UNIVERSITA' DEGLI STUDI DI MILANO – BICOCCA

Scuola di Dottorato in Scienze Mediche Sperimentali e Cliniche

Dottorato in Epidemiologia e Biostatistica – XXVI Ciclo

Development, validation and clinical utility of a long-term cardiovascular disease risk prediction model in the Italian population

Tutor: Prof. Marco M Ferrario

Tesi di dottorato di:

Giovanni Veronesi

Matr. 061970

Anno Accademico 2012-2013

SUMMARY

List of abbreviations:

AUC = Area Under the ROC-Curve; CAMUNI = CArdiovascular Monitoring Unit in Northern Italy; CHD = Coronary Heart Disease; CVD = CardioVascular Disease; MONICA = MONItoring of trends and determinants in CArdiovascular disease; IDI = Integrated Discrimination Improvement; NRI = Net Reclassification Improvement

INTRODUCTION

Current European and American guidelines for primary prevention of major coronary and stroke events recommend the use of a multivariable risk prediction model to identify high risk subjects[1, 2]. Several risk scores are available in different US[3, 4] and European[5] populations of middle-aged adults to estimate the risk of first fatal and non-fatal cardiovascular event over a 10 year time interval from a generally restricted number of risk factors, such as age, gender, lipids, systolic blood pressure, smoking habit and diabetes.

During the 2000s the 10-year risk prediction equation for the Italian population was developed as part of the Progetto CUORE[6], a project pooling 17 population-based cohorts enrolled between mid-1980s and early-1990s in different geographical areas, including the Brianza. The CUORE model has been adopted in clinical practice for risk stratification and statin reimbursement, but it was recently replaced by the European SCORE charts[7], although the latter does not consider non-fatal events in the prediction.

In a recent commentary on the utility of risk scores for primary prevention of cardiovascular disease in clinical practice, Grover and colleagues identified three important challenges[8]. First, primary prevention need to be moved towards the concepts of "lifetime"[9] and "long-term" risks[10], motivated also by the increasing life expectancy in western Countries. To this extent, 10-year risk prediction models are inadequate to distinguish between those at both low short-term and long-term risks, and those at low short-term but at elevated long-term risk due to the presence of non-optimal risk factors levels[11-13]. In the Framingham Study population, an unfavorable risk factor profile led to an increased 30-year risk of first cardiovascular event, independently on the age at the risk factors assessment[12]. In a cross-sectional study conducted in a representative sample of the Italian population, about 80% of individuals classified at low 10-year risk had increased lifetime risk according to US definition (>=40%), potentially leading to a consistent number of un-prevented

events that might have been prevented if lifetime risk had been considered[13]. This group was largely composed of women and young subjects, suggesting that long-term prediction models for risk stratification may be even more beneficial in populations at low incidence of cardiovascular disease[14]. As we write this document, there are only two long-term risk equations, one developed from the Framingham population in the US[12], and the other from a database of clinical records in the UK[15]. The previous experience with short-term models suggests that the development of a specific risk score in a low-incidence populations should be preferred with respect to re-calibration of models derived in high-incidence countries[16].

The second challenge is the assessment of the clinical utility of any given score, in particular of new ones. Subjects' stratification in risk categories is often based on arbitrary cut-points of absolute risk[8] originally proposed from the US population but that may show no benefit in clinical practice when applied in a different context[17]. Moreover, these cut-off values are the same for men and women, although the underlying risk distribution is not the same. The evaluation of the clinical benefit of long-term prediction by means of some standard measure[18] has not been provided so far and is therefore required[9].

The third challenge is bounded to the concept of "improvement" in risk prediction. The discrimination ability as measured by the Area Under the ROC curve (AUC) of most models based on traditional risk factors is in the range of 70%-80%[8]. Many efforts are nowadays dedicated to the contribution of novel markers, in particular to improve subjects' stratification and clinical utility[19]. At this stage, promising biomarkers are recommended for secondary screening of subjects at intermediate risk, due also to the costs of assessment[20, 21], while non-laboratory risk factors assessed in clinical practice at lower costs may be especially beneficial at a population level. Family history of coronary heart disease (CHD) and low socio-economic status are well-established independent risk factors with the same level of evidence as high-sensitivity CRP or fibrinogen[1, 2]. However,

despite the strong evidence coming from association studies, their contribution to risk prediction beyond traditional risk factors has been examined to a lesser degree and with controversial findings[22-27] over a short-term time interval only.

The aim of this PhD project was to develop a long-term cardiovascular disease risk prediction model intended to be used for primary prevention in clinical practice in Italy and potentially in other low-incidence, Southern European populations with similar characteristics. The work, as well as this document, has been structured in three main parts, roughly corresponding to the three challenges above mentioned. In the first section, ***model development and validation***, we focused on deriving the reference model for 20-year risk prediction of first coronary event or ischemic stroke, fatal or non-fatal, in the Italian population. Extending the range of risk prediction over 20 years is not a straightforward operation. First, although several studies have shown that a single measurement of risk factor is predictive of future events after 30 plus years[12, 28], behavioral changes and risk factors modification may affect model discrimination. Second, although an external validation, on a "new" set of subjects, of any score is recommended before adopting it in clinical practice[29], it is rarely performed in long-term prediction equations as it requires high-quality follow-up data, with a consistent event definition over-time in a large number of subjects possibly enrolled in different study cohorts. Previous long-term models only provided internal validation[12]. Finally, some authors suggested the potential need of accounting for the competing risk of non-CVD death when long-term models are used for risk stratification[12, 30].

In the second section we evaluated the ***clinical utility*** of the reference model for risk stratification, according to several strategies with contrasting public health aims, namely to reduce the fraction of events potentially "missed" by any preventive action, or to reduce un-necessary treatment. Subjects' stratification based on predicted risk was compared to a stratification based on the number of risk factors. The decision curve analysis[18] based on the Net Benefit was also provided.

Finally, in the last section, we evaluated the ***improvement*** in long-term risk prediction when ***family history of coronary heart disease and education*** are added to the reference model. Family history remains to date the most accessible way of measuring disease heritability, and it reflects both the genetic trait and the environment shared among household members[31]. Level of education is a frequently adopted proxy of social status, because it is easily measured, it remains stable over time, and it reflects both intellectual and material resources, as well as early lifetime conditions[32].
We hypothesized that the addition of these two time-invariant conditions in middle-aged adults, might actually improve long-term risk prediction beyond traditional and behavioral risk factors.

MATERIALS AND METHODS

Study population

The Brianza population comprises residents in 73 municipalities in the area between Milan and the Swiss border, Northern Italy. The CAMUNI (CArdiovascular Monitoring Unit in Northern Italy) study includes four independent population surveys carried out between 1986 and 1994 as part of either the WHO-MONICA Project (3 surveys[33]) or the PAMELA study[34]. Participation rates were 70.1%, 67.2%, and 70.8% for the three MONICA surveys, respectively, and 64% for the PAMELA Study, with no differences between men and women. Both the baseline screening and the follow-up for all the surveys were approved by the ethical committee of the Monza Hospital.

Baseline risk factors assessment

Cardiovascular risk factors were collected at baseline according to the standardized procedures and quality standards of the WHO-MONICA Project[35]. Serum total cholesterol, HDL-cholesterol and blood glucose were determined using the enzymatic method on a fasting blood sample. Systolic blood pressure was assessed twice, at 5 minutes apart, using a standard mercury sphygmomanometer; the study variable for systolic blood pressure is the average of the two measurements. A standardized interview was administered to participants by trained interviewers.

Information on the use of anti-hypertensive treatment in the previous two weeks was dichotomized as yes/no; similarly, cigarette smoking habit was dichotomized as current versus past/never smokers. Diabetes mellitus was defined using self-reported diagnoses, information on insulin and oral hypoglycemic treatments and fasting blood glucose exceeding 126 mg/dl. The presence at baseline of a previous history of myocardial infarction, unstable angina pectoris, cardiac revascularization or stroke was defined based on self-reported information.

Definition of family history of CHD and socio-economic position

In the first two MONICA surveys and in the PAMELA study cohort, the first-degree family history of coronary heart disease ("Has one or more of your first degree relatives suffered from coronary heart disease? ", with possible answers: yes/no) was ascertained at baseline as part of the interview, with no reference to age limit. The last MONICA survey included an age limit at 50 in the definition. The number of years of schooling ("How many years have you spent at school or in full time study?") was also investigated. As year of schooling are subject to modifications across different birth cohorts, we derived a three-class study variable (high, intermediate and low education) by comparing the years of schooling of any given subject with the distribution within his gender-specific birth cohort. Sample tertiles were used as cut-points, as previously described[36].

Study endpoint and follow-up procedures

The study endpoint is the occurrence of first major coronary event (myocardial infarction, acute coronary syndrome and coronary revascularization) or first ischemic stroke or carotid endarterectomy, fatal and non-fatal. Vital status and death certificates were available for 99% of the subjects. Suspected out-of-hospital deaths were investigated through interview of relatives. Suspected hospitalized coronary (discharge code ICD-IX 410 or 411 and ICD-IX CM 36.0-9 for coronary revascularization) and stroke events (ICD-IX 430-432, 434, 436; ICD-IX CM 38.01-39.22 or 39.50-39.52 with at least one 430-438 as discharge code, for carotid endarterectomy) were

identified through deterministic and probabilistic record linkages with regional hospital discharge databases, obtaining a satisfactory performance in case finding, as reported[16, 33, 37]. All acute events were investigated and validated according to the MONICA diagnostic criteria[35]; the ischemic subtype for stroke was attributed after review of the available clinical information.

Statistical methods

Model development and validation

The *derivation set* for model development consisted in the 35-69 years old men and women, free of cardiovascular disease at enrollment, participants to the CAMUNI study. The reference 20-year risk prediction model consisted in two gender-specific Cox regression models with age, total cholesterol, HDL-cholesterol, systolic blood pressure, anti-hypertensive treatment, cigarette smoking and diabetes. These predictors are core risk factors included in the 10-year CUORE Project score[6, 16] as well as in other well-established 10-year risk equations[3, 4]. After a preliminary check on linearity, total- and HDL-cholesterol were included in the model as categorical variables in four standard classes[4]. The interaction between systolic blood pressure and anti-hypertensive treatment was not statistically significant (p-value 0.84 in men and 0.12 in women, respectively). There was no evidence of any cohort effect in the full model, in men (3 df test p-value=0.2) nor in women (p-value=0.5). Finally, no violations in the proportional hazard assumption were observed using a standard test for time-dependent variables.

