after truncation. Similarly we truncated body mass index, lipids (Cholesterol, HDL, LDL, Triglycerides), homocysteine and creatinin levels by shifting values below the 1th and above the 99th centile to the truncation points.

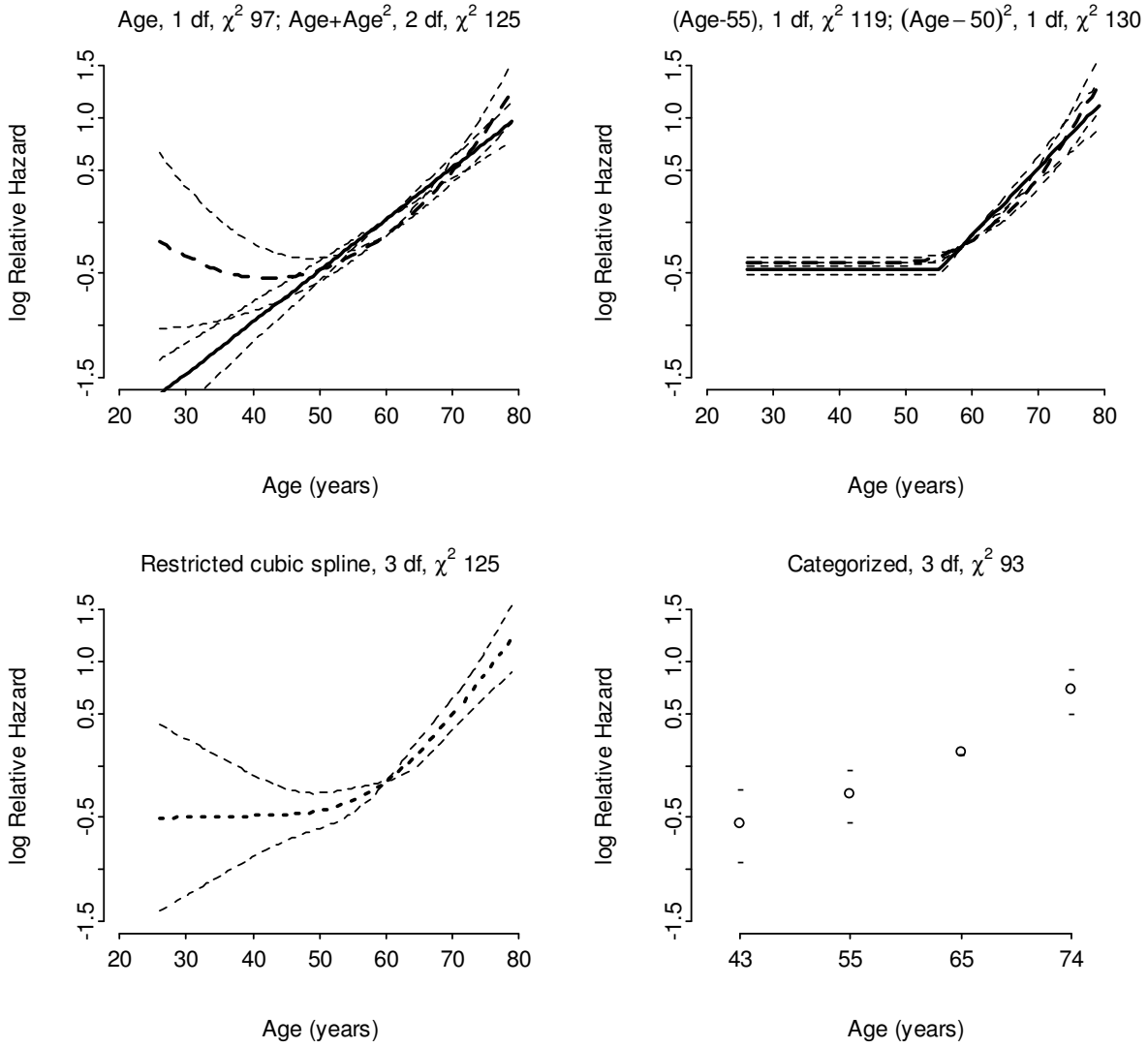

Fig 23.5 Transformations of age in univariate analysis of the SMART study. Upper left: age linear and age plus age squared; Upper right: age linear after 55 years ("Age–55)$_+$)" and age squared above 50 years ("Age–50)$_+^2$)"; Lower left: restricted cubic spline, 4 knots, 3 df; Lower right: age categorized in 4 groups.

23.4.2 Transforming continuous predictors

Age is an important predictor of cardiovascular events. We considered several age transformations (Fig 23.5, Table 23.4). In our cohort the Wald χ^2 of the linear fit was 97. Adding age^2 increased the χ^2 to125, but there was a biologically implausible increased risk below age 40 years. Based on visual inspection (Fig 23.5), it may be judged reasonable to assume no age effect till age 55, and a linear effect for age>55 years ("$(Age–55)_+$" variable, χ^2 119). A transformation such as $(Age–50)_+^2$ led to an even better model (χ^2 130, Fig 23.5). A restricted cubic spline with 3 df (4 knots) did not describe the relationship of age to outcome better (χ^2 125). Categorizing in quartiles has a clearly lower performance (χ^2 93). Such categorization should not be used because jumps in predictions are unnatural. Dichotomizing at age 60 years (close to the median of 61 years) led to a substantial decrease performance (χ^2 72, Table 23.4), illustrating that dichotomization is "a bad idea" [355].

Other continuous predictor variables were examined in a similar way; some examples are shown in Table 23.4. For creatinine, a log transformation gave the best fit (Fig 23.6). A linear coding of systolic blood pressure was reasonable, and diastolic blood pressure had no effect when we analyzed the conventional sphygmomanometry blood pressure measurements ('SYSTH' and 'DIASTH' variables). All analyses were repeated with multiply imputed data sets, with largely similar results.

Table 23.4 Impact of various codings of predictors in a univariate Cox regression models for the SMART study.

Predictor	Coding	Wald χ^2	df
Age	Linear	97	1
	Squared	125	2
	$(Age–55)_+$: linear effect after age 55	119	1
	$(Age–50)_+^2$: square effect after age 50	130	2
	Restricted cubic spline, 3 df	125	3
	<50, 50-59.9, 60-69.9, ≥70	93	3
	<60, ≥60	72	1
Creatinine	Linear	93	1
	Restricted cubic spline, 3 df	116	3
	Restricted cubic spline, 2 df	99	2
	Log	131	1
Blood pressure (conventional reading)	Linear systolic	15	1
	Restricted cubic spline systolic, 2 df	15	2
	Linear diastolic	0.7	1
	Restricted cubic spline diastolic, 2df	2	2
Previous symptomatic atherosclerosis	Sumscore 0-4	96	1
	Sumscore 0-5 (AAA=2)	119	1
	Separate terms	123	4
	Cerebral	36	1
	Coronary	19	1
	Peripheral	23	1
	Abdominal Aneurysm Aorta	96	1