Model calibration was assessed through the Grønnesby-Bogan goodness-of-fit test, which is the extension of the Hosmer-Lemeshow test to the survival setting[38]. The Area Under the ROC curve (AUC) defines a measure of model discrimination, as the probability that the risk score for an event is higher that the score in a subject who is a non-event:

$$AUC = P(Z_i > Z_j \,|\, D_i = 1,\ D_j = 0) \qquad [1]$$

As we are in a survival setting, we must acknowledge that *i*) the AUC must be defined within a certain follow-up time, i.e. AUC(t), as "events" and "non-events" must be defined within a certain follow-up time; and *ii)* censorship must be taken into account when estimating AUC(t), since because of censorship we might not able to see all the events within t. Therefore, we will estimate the AUC(t) according to the following formula[39]:

$$\hat{AUC}(t) = \frac{\hat{E}[(1-\hat{S}(t|Z_i)) * \hat{S}(t|Z_j) I(Z_i < Z_j)]}{\hat{E}[(1-\hat{S}(t|Z)) * \hat{S}(t|Z)} \quad [2]$$

where $\hat{S}(t|Z_i)$ is the fitted survival function for risk score Z_i, $I(Z_i < Z_j)$ is an indicator variable for $Z_i < Z_j$ and $\hat{E}$ is the estimated expected value. As the formula [2] is based on the fitted survival function, the AUC takes censorship into account. Similarly, model sensitivity and specificity in the top and bottom predicted risk quintiles were also computed taking censorship into account[39].

To assess the hypothesis of a loss in discrimination ability due to a longer prediction period, we estimated the 10-year predicted probability of event in our database, using the same set of risk factors but with shorter follow-up period, i.e. up to the end of 2002 for all the subjects (number of events: 234 in men, 79 in women). We then compared the estimated AUC(10) with AUC(20) by looking at their respective bootstrapped confidence intervals.

The *internal validation analysis* consisted in estimating over-optimism in discrimination through 1000 bootstrapped samples[29]; we then provide AUC-corrected values. For the *external validation* analysis, the validation set consisted in the 5307 (2418 men) subjects enrolled in the Latina (Rome) in the same time span as the Brianza cohorts (MATISS Study). The MATISS study[40] was also part of the CUORE Project and shared the same procedures for baseline risk assessment and follow-up procedures, including MONICA definition of acute events, as the derivation set. To assess the external validation, we evaluated the performance of the CAMUNI score in the validation set; the Framingham CVD risk score[3] was used for comparison. The absolute predicted risk from both

scores was re-calibrated to the 20-year risk observed in the validation set. We report the calibration slope[41] as a measure of calibration. The calibration slope is the beta-coefficient from a Cox model fitted in the validation set with the re-calibrated absolute risk as the only covariate; a value different from 1 is suggestive of a different strength in predictor effects. A calibration plot was also provided. The Area Under the ROC-curve (AUC), estimated as in [2], measured the discrimination ability for the CAMUNI and the Framingham risk scores in the validation set; the AUC was compared to the value estimated for the CAMUNI score in the derivation set, corrected for over-optimism. Finally, in a sensitivity analysis we considered the effect of the competing risk of non-CVD death on risk stratification based on our prediction model by considering model calibration with and without competing risks. A published SAS macro was used to estimate the 20-year absolute risk of first CVD event taking competing risk into account[42].

Clinical utility

To assess the clinical utility of the long-term model for risk stratification, we considered two different public health goals. One is to decrease the number of events occurring among those considered at "low-risk". If we assume that a subject classified at "high risk" will be targeted for prevention (either lifestyle intervention or treatment), any event occurring outside this category is "not-identified" or "missed" by the prevention strategy. The second strategy aims instead to reduce un-necessary treatment, by decreasing the number of non-events among those considered at "high-risk". Under the two scenarios, "high-risk "subjects are defined as those with predicted risk above a certain cut-off value. Clinical utility is defined in terms of *i)* fraction of "missed" events; *ii)* probability of event among those classified at high risk; and *iii)* false positive/true positive ratio, for several threshold values in the 20-year predicted risk. We also provide a decision curve analysis based on the net benefit:

$$\text{Net Benefit} = (\text{true positives} - w*\text{false positives})/n, \qquad [3]$$

where n is the sample size and the weight w represents the ratio between the harm of un-necessary treatment and the harm of missing a case at that given value of predicted risk[18].

Improvement in risk prediction

The analysis on the additional contribution of education and family history of CHD to long-term risk prediction was restricted to the first two MONICA-Brianza surveys and the PAMELA study, due to inclusion of an age limit at 50 in the definition of family history of CHD in the most recent MONICA survey. "Improvement" was defined in terms of association, change in discrimination and reclassification improvement over the reference model[19]. Change in discrimination was assessed as difference in the Area Under the ROC-Curve (Δ-AUC(20)) as well as Integrated Discrimination Improvement[43] (IDI). The Δ-AUC(20) is defined as the difference in AUC(20) for the new and the traditional model, both estimated is in [2] taking censorship into account. The IDI was defined as the net gain between the change in sensitivity and the change in (1-specificity) due to the "new" model with respect to the "old" or "reference" one:

$$IDI = (IS_{new} - IS_{old}) - (IP_{new} - IP_{old})$$

In the survival setting, IDI becomes IDI(t) and should be estimated taking censorship into account. Chambless et al[39] found that the difference between IS(t) and IP(t) for a given model can be interpreted as the proportion of variance explained by the model:

$$IS(t) - IS(t) = \frac{Var[S(t|Z)]}{S(t)*[1-S(t)]} = R^2(t)$$

This quantity can be estimated from the fitted survival term:

$$\hat{R}^2(t) = \frac{Var[\hat{S}(t|Z)]}{\hat{S}(t)*[1-\hat{S}(t)]} \qquad [4]$$

The estimator for IDI(t) becomes:

$$\hat{IDI}(t) = \hat{R}^2_{new}(t) - \hat{R}^2_{old}(t) \qquad [5]$$

where $\hat{R}^2_{new}(t)$ and $\hat{R}^2_{old}(t)$ are the proportion of variance explained by the new and the old model, respectively, both estimated as in [4].

Pencina et al. introduced the concept of improvement in reclassification ability due to a new model over the reference one[43]. If we assume that subjects can be classified in three categories, i.e. "low", "intermediate" and "high" risk based on their absolute risk predicted by the reference model, the new model might change risk stratification as follows:

		New Model		
		Low Risk	**Int Risk**	**High Risk**
Old model	**Low Risk**		**UP** (**Improvement if D=1**, **worsened if D=0**)	**UP** (**Improvement if D=1**, **worsened if D=0**)
	Int Risk	**DOWN** (**Improvement if D=0**, **worsened if D=1**)		**UP** (**Improvement if D=1**, **worsened if D=0**)
	High Risk	**DOWN** (**Improvement if D=0**, **worsened if D=1**)	**DOWN** (**Improvement if D=0**, **worsened if D=1**)	

where "UP" and "DOWN" mean a reclassification in a category at higher risk (upward) or lower risk (downward) then the original category, respectively; the grey cells identify no change in risk categories. Whether a reclassification determines an "improvement" in risk stratification, it depends on whether the subject is an event (D=1) or not (D=0); we can define the Net Reclassification Improvement among events and non-events as:

$$NRI_{events} = P(UP|D=1) - P(DOWN|D=1)$$

$$NRI_{non\ events} = P(DOWN|D=0) - P(UP|D=0)$$

A weighted sum will define the overall Net Reclassification Improvement[43]:

$$NRI_{overall} = w * NRI_{events} + (1-w) * NRI_{non\ events} \quad [6]$$

where the weight w may reflect a differential "cost-benefit" function for improvement in events and in non-events. In our application, w=0.5.

In the survival setting, a formula for NRI(t) which takes censorship into account has been proposed by Chambless et al[41] as an application of the Bayes' theorem:

$$NRI(t)_{events} = \left(\frac{P(D(t)=1|UP) * P(UP)}{P(D(t)=1)} - \frac{P(D(t)=1|DOWN) * P(DOWN)}{P(D(t)=1)} \right)$$

$$NRI(t)_{non\ events} = \left(\frac{P(D(t)=0|DOWN) * P(DOWN)}{P(D(t)=0)} - \frac{P(D(t)=0|UP) * P(UP)}{P(D(t)=0)} \right)$$

and for w=0.5 and $p = P(D(t)=1)$:

$$NRI(t)_{overall} = \frac{(P(D(t)=1|UP) - p) * P(UP) - (P(D(t)=1|DOWN) - p) * P(DOWN)}{p(1-p)} \qquad [7]$$

The quantities in [7] can be estimated as Kaplan-Meier survival estimates among those reclassified upward and downward, as well as in the overall sample, to take censorship into account.

In our analysis, the improvement in risk stratification due to the addition of family history of CHD and education over the traditional model was measured by a three-category Net Reclassification Improvement at 20 years (NRI(20)). To define the risk categories, we used the threshold values 10% and 20% in men, and 2% and 10% in women; these values were chosen based on the previous assessment of clinical utility. In a sensitivity analysis we considered different thresholds as NRI is sensible to the choice of the cut-off values[44], but the findings did not change substantially from those presented here. We provide also an estimate for the clinical NRI, defined as the NRI among those originally considered at intermediate risk by the reference model[39,43]. As there are no close forms for standard error estimators for Δ-AUC(20), IDI(20) and NRI(20) presented in [2], [5] and [7] respectively, we provided bootstrapped confidence intervals from 1,000 bootstrapped samples.

A new SAS package for risk prediction models

All the analyses were conducted using the SAS software, release 9.2. As there are no publicly available programs, the author developed a comprehensive SAS package [reSAS, Risk Estimation in Survival Analysis using SAS], with several macros to assess model calibration, discrimination, and internal validity, as well as to compare several models in terms of Δ-AUC(t), IDI(t) and NRI(t). All the relevant metrics were estimated taking censorship into account, as appropriate for the survival setting[38, 39]. Confidence intervals at a 95% nominal level were estimated from bootstrapping. The package and the macros are fully reported and described in the appendix.

RESULTS

The CAMUNI 20-year CVD risk score: development and validation

In the CAMUNI study (derivation set) n=5,426 (2,703 men) subjects were enrolled in the age range 35-69 years. N=205 subjects (3.8%; n=14 events) had at least one missing data; we considered data imputation (R *transcan* function[44]) and excluded only those with missing values in more than 4 covariates of interest (n=6 men and n=3 women). Finally n=120 men and n=45 women with a positive history of cardiovascular disease at baseline were also excluded, reducing the sample size to 2,574 men and 2,673 women. The validation set consisted in the 2,418 men and 2,889 women, aged 35-69 years and free of cardiovascular disease at baseline, enrolled in the MATISS study.

Main demographic characteristics and CVD risk factors for the derivation and the validation set, by gender, are shown in **Table 1**. Subjects in the validation set were about 1 year older on average, had a lower HDL-cholesterol and a higher systolic blood pressure than the derivation set, in both genders. The prevalence of smoking was 10% higher in men, and 11% lower in women.
In the derivation set, during a median follow-up time of 15 (interquartile range: 12-20), we observed 315 first CVD events in men (233 coronary events) and 123 in women (n=85 coronary

events). The Kaplan-Meier estimate for 20-year risk was 16.1% and 6.1% in men and women, respectively. In the validation set the median follow-up time was 17 years (interquartile range 13-20); the 20-year Kaplan-Meier risk was slightly lower than in the derivation set, in men (13.2%) and in women (5.6%).