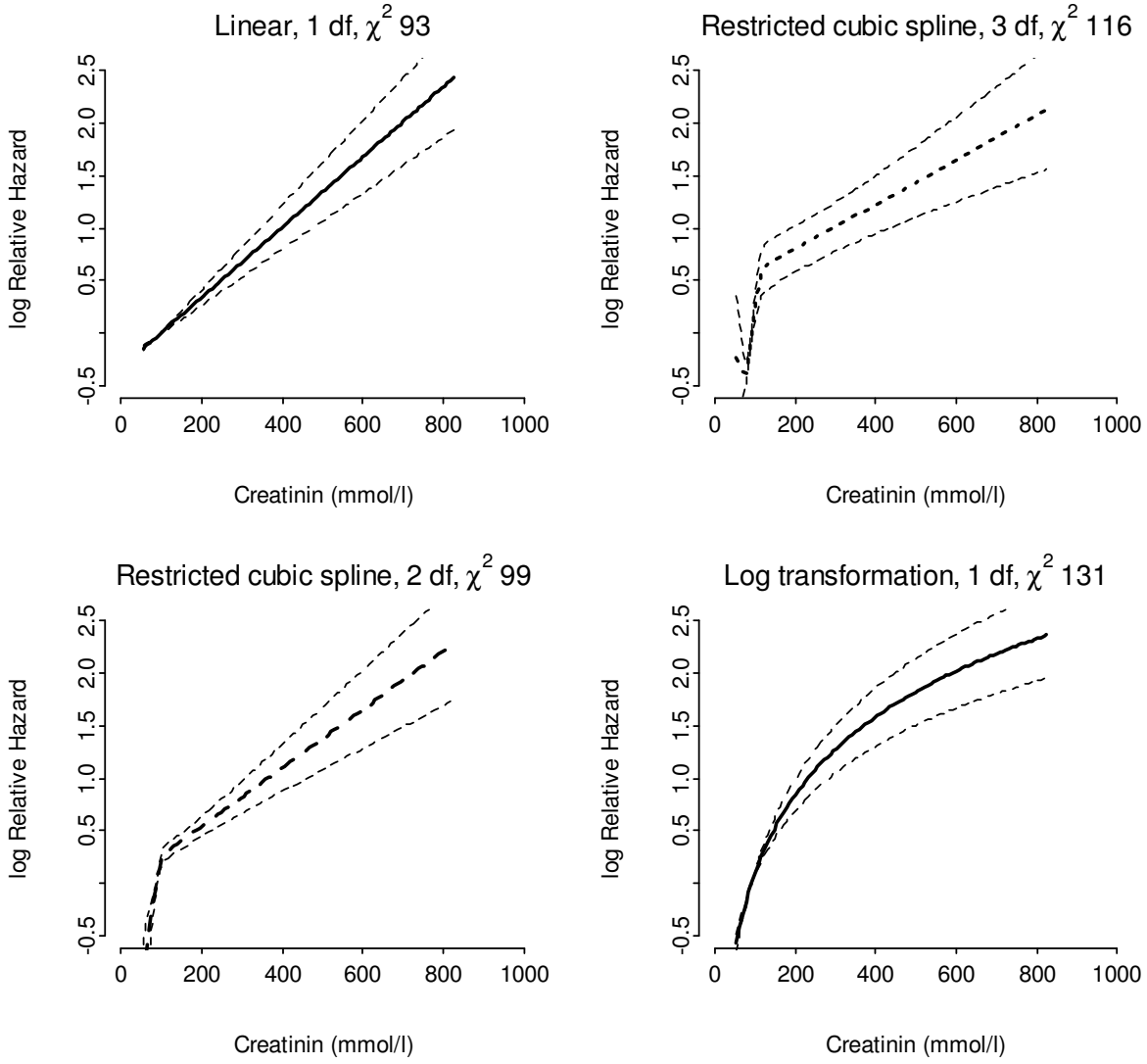

Fig 23.6 Transformations of creatinine in univariate analysis of the SMART study.

23.4.2 Combining predictors with similar effects

Combining predictors with similar effects can be an effective way to limit the degrees of freedom of predictors in a model (Chapter 10). In atherosclerotic patients several variables reflect the extent of atherosclerosis. The affected organs reflect the load of atherosclerosis in one particular patient. The location of symptomatic events (cerebral, coronary, abdominal aortic aneurysm (AAA), peripheral artery disease) can be entered separately in the model. For each parameter we would spend 1 df, resulting in a model χ^2 of 123 (4 df, Table 23.4). If we combine the presence of previous vascular events in 1 variable, simply by assuming equal weights for each condition, the model χ^2 is 96 (1 df). The difference of the two models is a χ^2 of 27, which is highly significant at 3 df. Separate terms hence lead to a much better fit. When we test for the separate contributions of each localization it appears that the contribution of an AAA is considerably higher than the contribution of the other localizations. If we attribute 2 points for the presence of an AAA, the sumscore performs remarkable better (range 0 – 5, model χ^2 119, close to 123 for separate terms, Table 23.4).

23.5 Model specification

A full, main effects model was defined which included the common demographics age and sex, important classical risk factors (smoking status, alcohol use, body mass index, blood pressure, lipid levels and diabetes), the sum score for previous symptoms of atherosclerosis, and finally markers of the extent of the atherosclerotic process (including hyperhomocysteinemia, creatinin, intima media thickness of the carotid artery, carotid artery stenosis, and albuminuria). We focused on systolic blood pressure since recent publications stress the more important role of systolic rather than diastolic blood pressure in predicting cardiovascular events [428]. The full model consisted of 14 predictors, with several having rather limited contributions (Table 23.5). Predictors with a large prognostic strength were age (χ^2 33), the localization of the symptom of atherosclerosis (sumscore χ^2 36), and the marker of renal damage creatinin (χ^2 23). Other characteristics had much smaller prognostic relevance, with some impact of the general marker of atherosclerosis intima media thickness (χ^2 9.9), but a minor contribution of homocysteine. The classical risk factors had at most a χ^2 of 6 (for HDL) and hence hardly contributed to the model predictions.

We tested interactions between the predictors and gender by including cross-product terms with predictors in the selected model (overall χ^2 15, 10 df, p=0.14). The strongest interaction was between sex and the sumscore for previous symptomatic atherosclerosis (χ^2 8.1, 1 df, p=0.004). In all, the interactions were not considered relevant enough to include an interaction term with sex in the model. We also tested proportionality of hazards. The overall test was not significant (overall χ2 12, df 10, p=0.27, `cox.zph` function). Detailed results of the assessment of interactions are provided at the web.

23.5.1 Selection

We judged our sample size as large enough to allow for some model reduction for easier practical application (460 events, full model with 17 degrees of freedom, ignoring that the coding of predictors also consumed some degrees of freedom). One approach was to apply a backward selection procedure with a higher than standard p-value. We used Akaike's Information Criterium (AIC), which implies a p-value < 0.157 for selection of predictors with 1 df [14].

A promising alternative is to apply the Lasso method, which achieves selection of predictors by shrinking some coefficients to zero by setting a constraint on the sum of the absolute standardized coefficients [435]. The Lasso model was found to be optimal with 10 predictors, but in this model, the coefficient of homocysteine was close to zero. With more shrinkage, this predictor was dropped, and the same set of 9 predictors was selected as in the stepwise selection procedure with AIC (Table 23.6).

Table 23.5 Hazard ratios (HR) and contribution to Cox regression model (χ^2 and df) of the predictors in a full model for cardiovascular events in the SMART study. A single imputed data set was used with n=3873.

Predictor	HR [95% CI]*	χ^2	df
$(Age–50)_+^2$ (years above 50)	1.5 [1.3–1.7]	39	1
Gender (male)	0.9 [0.7–1.2]	0.1	1
Classical risk factors			
Smoking		1.1	2
Never	0.9 [0.7–1.2]		
Former	1		
Current	1.1 [0.7–1.6]		
Alcohol		1.1	2
Never	1.2 [0.8 –1.6]		
Former	1		
Current	1.1 [0.8–1.4]		
Body mass index (kg/m^2)	0.9 [0.8–1.0]	3.2	1
Systolic blood pressure (mm Hg)	1.0 [0.9–1.2]	0.3	1
HDL	0.8 [0.7–1.0]	5.4	1
Diabetes	1.3 [1.0–1.8]	4.5	1
Previous symptomatic atherosclerosis			
Sumscore (AAA 2 points)	1.4 [1.3–1.6]	37	1
1			
2			
3			
4			
5			
Markers of atherosclerosis			
Homocysteine (mmol/l)	1.0 [0.9–1.1]	0.2	1
Creatinin (mmol/l)	1.2 [1.1–1.3]	24	1
Albumin		5.2	2
No	0.8 [0.6–1.0]		
Micro	1		
Macro	1.1 [0.7–1.7]		
Intima media thickness (mm)	1.2 [1.1–1.3]	10	1
Carotid artery stenosis >50%	1.2 [1.0–1.5]	3.6	1

* Hazard ratio [95% confidence interval] refers to interquartile range for continuous predictors

Table 23.6 Cox regression coefficients in the full model, a stepwise selected model (using Akaike's Information Criterion), and in the Lasso model.