The beta-coefficients for the CAMUNI 20-year CVD risk score in the derivation set are provided in **Table 2**. All the risk factors were statistically significant, except for anti-hypertensive treatment, though its point estimate reflected a 30% increase in hazard in both men and women; the variable was retained in the model for comparability with the short-term CUORE model. There were no significant differences in the set of beta estimates for the 20-year model as compared to those from the 10-year risk model for the risk factors in the model (data not shown). The model calibration in the derivation set was satisfactory, in men (Grønnesby-Bogan goodness-of-fit chi-square 6.7, p-value=0.67) and in women (chi-square 9.6, p-value=0.38); the calibration plot, comparing the average predicted risk among deciles of observed risk, is available as **Figure 1**.

In the derivation set, after the correction for over-optimisms we found no statistically significant difference in the overall discrimination ability between long- and short-term prediction models, in men (AUC(20)=0.736 vs. AUC(10)=0.731) and in women (AUC(20)=0.801 vs. AUC(10)=0.816; **Table 3)**. Only 5% of 20-year events in men occurred among subjects with a predicted risk below the 20^{th} percentile (bottom quintile); the corresponding figure in women is 2%. The relative risk of event for being above the 80^{th} percentile vs. below the 20^{th} percentile of 20-year risk was 9.5 (i.e. 35.1/3.7) in men and 22.4 (i.e. 20.2/0.9) in women. Finally, the value of the 80^{th} percentile for 20-year risk was more than twice as high than the similar percentile for 10-year risk in men (26.8 vs. 10.8) and more than three times as high in women (10.1 vs. 3.0). A similar consideration holds for the 20^{th} percentile of risk or the median value.

Main findings from the *external validation* analysis are reported in **Table 4**, for the CAMUNI risk score as compared to the Framingham risk score. The calibration slope for the CAMUNI score in the validation set did not significantly differ from 1 in men (1.07; 95% confidence interval 0.91-1.23) nor in women (1.00; 0.83-1.16). The Framingham risk score performed equally well in men (1.06; 0.90-1.22) but worse in women (1.32; 1.10-1.55). A lack of calibration in women in the validation set for the Framingham risk score is also visible in the calibration plot (**Figure 2)**, in particular when the observed 20-year risk is above 5%. In the derivation set, the over-optimism corrected AUC(20) for the CAMUNI model was 0.737 in men and 0.801 in women (**Table 4**); corresponding figures in the validation set were 0.732 (95% CI: 0.727-0.738) in men, and 0.801 (0.794-0.808) in women. The Framingham risk score performed less well in men (0.722; 0.717-0.727) and in women (0.705; 0.699-0.711).

Finally, we considered the potential impact of the competing risk of non-CVD death on risk stratification based on the prediction model[12, 30]. In the derivation set, the Kaplan-Meier estimate of 20-year risk of first cardiovascular event adjusted for competing risk[42] was 14.9 in men and 5.9 in women. The calibration for the 20-year predicted risk from the standard Cox model was satisfactory except for the last decile of predicted risk in men (data not shown). In addition, the analysis by age strata did not reveal any clear pattern of risk overestimation by the standard Cox model (**Table 5**). These two findings somehow reflects the work by Wolbers and colleagues, which reported a satisfactory calibration for the standard Cox model up to the age of 75 in a frail population[30]. Thus in our population of 35-69 years old the competing risk of non-CVD death is likely not to affect CVD risk stratification in a clinically meaningful way.

Clinical utility analysis

Table 6a and **Table 6b** describe strategies for the identification of high-risk subjects, based on predicted 20-year risk, in men and women respectively. A cut-off value of 10% twenty year risk in

men would result in a 9% of "missed" events (i.e. events among those with predicted risk below the cut-point), with a probability of event of 23% and one true positive for every 3.4 false positives (**Table 6a**). In the second scenario, by choosing the 20% twenty year risk threshold value, the fraction of missed events was 36%. Note that about 30% of events occurred for a predicted 20-year risk between 20% and 30%. Finally, using the number of risk factors to define high risk subjects would result in a higher fraction of missed events, with no changes in specificity or in the prevalence of subjects at high risk.

Among women, a cut-off value of 2% would result in a 5% of missed events, with a probability of event of 9% and a true positive for every 10.1 false positive women (**Table 6b**). In the second scenario, the probability of event among those with absolute risk greater than 10% was 20.4%, with a true positive for every 3.9 false positives. However, the fraction of missed events would be 32%; this number can be reduced by lowering the cut-off value to 8%. By considering at high risk those with 2 or more risk factor would result in a higher fraction of missed events, with no gain in specificity or in the probability of event in the group. **Figure 3** illustrates the decision curve analysis based on the Net Benefit, for men (left) and women (right). The figure suggests a greater net benefit for the predicted risk with respect to the number of risk factors over the whole range of values, thus generalizing the findings from **Table 6a** and **Table 6b**.

Improvement in risk prediction due to family history of CHD and education

The analysis of the additional contribution of family history of CHD and education was restricted to the 4,099 subjects enrolled in the first two MONICA-Brianza surveys or in the PAMELA study in the age range 35-69 years. 130 subjects with a positive history of CVD at baseline were excluded, as well as subjects with missing data on covariates of interest (n=11). The available sample size for the analysis was 1,941 men and 2,015 women. The prevalence of family history of CHD at baseline was 27% in men and 34% in women; 42% of men and 37% of women were in the low education

group. During a median follow-up time of 18 years (interquartile range: 12-20), we observed 254 first CVD events in men (188 coronary events) and 102 in women (68 coronary events). The Kaplan-Meier estimate for 20-year risk was 16.7% and 6.4% in men and women, respectively.

In men, education was associated with incidence of cardiovascular events (2 df p-value=0.049) when controlling for age; in particular, less educated men had a significant 40% risk excess when compared to more educated subjects (95% Confidence Interval: 1.01, 1.88; **Table 7**). After the adjustment for traditional risk factors and family history of CHD, the association remained statistically significant (p-value: 0.03). We observed a 40% risk excess for less educated women as well; the association however was not significant, and partially mediated by traditional risk factors. In men, the age-adjusted hazard ratio for family history of CHD was 1.55 (95% CI: 1.20; 2.02); further adjustment for traditional risk factors did not modify the estimate. No association was present among women.

The model calibration was satisfactory, in men (Grønnesby-Bogan goodness-of-fit chi-square below 20 for all the models, all p-values greater than 0.2) and in women (all chi-squares less than 5; see **Table 3**). The AUC(20) for the reference model was 0.7508 in men and 0.8358 in women. In men, the inclusion of either education or family history of CHD modesty increased model's discrimination (**Table 8**), while the improvement was more evident when both were added (Δ-AUC: 0.01; 95% CI 0.002-0.02; IDI: 0.01; 95% CI 0.001-0-024). Among women, the change in discrimination was about one-fifth the level for men for any model, and no metric was significantly different from zero.

Table 9 reports the reclassification metrics, in men and women, for the overall population and considering only those classified at intermediate risk from the reference model. In men, the addition of both education and family history of CHD led to an overall NRI of 5.8% (95% CI: 0.2%-15.2%). Moreover, about 30% of those at intermediate risk were reclassified; the NRI among cases was

12%, while the overall NRI was 20.1% (95%CI: 0.5%-44%). Among women, no significant change in reclassification was observed, in the overall population (NRI = -1.4%) nor considering only those at intermediate risk (NRI = 6.6%, not significant). Only 5% to 7% of women were reclassified either upward or downward by the different models.

Figure 4 illustrates the reclassification plot due to the addition of both family history and education to the reference model, in men (left) and women (right). The 20-year probability of event among the 134 men reclassified upward was 23% (26 CVDs), rising to 31% considering only those at intermediate risk according to the reference model (73 men, 19 CVDs); i.e. about 1 event every 3 subjects. The probability of event among those reclassified downward was 13%. In women, the probability of event among those reclassified upward (n=48, 4 CVDs) and those reclassified downward (n=54, 3 CVDs) were 9% and 12%, respectively.

DISCUSSION

We illustrate here the development of a 20-year prediction model of first major coronary or ischemic stroke event in a Northern Italian population of men and women aged 35 to 69 years at baseline. To our knowledge, this is the first long-term prediction model in a low-incidence, southern European population. Based on the findings from the external validation analysis, the risk score seems to be appropriate for long-term risk prediction in Italy and, more generally, in low-incidence populations. As in the Framingham study, in our population the long-term predicted risk was more than simply *n*-times the short-term risk prediction[12]. In addition in the age range 35 to 49 years, the long-term predicted risk in subjects with 1 or more non-optimal or elevated risk factors (defined as in Lloyd-Jones at al.[9] was 3-times the short-term risk in men, and 4-times in women (**Figure 5**). This conveys the importance of long-term prediction for early identification of young subjects and women at increased likelihood of event during their remaining lifespan.

Risk scores are an attempt to predict an individual outcome based on group average among subjects

sharing the same levels of risk factors. Two recent debates – the lack of concordance between different risk calculators[46] and the severe risk overestimation by the new risk score adopted by 2013 American College of Cardiology/American Heart Association CVD primary prevention guidelines[47-50] – highlighted the importance of calibrating the model on the risk of the underlying population. This finding is not completely new[16], and it justifies the need of developing specific scores for populations at different disease incidence. By far less attention has been paid so far on how thresholds of predicted risk for subjects' stratification are chosen, often arbitrarily[8], potentially limiting the clinical utility of risk prediction models[17]. According to 2013 US guidelines[47], 45% and 23% of Caucasian men and women are above the recommended threshold of predicted risk for statin prescription, respectively. However, there are no indications on sensitivity and specificity of such a stratification, nor on cost implications of potentially treating about 1 middle-aged man out of 2. In this research project, we considered two strategies for the identification of "high-risk" subjects with contrasting public health goals, either to decrease the fraction of missed events or to decrease un-necessary treatment. These can be implemented by choosing threshold values for the predicted risk driven by either sensitivity or by specificity, respectively. Despite the lowering costs of statin treatment with respect to the costs of one un-prevented event, the high sensitivity scenario was not cost-effective over a 10-year period[51]. These two scenarios might be combined to adopt a more complex risk stratification, as often present in clinical practice[1-2, 14]. For instance, if we consider at "low-risk" the 36% of men with 20-year absolute risk less than 10%, the fraction of missed events would be 9%, i.e. 31 first events in 20 years. About 31% of men with absolute risk between 10% and 20% could be addressed for lifestyle modification or treatment according to the presence of specific risk factors; this category accounts for about 20% of cases. Finally, the 33% of men with predicted risk above the 20% could be targeted with treatment intervention; they account for 68% of events, and out of 3.2 treated men, one is a case. A similar stratification can be provided for women, with different threshold values reflecting gender-specific underlying risk as for the cardiovascular age assessment[8].

The analysis of the additional contribution of family history of CHD and education to a reference model with established CVD risk factors gives the opportunity here to discuss the concept of "improvement" in risk prediction from the statistical perspective. Although family history and education are well-established independent risk factors for major cardiovascular events[1,2], association alone is not enough to warrant the inclusion in any risk score[19]. Discrimination statistics such as the Δ-AUC and the IDI define "improvement" in terms of an increased ability of separating events from non-events. These measures however get smaller and smaller as the discrimination ability of the reference model increases, no matter how strong the additional predictor is; furthermore, a Δ-AUC=0.02 has no straightforward clinical interpretation[52]. The NRIs statistics define "improvement" in terms of a better stratification of subjects in risk categories, which ultimately leads to a more appropriate clinical decision on treatment allocation. Thus the NRI assess the clinical value of the additional information due to the new markers, especially when its separate components are also provided (NRI among events and among non-events; see [6] above). However, Pepe and colleagues pointed out that the null hypothesis of no association is equivalent to the null hypothesis of no "improvement", no matter how defined[53]. As in the logistic regression setting the distribution of "improvement" metrics under the null may not be normal[53, 54], the null hypothesis of no "improvement" should be tested through a standard likelihood test comparing two nested models[53].