Predictor	Full	Stepwise (AIC)	Lasso
$(\text{Age}–50)_+^2$ (years above 50)	0.0013	0.0013	0.0012
Gender (male)	–0.049	not selected	not selected
Smoking		not selected	not selected
Never	0		
Former	0.13		
Current	0.21		
Alcohol		not selected	not selected
Never	0		
Former	–0.15		
Current	–0.11		
Body mass index (kg/m^2)	–0.025	–0.026	–0.001
Bloodpressure (mm Hg)	0.0012	not selected	not selected
HDL	–0.37	–0.39	–0.16
Diabetes	0.23	0.23	0.11
Previous vascular disease	0.34	0.35	0.33
Homocysteine (mmol/l)	0.0042	not selected	not selected
Log(creatinin) (mmol/l)	0.68	0.71	0.71
Albumin			
No	0	0	0
Micro	0.22	0.24	0.13
Macro	0.35	0.35	0.20
Intima media thickness (mm)	0.55	0.56	0.50
Carotid artery stenosis >50%	0.20	0.22	0.16

23.6 Model estimation, performance, validation, and presentation

23.6.1 Model estimation

Regression coefficients were first estimated as default with Cox regression analysis, i.e. by maximizing the log-likelihood of the fit of the model to the data . The coefficients of the 9 predictors in the stepwise backward selected model were rather similar to their corresponding coefficients in the full model (Table 23.6). In contrast, the Lasso model shrunk coefficients of weaker predictors such as BMI, HDL, diabetes and albumin considerably towards zero. The effects of strong predictors, such as age, sumscore for atherosclerosis, creatinin, IMT and carotid artery stenosis, were comparable with the maximum likelihood estimates, but effects of weaker predictors are shrunk considerably (Fig 23.7).

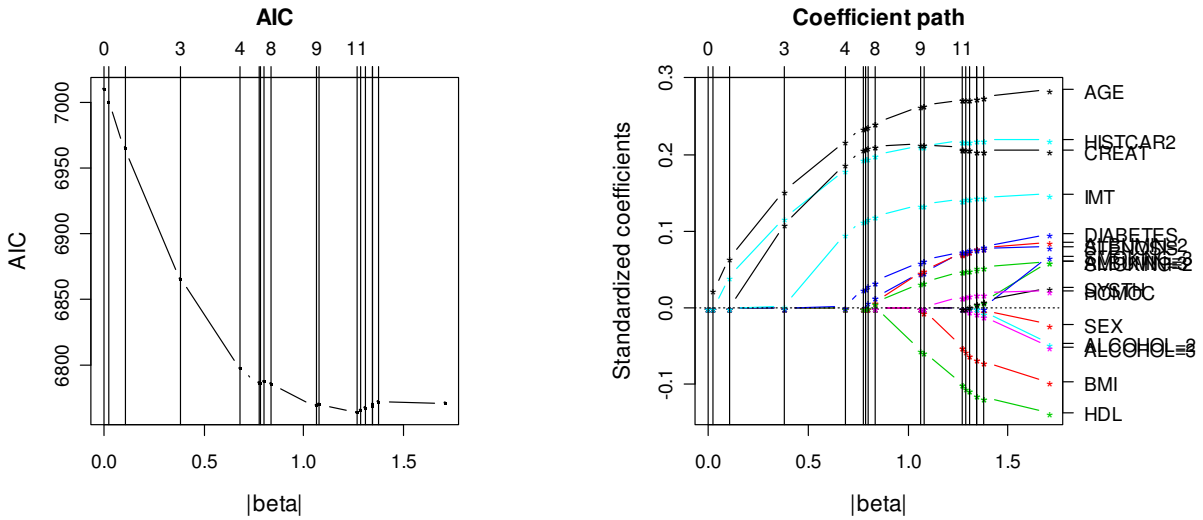

Fig 23.7 Lasso path with increasing sum of the absolute standardized coefficients (|beta|). The optimal AIC is obtained with 11 predictors, but differences are small between models with 9 to 12 predictors. The coefficient path shows that predictors have effects other than zero with higher |beta|.