In our perspective cohort study with a long follow-up period in which subjects are exposed to censorship, we estimated a comprehensive set of discrimination and reclassification metrics, as appropriate in the survival setting[39]. These estimators had less bias, smaller variance and mean squared error than the original ones which ignore censorship[39]. As there are no close forms for standard errors, we provided confidence intervals based on bootstrap, which may be slightly more conservative than the nominal 95% level[39]. To investigate the asymptotic properties of Δ-AUC(t), IDI(t) and NRI(t) could be the topic of future research. Our findings were quite consistent from Pepe and colleagues perspective. In men, after the adjustment for established risk factors, low

education and positive family history of CHD were associated with the study endpoint (table 7). The addition of both factors to the reference model significantly improved discrimination (table 8) and risk stratification (table 9), as bootstrapped confidence intervals for these quantities did not contain 0. Considering the subgroup at intermediate risk according to the reference model, the NRI among cases was 12%, the overall NRI was 20.1% (95%CI: 0.5%-44%), and about 1 every 3 men reclassified upward is expected to experience a CVD event in 20 years. In women, to a null finding in the association for both education and family history of CHD (table 7) corresponded a modest and not-significant change in discrimination and in risk stratification. The age-adjusted hazard ratios for low education were similar in men and women (1.38 vs. 1.40, respectively; table 7), but the lower number of events as well as the presence of wider social inequalities in risk factors distribution with respect to men[55] may explain the non-statistically significant result in women. For family history of CHD, we acknowledge the absence of the age limit in our definition. In the last MONICA-Brianza survey (not included here), where family history was defined within the age limit of 50 years, the age-adjusted hazard ratio in women raised to 1.59, as compared to 0.96 in Table 7. The difference in hazard ratios was less evident in men (1.64 vs. 1.55).

We briefly discuss strengths and limitations of the current analysis. Our sample comprises subjects drawn from a representative northern Italian population, with a satisfactory participation rate. The underlying population is characterized by high levels of industrialization and urbanization, with one of the highest average incomes in Italy. We also mention a high-quality of follow-up procedures, including case ascertainment for non-fatal events[37] and a consistent event validation according to MONICA criteria. Also, the Standardized Incidence Rate, a measure comparing the expected and observed number of events in the cohort using rates from the underlying population, was above 1 over the whole follow-up period[33]. Together with measures of internal validity of the predictive model, we provide a formal external validation analysis, which has been an issue for previous long-term models[12, 15]. Finally, the study endpoint reflects the clinical need to treat the “global” ischemic

risk of a given patient, and not its separate components[47]. Potential study limitations include the definition of positive family history of CHD based on self-reported data without a formal validation. The self-reported definition is likely to be used in clinical practice[15, 31] and it has been adopted by other observational prospective studies, either with[22, 26] or without age limits[23]. The lack of age limit in our definition may have resulted in a lower sensitivity for positive family history potentially biasing the association with the study endpoint toward the null, as mentioned above. In more recent data from the same Brianza area[56], the prevalence of self-reported family history of CVD was 28% in men (age limit 55) and 35% in women (age limit 65). The comparison with the prevalence reported in our population (27% and 34% in men and women, respectively) suggests a non-differential misclassification by gender in our data.

CONCLUSION

During the PhD program the author developed the first model to predict long-term risk of first major ischemic cardiovascular event in a low-incidence, Southern European population. The prediction model has been internally and externally validated, and its clinical utility has been formally assessed at different thresholds of predicted risk for clinical decisions. The clinical utility analysis should be part of the validity assessment of any new predictive model. The statistical implications of assessing the "improvement" in risk prediction were discussed and illustrated through a paradigmatic analysis of two indicators of disease heritability and social status. A new SAS package, Risk Estimation in Survival Analysis using SAS [reSAS], detailed in the appendix, has been specifically developed by the author for the SAS software release 9.2, and is available to other researchers.

REFERENCES

1. Perk J, De Backer G, Gohlke H, Graham I, Reiner Z, Verschuren WM, et al. European Guidelines on cardiovascular disease prevention in clinical practice (version 2012). *Eur Heart J 2012*;33:1635-1701

2. Greenland P, Alpert JS, Beller GA, Benjamin EJ, Budoff MJ, Fayad ZA, Foster E, Hlatky MA, Hodgson JMcB, et al. 2010 ACCF/AHA guideline for assessment of cardiovascular risk in asymptomatic adults: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol* 2010;56:e50–103.

3. D'Agostino RB, Ramachandran SV, Pencina MJ, Wolf PA, Cobain M, Massaro JM, Kannel WB. General cardiovascular risk profile for use in primary care. The Framingham Heart Study. *Circulation* 2008;117:743-753

4. Chambless LE, Folsom AR, Sharrett AR, Sorlie P, Couper D, Szklo M, Nieto FJ. Coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC) study. *J Clin Epidemiol*. 2003;56:880-90

5. Conroy RM, Pyorala 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

6. Palmieri L, Panico S, Vanuzzo D, Ferrario M, Pilotto L, Sega R, Cesana G, Giampaoli S per il gruppo di ricercar del Progetto CUORE. Evaluation of the global cardiovascular absolute risk: the Progetto CUORE individual score. *Ann Ist Super Sanità* 2004;40:393-399.

7. Agenzia Italiana del Farmaco. Modifica alla Nota AIFA n.13 09/04/2013. Gazetta Ufficiale n. 83, 9 Aprile 2013

8. Grover SA, Lowensteyn I. The challenges and benefits of cardiovascular risk assessment in clinical practice. *Can J Cardiol* 2011;27:481-487

9. Lloyd-Jones DM, Leip EP, Larson MG, D'Agostino RB, Beiser A, Wilson PWF, Wolf P, Levy D. Prediction of lifetime risk for cardiovascular disease by risk factor burden at 50 years of age. *Circulation* 2006;113:791-798

10. Lloyd-Jones DM. Short-term versus long-term risk for coronary artery disease: implications for lipid guidelines. *Curr Opin Lipidol* 2006;17:619-625

11. Daviglus ML, Stamler J, Pirzada A, Yan LL, Garside DB, Liu K, Wang R, Dyer AR, Lloyd-Jones DM, Greenland P. Favourable cardiovascular risk profile in young women and long-term risk of cardiovascular and all-cause mortality. *JAMA* 2004;292:1588-1592

12. Pencina MJ, D'Agostino RB, Larson MG, Massaro JM, Vasan RS. Predicting the 30-year risk of cardiovascular disease. The Framingham Heart Study. *Circulation* 2009;119:3078-3084

13. Di Castelnuovo A, Costanzo S, Persichillo M, Olivieri M, de Curtis A, Zito F, Donati MB, de Gaetano G, Iacoviello L. Distribution of short and lifetime risks for cardiovascular disease in Italians. *Eur J Prev Cardiol* 2011;19:723-730.

14. Mosca L, Benjamin EJ, Berra K, Bezanson JL, Dolor RJ, et al. Effectiveness-based guidelines for the prevention of cardiovascular disease in women–2011 update: a guideline from the American Heart Association. *J Am Coll Cardiol* 2011;57:1404–1423

15. Cox JH, Coupland C, Robson J, Brindle P. Derivation, validation and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular diseases: cohort study using QResearch database. *BMJ* 2010;341:c6624.

16. Ferrario MM, Chiodini P, Chambless LE, Cesana G, Vanuzzo D, Panico S, Sega R, Pilotto L, Palmieri L and Giampaoli S for the CUORE Project Research Group. Prediction of coronary events in a low incidence population. Assessing accuracy of the CUORE Cohort Study prediction equation. *Int J Epidemiol* 2005;34:413-21

17. Collins GS, Altman DG. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. *BMJ* 2012;344:e4181.

18. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *Med Decis Making* 2008;8:53

19. Hlatky MA, Greenland P, Arnett DK, et al. on behalf of the American Heart Association Expert Panel on Subclinical Atherosclerotic Diseases and Emerging Risk Factors and the Stroke Council. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation* 2009;119:2408 –2416.

20. The Emerging risk factors collaboration. C-Reactive Protein, Fibrinogen, and cardiovascular disease prediction. *N Engl J Med* 2012;367:1310-20.

21. Wierzbicki AS. New directions in cardiovascular risk assessment: the role of secondary risk stratification markers. *Int J Clin Pract* 2012;66:622-630

22. Sivapalaratnam S, Boekholdt SM, Trip MD, Sandhu MS, LUben R, Kastelein JP, Wareham NJ, Khaw K. Family history of premature coronary heart disease and risk prediction in the EPIC-Norfolk prospective population study. *Heart* 2010;96:1985-1989

23. Yeboah J, McClelland RL, Polonsky TS, Burke GL, Sibley CT, O'Leary D, Carr JJ, Goff DC Jr, Greenland P, Herrington DM. Comparison of novel risk markers for improvement in cardiovascular risk assessment on intermediate-risk individuals. *JAMA* 2012;308:788-795

24. Ramsay SE, Morris RW, Whincup PH, Papacosta AO, Thomas MC and Wannamethee SG. Prediction of coronary heart disease risk by Framingham and SCORE risk assessments varies by socioeconomic position: results from a study in British men. *Eur J Cardiovasc Prev Rehabil* 2011;18:186-193.

25. Franks P, Tancredi DJ, Winters P et al. Including socioeconomic status in coronary heart disease risk estimation. *Ann Fam Med* 2010;8:447-53

26. Woodward M, Brindle P, Tunstall-Pedoe H, for the SIGN group on risk estimation. Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). *Heart* 2007;93:172-176.

27. Fiscella K, Tancredi D, Franks P. Adding socioeconomic status to Framingham scoring to reduce disparities in coronary risk assessment. *Am Heart J* 2009;157:988-994.

28. Menotti A, Lanti M, Kafatos A, Nissinen A, Dontas A, Nedeljkovic S, Kromhout D, for the Seven Countries Study. The role of baseline casual blood pressure measurement and of blood pressure changes in middle age in prediction of cardiovascular and all-cause mortality occurring late in life: a cross-cultural comparison among the European cohorts of the Seven Countries Study. *J Hypertens* 2004;22:1683-1690

29. Harrel FE, Lee KL and Marck DB. Tutorial in biostatistics: multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statist med* 1996;15:361–387

30. Wolbers M, Koller MT, Witteman JCM, Steyerberg EW. Prognostic models with competing risks. Methods and application to coronary risk prediction. *Epidemiology* 2009;20:555-561

31. Banerjee A. A review of family history of cardiovascular disease: risk factor and research tool. *Int J Clin Pract* 2012;66:536-543

32. Galobardes B, Shaw M, Lawlor DA, et al. Indicators of socioeconomic position (part 1). *J Epidemiol Community Health* 2006;60(1):7-12.

33. Ferrario M, Sega R, Chatenoud L, et al. Time trends of major coronary risk factors in a northern Italian population (1986-1994). How remarkable are socio-economic differences in an industrialised low CHD incidence country? *Int J Epidemiol* 2001;30:285-291.

34. Cesana GC, De Vito G, Ferrario M, et al. Ambulatory blood pressure normalcy: The PAMELA Study. *J Hypertension Suppl* 1991;9:17-23.

35. WWW-publications from the WHO MONICA Project. MONICA Manual. Available at: http://www.thl.fi/publications/monica/manual/index.htm.

36. Karvanen J, Veronesi G, Kuulasmaa K. Defining thirds of schooling years in population studies. *Eur J Epidemiol* 2007;22:487-492.

37. Fornari C, Madotto F, Demaria M, et al. Record-linkage procedures in epidemiology: an Italian multicentre study. *Epidemiol Prev* 2008;32:79-88.

38. May S., Hosmer DW. A simplified method of calculating an overall Goodness-of-Fit test for the Cox proportional hazards model. *Lifetime Data Anal* 1998;4:109-120

39. Chambless LE, Cummiskey CP, and Cui G. Several methods to assess improvement in risk prediction models: extension to survival analysis. *Statist med* 2011;30:22–28

40. Giampaoli S, Poce A, Sciarra F, et al. Change in cardiovascular risk factors during a 10-year community intervention program. *Acta Cardiologica* 1997;5:372-379.

41. Steyerberg EW. Clinical prediction models. 2009 Springer Science + Business Media, LLC. New York, US.

42. Rosthøj S, Andersen PK, Abildstrøm SZ. SAS macros for estimation of the cumulative incidence functions based on a Cox regression model for competing risks survival data. *Comput Meth Programs Biomed* 2004;74:69-75

43. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr., Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157-72.

44. Cook NR. Clinical relevant measures of fit? A note of caution. *Am J Epidemiol* 2012;176(6):488-491

45. Harrel FE. R reference manual: Package 'Hmisc'. Available at http://cran.r-project.org/web/packages/Hmisc/Hmisc.pdf

46. Allan GM, Nouri F, Korownyk C, Kolber MR, Vandermeer B, McCormack J. Agreement among cardiovascular disease risk calculators. *Circulation* 2013;127:1948-1956

47. Goff Jr DC, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk, Journal of the American College of Cardiology (2013), doi: 10.1016/j.jacc.2013.11.005

48. Ridker PM, Cook NR. Statins: new American guidelines for prevention of cardiovascular disease. *Lancet* 2013;382:1762-1765

49. Kovalchik S. New cardiovascular disease risk calculator is too risky. Avalable at: http://www.significancemagazine.org/details/webexclusive/5592881/New-cardiovascular-disease-risk-calculator-is-too-risky.html [accessed january 2014]

50. Kolata G. Risk calculator for cholesterol appears flawed. Published on nov, 17 2013 in the New York Times. Available at: http://www.nytimes.com/2013/11/18/health/risk-calculator-for-cholesterol-appears-flawed.html?_r=1& [accessed January 2014]

51. Greving JP, Visseren FLJ, de Wit GA, et al. Statin treatment for primary prevention of vascular disease: whom to treat? Cost-effectiveness analysis. *BMJ* 2011;342:d1672.

52. Pencina MJ, D'Agostino RB, Pencina KM, Janssens CJW, Greenland P. Interpreting incremental value of markers added to risk prediction models. Am J Epidemiol 2012;176:473-81.

53. Pepe MS, Kerr KF, Longton G, Wang Z. Testing for improvement in prediction model performance. *Statist Med* 2013:32:1467-82

54. Kerr KF, McClelland RL, Brown ER, Lumley T. Evaluating the incremental value of new biomarkers with Integrated Discrimination Improvement. *Am J Epidemiol* 2011:174:364-74

55. Veronesi G, Ferrario MM, Chambless LE, Sega R, Mancia G, Corrao G, Fornari C, Cesana G. Gender differences in the association between education and the incidence of cardiovascular events in Northern Italy. *Eur J Public Health* 2011;21(6):762-7

56. Istituto Superiore di Sanità. Cardiovascular Epidemiology Observatory, risk factors distribution: family history of cardiovascular disease. Available at: http://www.cuore.iss.it/eng/factors/family.asp

TABLES AND FIGURES

Table 1. Baseline characteristics (mean (SD) or %) of the study population and number of incident events, by gender. Men and women, 35-69 years old, CVD-free at baseline. Derivation set (MONICA-Brianza and PAMELA Study) and validation set (MATISS Study).

	Men			Women		
Variable	**Derivation set**	**Validation set**	**p**	**Derivation set**	**Validation set**	**p**
N	2574	2418	-	2673	2889	-
Age (years)	50.8 (9.1)	51.9 (9.4)	***	50.3 (9)	51.4 (9.4)	***
Total Cholesterol (mg/dl)	223 (42.5)	221.9 (39)	ns	222.9 (43.5)	220.5 (38.5)	*
HDL-Cholesterol (mg/dl)	50.6 (13.2)	49.1 (12.1)	***	61.5 (14.8)	54.4 (11.9)	***
Body Mass Index	26.2 (3.5)	27.2 (3.6)	***	25.6 (4.7)	29.2 (4.9)	***
Systolic Blood Pressure (mmHg)	134.8 (19.3)	138.3 (18.7)	***	131.6 (20.2)	138.8 (20.6)	***
Anti-hypertensive treatment (%)	11.8	6.6	***	16.0	15.3	ns
Glycaemia (mg/dl)	97.9 (23.8)	95.2 (22.7)	***	91.3 (21.6)	91.6 (22.8)	ns
Diabetes (%)	6.7	5.3	*	4.0	5.3	*
Current smoker (%)	37.1	47.9	***	19.6	8.1	***
Incident coronary event (n)	233	187	-	85	68	
Incident ischemic stroke (n)	99	59	-	43	53	
Incident CVD event (n)	315	238	-	123	119	
CVD Event Rate°	8.5	6.5	-	2.9	2.5	
20-year absolute risk of CVD^	16.1	13.2	-	6.1	5.6	

°: per 1000 person-years. ^: Kaplan-Meier Estimate. p-value testing the difference in risk factor distribution between the two sets of data; ***:<.0001; **:<.01; *:<.05. ns = not significant. ^: Kaplan-Meier estimate.

Table 2: Beta-coefficients, standard errors and baseline survival for the CAMUNI 20-year risk prediction model in the derivation set (MONICA-Brianza and PAMELA Study). Men and women, 35 to 69 years old, free of CVD at baseline.

	Men			Women		
	Beta	**SE**	**p-value**	**Beta**	**SE**	**p-value**
Age (years)	0.058	0.008	<.0001	0.084	0.014	<.0001
Total Cholesterol^						
200-240 mg/dl	0.388	0.161		0.553	0.287	
240-280 mg/dl	0.690	0.167	<.0001	0.607	0.310	0.027
> 280 mg/dl	0.923	0.198		0.996	0.328	
HDL-Cholesterol°						
<45 mg/dl	0.403	0.160		0.804	0.250	
45-50 mg/dl	0.367	0.186	0.013	0.364	0.309	0.015
50-60 mg/dl	0.024	0.177		0.261	0.225	
Systolic Blood Pressure (mmHg)	0.011	0.003	0.0003	0.015	0.005	0.001
Anti-hypertensive treatment (yes/no)	0.247	0.154	0.11	0.267	0.209	0.20
Smoking (yes/no)	0.521	0.117	<.0001	0.994	0.216	<.0001
Diabetes (yes/no)	0.744	0.163	<.0001	1.020	0.249	<.0001

SE = Standard Error. ^: reference group: total cholesterol<=200 mg/dl. °: reference group: HDL-cholesterol >60 mg/dl. *: at the mean value for continuous RFs, and at the reference class for categorical variables.
The risk model should be used within the following range for continuous risk factors: total cholesterol 135-330 mg/dl; HDL-cholesterol 30-100 mg/dl; systolic blood pressure 100-190 mmHg.

Table 3. Discrimination ability in the derivation set (MONICA-Brianza and PAMELA Study) for the 10-year and the 20-year risk prediction models. Men and women, 35-69 years old, CVD-free at baseline.

	Men		Women	
	10-year risk	**20-year risk**	**10-year risk**	**20-year risk**
AUC (95% CI)	0.731 (0.702; 0.761)	0.737 (0.713; 0.764)	0.814 (0.779; 0.853)	0.801 (0.771; 0.833)
Subjects with predicted risk below the 20th percentile				
20th percentile of risk	2.3	6.3	0.3	1.1
Fraction of events* (%)	4.4	5.1	1.4	2.0
Probability of event in the group^ (%)	0.8	3.7	0.2	0.9
Subjects with predicted risk above the 80th percentile				
80th percentile of risk	10.8	26.8	3.0	10.1
Sensitivity* (%)	49.9	45.6	68.7	62.0
Specificity (%)	82.4	85.5	81.1	83.1
Probability of event in the group^ (%)	19.4	35.1	7.5	20.2

The Area Under the ROC-curve (AUC) was estimated taking censorship into account, and adjusting for over-optimism (n=1000 bootstrap).
*: Probability of belonging to the group, given that the subject is a case. ^: Kaplan-Meier estimate of the probability of event in the group.

Table 4. External validation analysis for the CAMUNI score: calibration slope in the validation set, and discrimination ability in the derivation and in the validation sets. Men and women, 35-69 years old, CVD-free at baseline.

	Men	**Women**
Calibration slope (95% CI)		
Validation set°		
CAMUNI Risk Score	1.07 (0.91; 1.23)	1.00 (0.83; 1.16)
Framingham Risk Score	1.06 (0.90; 1.22)	1.32 (1.10; 1.55)
Discrimination [AUC (95%CI)]		
Derivation set^	0.737 (0.713; 0.764)	0.801 (0.771; 0.833)
Validation set°		
CAMUNI Risk Score	0.732 (0.727; 0.738)	0.801 (0.794; 0.808)
Framingham Risk Score	0.722 (0.717; 0.727)	0.705 (0.699; 0.711)

AUC = Area under the ROC Curve. ^: corrected for over-optimism. °: the CAMUNI score and the Framingham risk score were re-calibrated to the observed 20-year risk in the validation set.

Table 5: Comparison between observed and predicted 20-year risk of CVD in the derivation set taking into account the competing risk of non-CVD death, according to different age groups at baseline. Men (left) and women(right), 35-69 years old at baseline, free from CVD at baseline

	Men			**Women**		
Age	Observed 20-year risk°	Predicted 20-year risk, with no competing risk[1]	Predicted 20-year risk, with competing risk[2]	Observed 20-year risk°	Predicted 20-year risk, with no competing risk[1]	Predicted 20-year risk, with competing risk[2]
35-44	8.1	7.4	7.2	1.9	1.4	1.4
45-54	11.2	15.1	14.0	3.5	4.3	4.2
55-69	24.5	28.8	23.4	12.1	12.7	14.3

°: Kaplain-Meier estimate of 20-year risk, adjusted for competing risk of non CVD death
Predicted 20-year risk: average predicted 20-year risk 1: ignoring competing risk of non-CVD death; 2 taking the competing risk of non-CVD death into account

Table 6a. Identification of high risk men based on the 20-year risk prediction model with respect to the number of risk factors, according to strategies aiming to *i)* reducing the fraction of missed events; and *ii)* reducing un-necessary treatment. Men, 35-69 years old, CVD-free at baseline; derivation set (MONICA-Brianza and PAMELA Study).