23.6.2 Model performance

Discrimation of the final model was indicated by the *c* statistic, which was 0.693 (95% CI 0.65-0.73). Discrimination was further illustrated by dividing the predictions in quartiles, and plotting the Kaplan-Meier curves of these 4 groups (Fig 23.8). We note that patients in the lower quartile had a considerably poorer chance of being free of cardiovascular events during follow-up: around 75% at 5 years of follow-up, and near 50% at 9 years of follow-up.

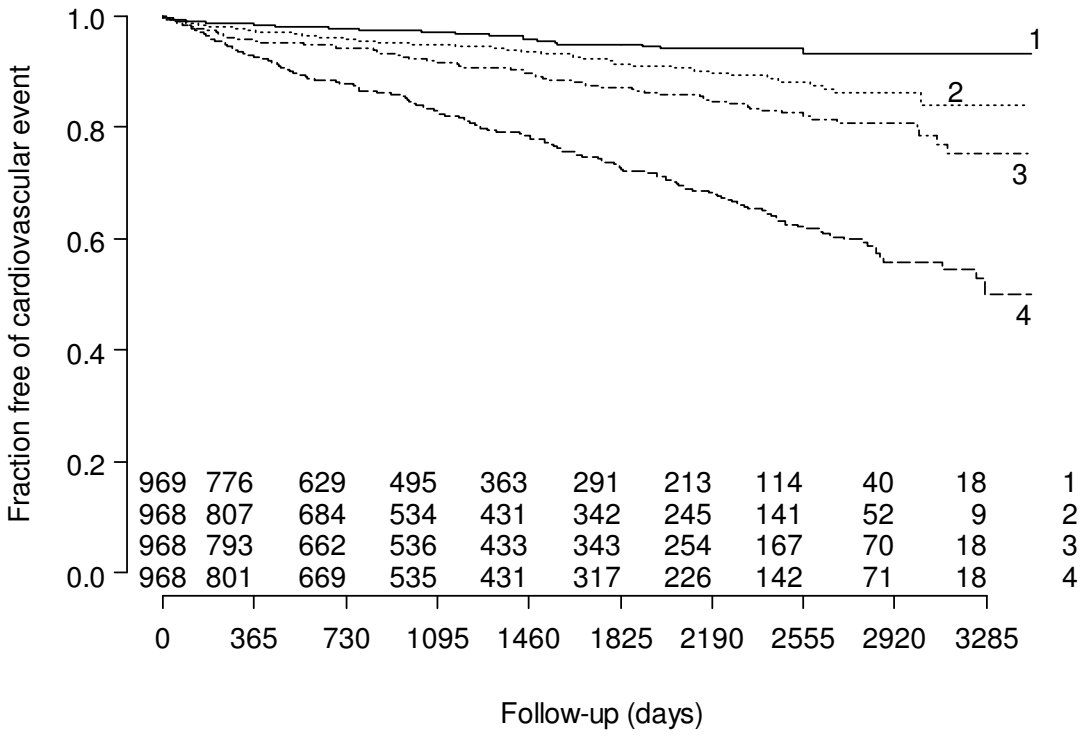

Fig 23.8 Fraction free of cardiovascular event according to quartiles of the linear predictor. Numbers at risk are indicated for to the upper to lower quartile (numbered 1 – 4).

23.6.3 Model validation: stability

We used a bootstrap resampling procedure to study the stability of our stepwise selected model, and to quantify the optimism of our modeling strategy. We found that age and localization of symptoms were strong predictors and were always selected when we repeated stepwise selection procedure in 200 bootstraps (Table 23.7). In contrast, sex, smoking, alcohol and systolic blood pressure were selected in only 26%, 40%, 36% and 57% of the bootstrap samples respectively, consistent with their exclusion from the stepwise model. Albumin, HDL and IMT were selected in the majority of the bootstraps, but not in all.