	Men at high risk		Fraction of missed events (%)	Specificity (%)	Probability of event* (%)	FP/TP Ratio
	n	%				
***Strategy a*: reduce the fraction of missed events**						
All	2574	100.0	0.0	-	16.1	5.2
1+ Major Risk Factor[#]	1842	71.6	13.7	32.5	19.5	4.1
20-year absolute risk > 10%	1645	63.9	9.1	41.2	22.9	3.4
20-year absolute risk > 15%	1169	45.4	22.1	60.9	27.7	2.6
***Strategy b*: reduce un-necessary treatment**						
2+ Major Risk Factors[#]	828	32.2	50.4	73.6	24.9	3.0
20-year absolute risk > 20%	841	32.7	35.7	73.7	31.7	2.2
20-year absolute risk > 30%	415	16.1	62.6	88.9	37.4	1.7

"Missed" events are events occurring among men not classified at "high risk", i.e. with 20-year absolute risk (or a number of risk factors) below the cut-off point.
*: Kaplan-Meier estimate of the probability of event in the group (positive predicted value).
FP = Number of False Positives; TP = Number of True Positives
#: total cholesterol>240 mg/dl; HDL-cholesterol <40 mg/dl; systolic blood pressure >160 mmHg; smoking; diabetes

Table 6b. Identification of high risk women based on the 20-year risk prediction model with respect to the number of risk factors, according to strategies aiming to *i)* reducing the fraction of missed events; and *ii)* reducing un-necessary treatment.
Women, 35-69 years old, CVD-free at baseline; derivation set (MONICA-Brianza and PAMELA Study).

	Women at high risk		Fraction of missed events (%)	Specificity (%)	Probability of event* (%)	FP/TP Ratio
	n	%				
***Strategy a*: reduce the fraction of missed events**						
All	2673	100.0	0.0	-	6.1	15.3
1+ Major Risk Factor[#]	1654	61.9	17.7	40.1	8.2	11.3
20-year absolute risk > 2%	1733	64.8	4.5	37.4	9.0	10.1
20-year absolute risk > 5%	1067	39.9	14.7	63.2	13.1	6.6
***Strategy b*: reduce un-necessary treatment**						
2+ Major Risk Factors[#]	640	23.9	42.3	79.5	14.8	5.8
20-year absolute risk > 8%	698	26.1	22.7	77.1	18.2	4.5
20-year absolute risk > 10%	545	20.4	32.1	82.7	20.4	3.9

"Missed" events are events occurring among women not classified at "high risk", i.e. with 20-year absolute risk (or a number of risk factors) below the cut-off point.
*: Kaplan-Meier estimate of the probability of event in the group (positive predicted value).
FP = Number of False Positives; TP = Number of True Positives
#: total cholesterol>240 mg/dl; HDL-cholesterol <50 mg/dl; systolic blood pressure >160 mmHg; smoking; diabetes

Table 7: Association between education and family history of CHD with the onset of first major coronary event or ischemic stroke during follow-up in the Brianza population. Men (left) and women (right), 35-69 years old at baseline, free from CVD at baseline

	Men			**Women**		
	Age-adjusted	**Traditional RFs-adjusted**	**Full model**	**Age-adjusted**	**Traditional RFs-adjusted**	**Full model**
Education						
High Education	**Ref***	Ref	**Ref***	Ref	Ref	Ref
Intermediate Education	**1.00 (0.69; 1.46)**	0.90 (0.62; 1.32)	**0.93 (0.63; 1.35)**	1.17 (0.72; 1.90)	1.18 (0.71; 1.94)	1.18 (0.72; 1.95)
Low Education	**1.38 (1.01; 1.88)**	1.29 (0.94; 1.78)	**1.35 (0.98; 1.85)**	1.40 (0.83; 2.36)	1.26 (0.73; 2.15)	1.24 (0.73; 2.12)
Family history of CHD	**1.55 (1.20; 2.02)†**	**1.52 (1.16; 1.98)†**	**1.55 (1.19; 2.03)†**	0.96 (0.63; 1.45)	0.83 (0.54; 1.27)	0.83 (0.55; 1.27)

In the table: Hazard Ratios (95% Confidence Intervals) from Cox Proportional Hazards model
Traditional RFs-Adjusted Hazard Ratio: age, total cholesterol, HDL-cholesterol, systolic blood pressure, anti-hypertensive treatment, smoking, diabetes
Full model: model with traditional RFs (as above) plus education and family history of CHD
p-value testing the association between education (2df chi-square test) and family history of CHD (1df chi-square test) with first coronary event or ischemic stroke during follow-up: *=<0.05; †:=<0.001

Table 8: Model calibration and improvement in discrimination due to the addition of education, family history of CHD, or both to the reference model in the Brianza population. Men and women, 35-69 years old at baseline, free from CVD at baseline

	Model Calibration°	Change in discrimination	
		Δ-AUC (95%CI)	IDI (95%CI)
Men			
Reference model	7.7	Ref^	Ref
Reference + education	6.6	**0.004 (0; 0.013)**	0.003 (-0.001; 0.012)
Reference + family history of CHD	6.7	**0.005 (0; 0.015)**	**0.007 (0; 0.022)**
Reference + education & Family history of CHD	12.2	**0.010 (0.002; 0.02)**	**0.010 (0.001; 0.024)**
Women			
Reference model	4.8	Ref*	Ref
Reference + education	2.8	0.001 (-0.001; 0.006)	0.001 (-0.003; 0.009)
Reference + family history of CHD	2.2	0.001 (0; 0.01)	0.000 (-0.002; 0.007)
Reference + education & Family history of CHD	3.1	0.002 (-0.001; 0.008)	0.002 (-0.005; 0.007)

The reference model includes: age, total cholesterol, HDL-cholesterol, systolic blood pressure, anti-hypertensive treatment, smoking, diabetes
°: We report the chi-square values for the Gronnesby-Borgan goodness-of-fit test. Values above 20 suggest a lack of calibration
AUC: Area Under the ROC Curve (difference from the reference model value). IDI = Integrated Discrimination Improvement (%)
AUC for the reference model: ^ = 0.7508 (men) and * = 0.8358 (women)

Table 9: Probability of reclassification and Net Reclassification Improvement metrics over the reference model due to the addition of education, family history of CHD, or both, in the Brianza population. Men (left) and women (right), 35-69 years old at baseline, free from CVD at baseline.

	Men			**Women**		
	Reference model +…			**Reference model +…**		
	Education	**Family History**	**Education & Family history**	**Education**	**Family History**	**Education & Family history**
All subjects						
Reclassified upward (%)	4.8	6.2	6.9	2.1	1.6	2.4
Reclassified downward (%)	6.5	6.4	9.4	2.1	2.2	2.7
NRI, events (%)	0.9	2.4	2.3	-4.5	-2.2	-1.6
NRI, non-events (%)	2.2	0.7	3.5	-0.3	0.5	0.2
NRI, overall (%; 95% CI)	3.1 (-1.4; 12.1)	3.1 (-3; 11.3)	**5.8 (0.2; 15.2)**	-4.9 (-15.2; -2.1)	-1.7 (-13.2; 0.5)	-1.4 (-12.2; 3.7)
Subjects at intermediate risk*						
Reclassified upward (%)	8.6	11.4	12.1	3.2	2.5	3.3
Reclassified downward (%)	10.3	9.9	17.4	2.5	3.1	3.6
NRI, events (%)	3.8	14.9	11.8	-0.8	1.6	6.1
NRI, non-events (%)	2.6	0.9	9.3	-0.7	0.7	0.6
NRI, overall (%; 95% CI)	6.4 (-10.4; 22.3)	15.7 (-1.7; 38)	**20.1 (0.5; 44)**	-1.4 (-36.2; 3.3)	2.4 (-10.2; 32.2)	6.6 (-13.9; 32.3)

The reference model includes: age, total cholesterol, HDL-cholesterol, systolic blood pressure, anti-hypertensive treatment, smoking, diabetes
*: Intermediate risk defined as 20-year predicted risk from the reference model between 10% and 20% in men; and between 2% and 10% in women. NRI = Net Reclassification Improvement

Figure 1: Calibration plot for the CAMUNI 20-year CVD risk prediction model in the derivation set (MONICA-Brianza and PAMELA Study). Men (left) and women (right), 35 to 69 years old, free of CVD at baseline.

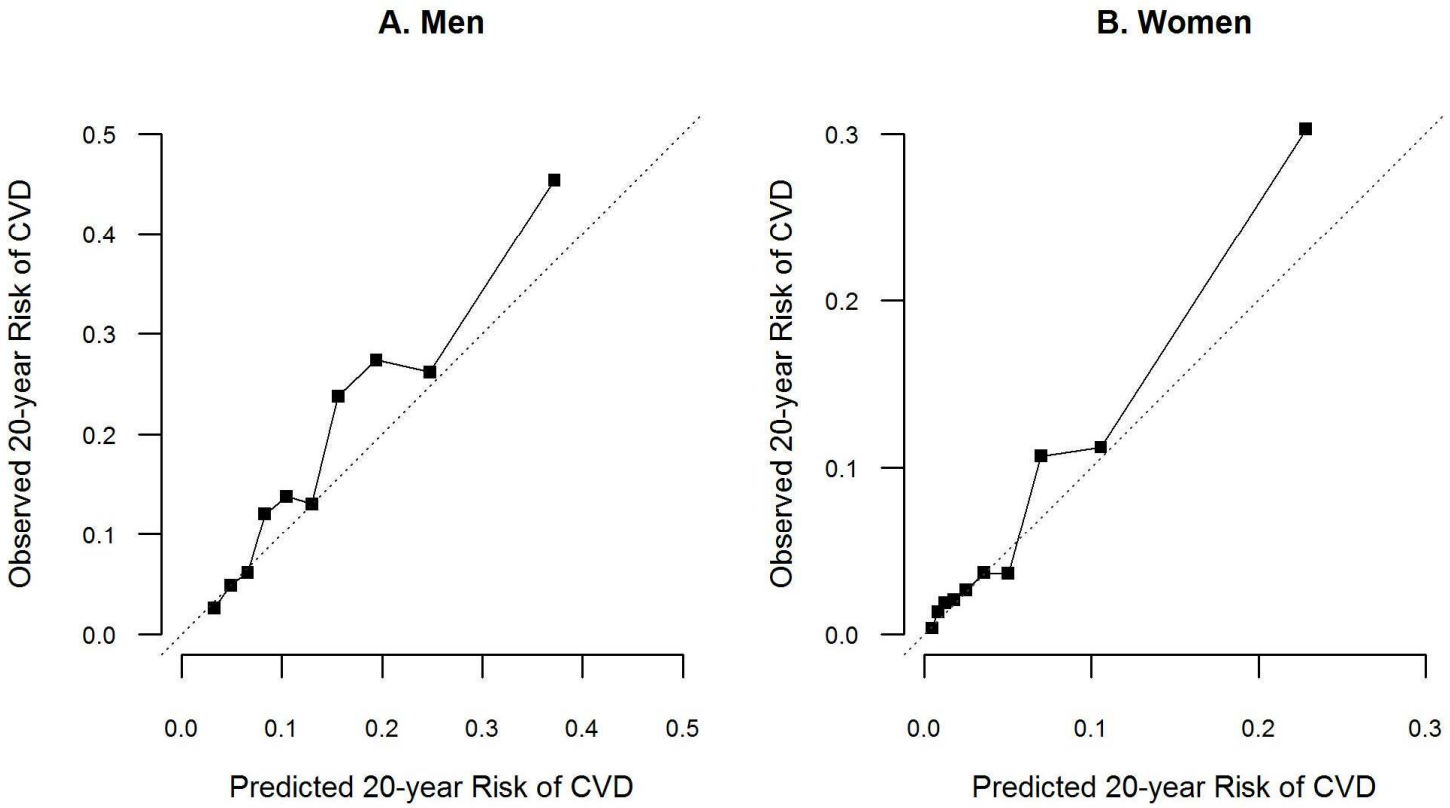


Figure 2: Calibration plot in the validation set for the CAMUNI and the Framingham risk scores.
Men (left) and women (right), 35 to 69 years old, free of CVD at baseline.

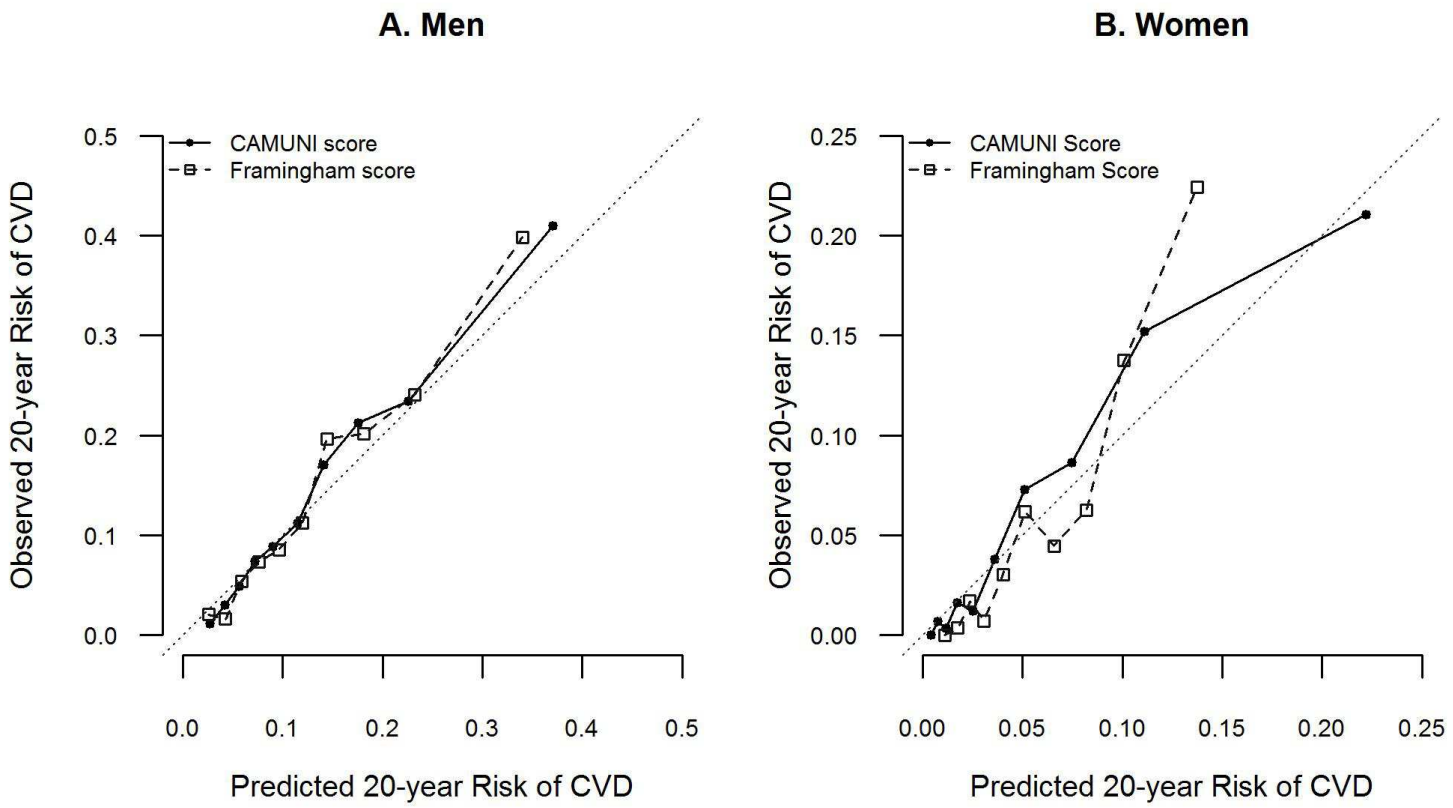
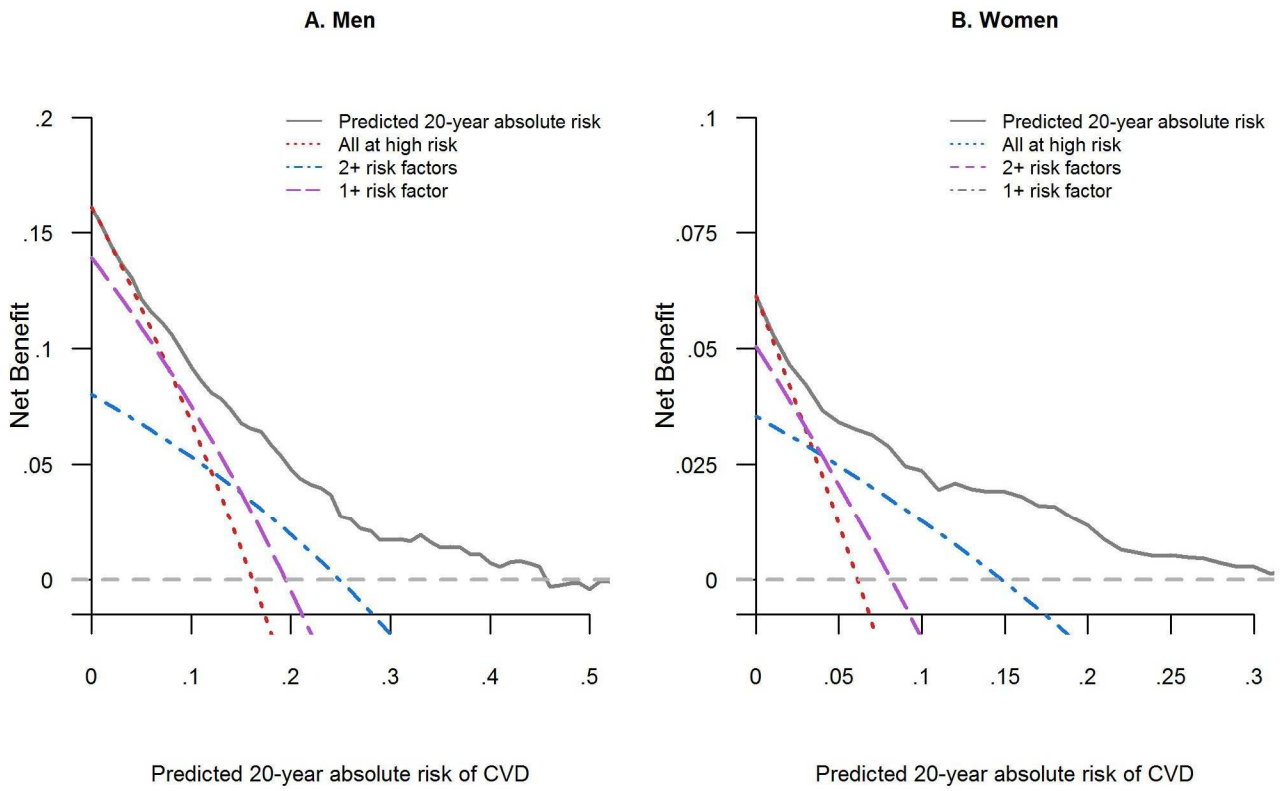
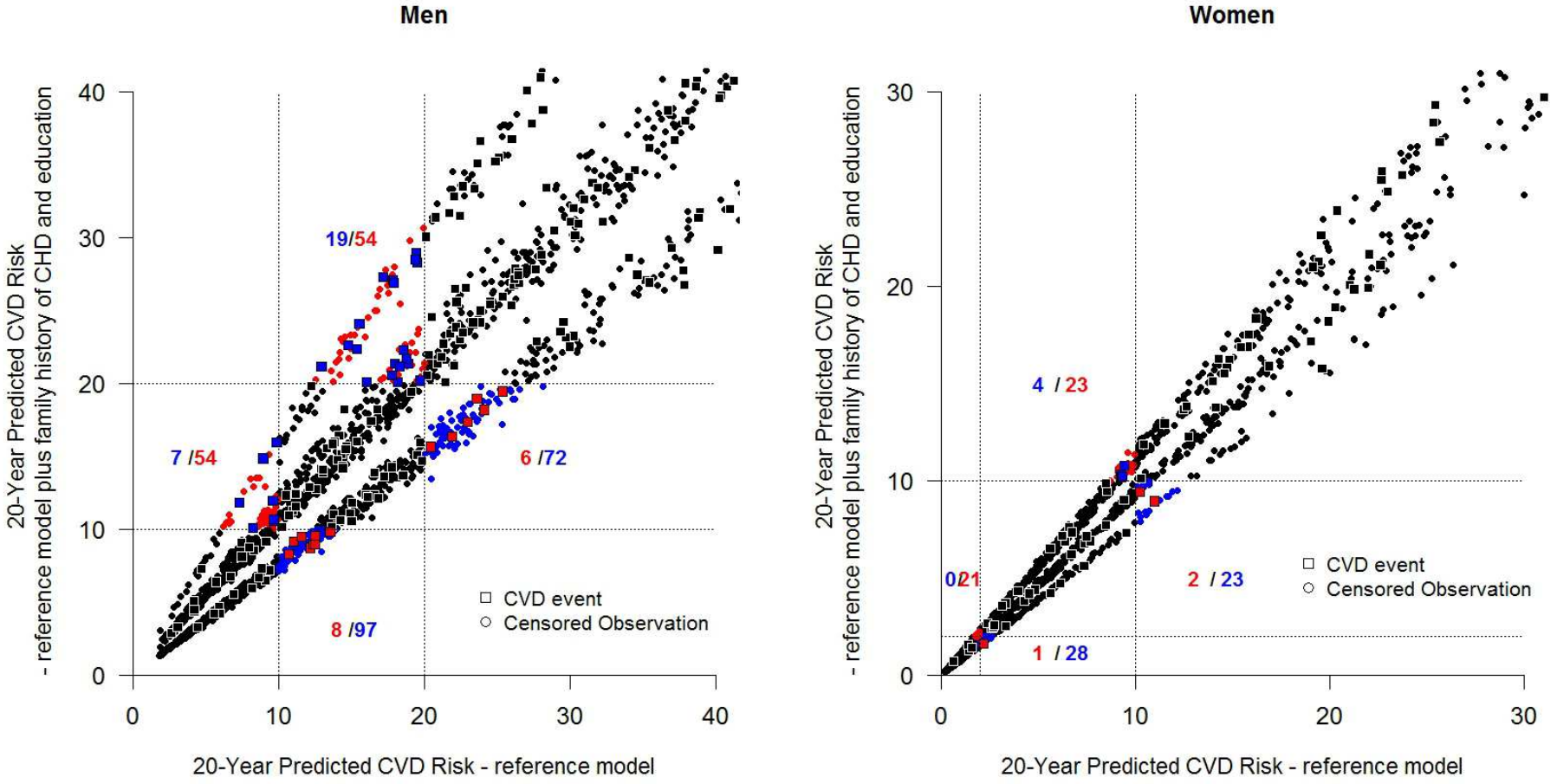


Figure 3: Decision curve for the CAMUNI 20-year risk prediction model in the derivation set, as compared to a risk stratification based on the number of risk factors. Men (left) and women (right), 35 to 69 years old, free of CVD at baseline.