There was a clear association between the estimated effect of a predictor according to the Lasso and the frequency of selection in the bootstrap procedure. The coefficients for BMI, HDL, and diabetes were considerably reduced according to the Lasso, and indeed these were not selected in 40%, 22%, and 32% of the bootstrap samples. Instead of excluding the predictor, which is equivalent to setting the coefficient to zero, the coefficient was shrunk towards zero. The coefficients of age, localization of symptoms, creatinin and IMT were virtually not affected by the Lasso, consistent with their selection in over 95% of the bootstraps.

Table 23.7 Frequency of selection of predictors from the full model in the backward stepwise selected model (with AIC). Results are shown for the first 20 bootstrap samples.

AGE	SEX	SMOKING	ALCOHOL	SYSTH	BMI	HDL	DIABETES	HISTCAR2	HOMOC	CREAT	ALBUMIN	STENOSI	IMT
*	*							*		*	*		*
*	*		*	*			*	*	*	*	*	*	*
*					*	*	*	*	*	*	*		*
*		*	*	*			*	*	*	*	*		*
*		*	*		*	*	*	*		*	*	*	*
*	*		*		*	*	*	*		*	*	*	*
*					*	*	*	*		*	*	*	*
*	*		*	*	*			*		*	*	*	*
*		*	*	*	*		*	*		*		*	*
*				*		*		*		*	*	*	*
*		*		*	*	*		*		*	*	*	*
*						*		*		*	*	*	*
*		*	*			*		*	*	*	*		*
*			*	*		*	*	*	*	*		*	*
*					*	*		*		*	*		*
*	*		*	*	*	*	*	*		*	*		*
*		*	*		*		*	*		*	*	*	
*	*			*	*	*	*	*		*			*
*					*	*	*	*		*	*		*
*				*	*	*	*	*		*	*	*	*
	100	26	40	36	57	60	78	68	100	38	99	84	63	95

23.6.4 Model validation: optimism

The c-statistic was expected to decrease from 0.693 to 0.680, or a decrease of 0.013, in a bootstrap procedure with repeated selection of predictors in every bootstrap sample. We estimated the required shrinkage for the coefficients in the stepwise selected model as 0.94, suggesting that each coefficient should be reduced by 6% to correct for optimism of the modelling process. Instead of using this shrinkage factor for the final model, we used the Lasso coefficients, which reduce coefficients for weak predictors more than for strong predictors. In all, the bootstrap validation procedure showed some instability of the model specification, and a modest amount of optimism in the final model.

23.6.5 Model presentation

The results of the modeling process can be presented in various ways. From Table 23.4 we learn about the relative contributions of each predictor to the model. For a survival model such as the SMART prediction model, an attractive way is to present the model as a nomogram (Fig 23.9). In the nomogram, we can judge the relative importance of each predictor by the number of points attributed over the range of the predictor, and we can calculate 3-year and 5-year survival estimates. Survival relates here to the probability of being free of a cardiovascular event.

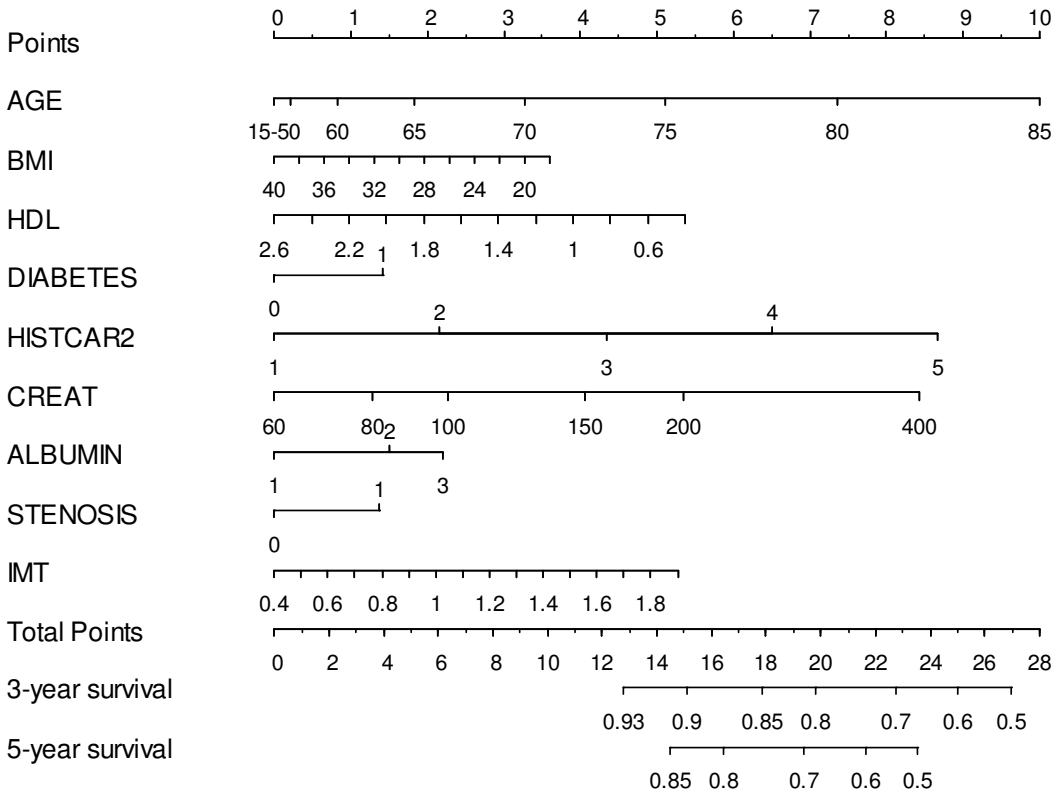

Fig 23.9 Nomogram to calculate predicted 3-year and 5-year survival (probability of being free of a cardiovascular event). Coefficients are based on the Lasso model. For example, a 75-year old patient, with a BMI of 28, HDL 1, no diabetes, previous aortic aneurysm but no other symptoms of atherosclerosis (HISTCAR2=2), a creatinin value of 100, low albumin, no carotid stenosis, IMT of 1 mm, has a total points score of 5 + 2 + 4 + 0 + 2 + 2 + 0 + 0 + 2 = 17. This corresponds to predicted 3 and 5-year survival of approximately 85% and 80% respectively.

23.7 Concluding remarks

This case study illustrates how a prediction model can be developed and internally validated for a survival analysis problem. We recognize that not all modeling steps could be considered in the bootstrap procedure for internal validation. Further external validation is necessary in the same setting (with more recent patients) and in other settings (to assess transportability).

We note a distinction between risk factors in the general population (without cardiovascular disease) and prognostic factors in patients with symptomatic disease. Classical risk factors such as smoking, alcohol use, BMI, blood pressure, HDL and diabetes, had very limited prognostic value in the clinical setting. These characteristics are hence not useful to predict future events once cardiovascular disease has developed. Indicators of previous symptomatic cardiovascular disease and the extent of atherosclerosis were more useful. This finding is similar to findings in the GUSTO-I sample, where e.g. smoking was associated with a better outcome after acute MI.

Questions

23.1 Composite outcomes (section 23.2.2 and Table 23.1)

Outcomes were combined in the presented analyses.

a) What does this imply about the effects of the predictors for each outcome?
b) How could this be tested? See Glynn & Rosner [140]

23.2 Missing values (Fig 23.1)

a) Some might argue to exclude patients with many missing values. What would be a reasonable number as maximum of missing values per patient in this analysis?
b) We note that missing values occur together for some predictors. We could also choose to exclude patients with missing values ('NA') in specific predictors. Which would you choose?

23.3 Effects of Lasso versus stepwise selection (Table 23.6)

We select the same predictors with a Lasso procedure as with stepwise selection using AIC.

a) How is it possible to obtain the same selection with these very different methods?
b) The effect of age is similar with both methods, while the effect of BMI is very weak according to the Lasso. How is this possible? Consider also the validation in Table 23.7.