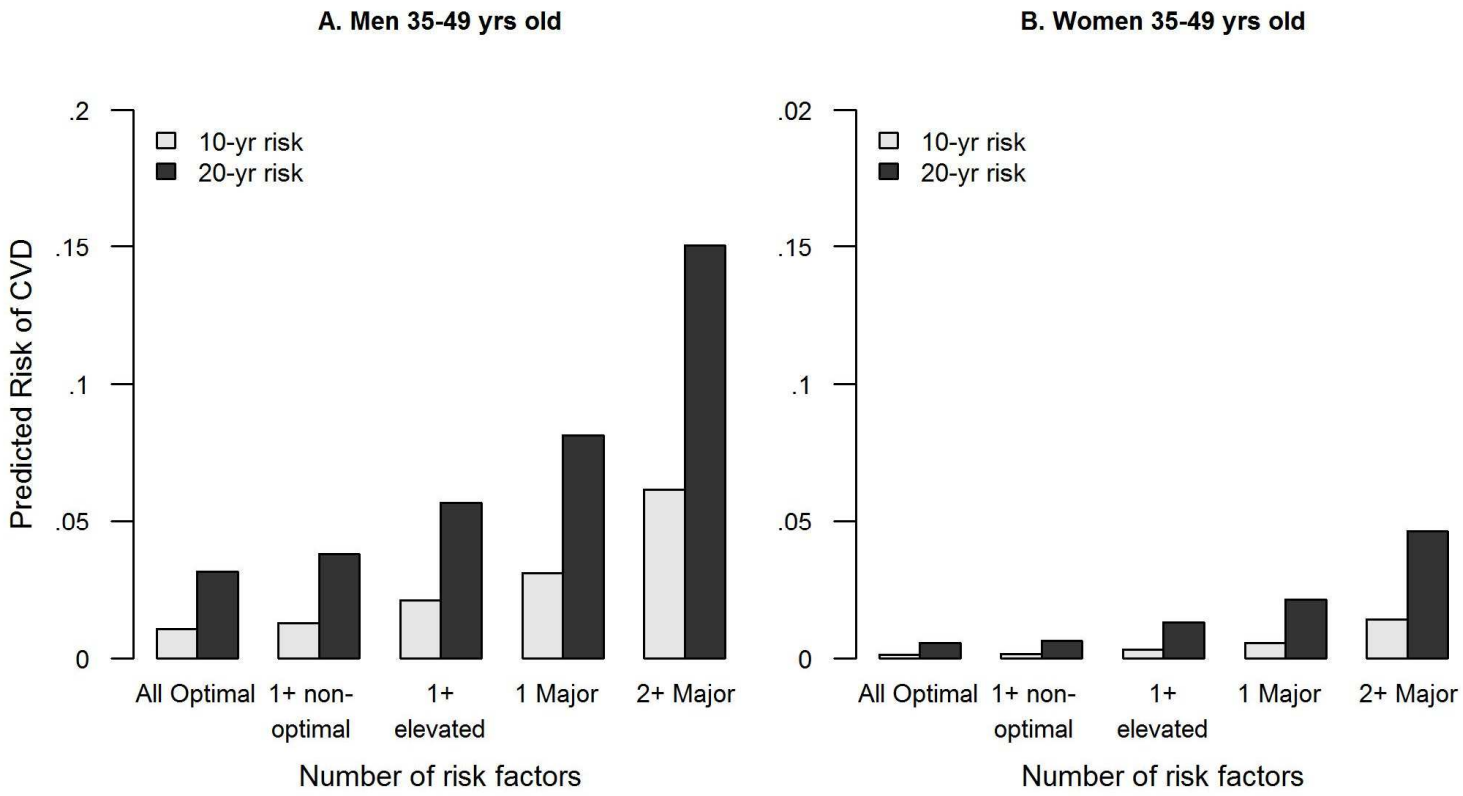
Net Benefit: (TP-w*FP)/n, where TP = True Positive; FP = False Positive; w = (Absolute risk threshold)/(1- (Absolute risk threshold)); n=sample size
Number of risk factors: total cholesterol>240 mg/dl; HDL-cholesterol <40 [men] or <50 [women] mg/dl; systolic blood pressure >160 mmHg; smoking; diabetes

Figure 4: Reclassification plot for the model with family history of CHD and education, with respect to the reference^ 20-yer risk prediction model. Men (left) and women (right), 35 to 69 years old, free of CVD at baseline. The MONICA-Brianza and PAMELA Study



^: The reference model includes age, total cholesterol, HDL-cholesterol, systolic blood pressure, anti-hypertensive treatment, smoking and diabetes.

Figure 5: Distribution of predicted 10-year and 20-year risk of first major CVD event, according to the number of risk factors.
Men (left) and women (right), 35 to 49 years old, free of CVD at baseline



Risk factors stratification derived from Lloyd-Jones[9].
All optimal: total cholesterol <180 mg/dl, HDL-Cholesterol >= 40 mg/dl [men] or >= 50 mg/dl [women], blood pressure <120/80 mmHg, non smoker, non diabetic;
1+ non-optimal: total cholesterol 180 to 199 mg/dl, systolic blood pressure 120 to 139 mmHg, diastolic blood pressure 80 to 89 mmHg, non smoker, non diabetic
1+ elevated: total cholesterol 200 to 239 mg/dl, systolic blood pressure 140 to 159 mmHg, diastolic blood pressure 90 to 99 mmHg, non smoker, non diabetic
Major risk factor: total cholesterol >=240 mg/dl, HDL-Cholesterol <40 mg/dl [men] or <50 mg/dl [women], systolic blood pressure>=160 mmHg or treatment, diastolic blood pressure >=100 mmHg, smoker, or diabetic

APPENDIX: THE reSAS PACKAGE

The "reSAS" is a SAS package written by the author which includes several macros to assess calibration (Grønnesby-Bogan goodness-of-fit test), discrimination [AUC(t)] and the internal validity of a given prediction model in the survival setting, as well as to compare two models in terms of discrimination [Δ-AUC(t), IDI(t)] and risk stratification [NRI(t)]. All these quantities have been described in the methods section.

The underlying survival model for all the macros is Cox proportional hazards, with time-on-study on the time scale (macro variable TIME). The specific assumptions for the Cox model need to be tested separately. In addition, as AUC(t), IDI(t) and NRI(t) are computed at a given time t, the macro variable TIME_STOP needs to be specified on the same time scale as TIME; if the survival time is in years, a TIME_STOP = 10 will return the AUC at 10-year time interval from baseline (AUC(10)).

The candidate models (reference plus all the additional models) need to be listed in a SAS dataset before running the macros, as below:

```
DATA MODEL_LIST;
infile datalines delimiter=',';
LENGTH MODEL $175. LABEL $25.;
INPUT NUM MODEL LABEL;
DATALINES;
1, AGE SEX SBP TRATT DIAB SMK, REF_MODEL,
2, AGE SEX SBP TRATT DIAB SMK choldl hdldl hdldl, TC_HDL,
3, AGE SEX SBP TRATT DIAB SMK choldl hdldl hdldl*SEX, TC_HDL_INT,
4, AGE SEX SBP TRATT DIAB SMK CT_CL2 CT_CL3 CHDL_CL1 CHDL_CL2 CHDL_CL1 CHDL_CL2,
CLASS_TC_HDL,
5, AGE SEX SBP TRATT DIAB SMK CT_CL2 CT_CL3 CHDL_CL1 CHDL_CL2 CHDL_CL1*SEX
CHDL_CL2*SEX, CLASS_TC_HDL_INT
RUN;
```

The first row is referring to the reference model, while the other rows are relative to the additional contribution of total- and HDL-cholesterol, either as continuous variables (model 2) or in classes (model 4), with a sex*HDL-cholesterol interaction (model 3 and model 5, respectively, for continuous or classes variables). The last column in the MODEL_LIST dataset is a label to identify the model in each output dataset. The macros can handle interactions as well as class variables; in this latter case, dummies need to be created in advance in the analysis dataset.

The reference model should be specified as macro variable REF_MODEL in the macros computing Δ-AUC(t), IDI(t) and NRI(t).

Bootstrapped confidence intervals for AUC(t), Δ-AUC(t), IDI(t) and NRI(t) can be obtained from the OVERALL_ANALYSIS macro. The macro produces an output dataset with the parameters'

estimates in each bootstrapped sample; confidence intervals can be the obtained using a standard program. In this macro, by setting the macro variable NUMBOOT = 0, AUC(t), Δ-AUC(t), IDI(t) and NRI(t) are computed only on the original dataset.

Each macro produces output datasets, which can be easily exported in excel for tables production. The CALIB_PLOT macro produces a standard calibration plot for deciles of predicted risk, as well as a SAS dataset with observed and predicted risk, to customize the plot if needed, either in SAS or in a different environment. The ROC_PLOT macro produces a standard plot of AUC contrasting each tested model with the reference ones well as a SAS dataset with observed and predicted risk, to customize the plot if needed, either in SAS or in a different environment.
All the macros in the package were written using SAS Release 9.2.

List of macros in the package:

1A. Model calibration – Grønnesby-Bogan goodness-of-fit test
1B. Model calibration – Calibration plot of expected vs. predicted risk of event at time t
2A. Model discrimination – AUC(t), Δ- AUC(t), IDI(t)
2B. Model discrimination – plot the ROC curve
3. Net Reclassification Improvement – NRI(t), clinical NRI(t), continuous NRI(t)
4. Bootstrapped Confidence Intervals for AUC(t), Δ-AUC(t), IDI(t), NRI(t)
5A. Internal validation analysis: over-optimism.
5B. Internal validation analysis: calibration slope.

References for the quantities estimated in the macros:

Model calibration, goodness-of-fit test, calibration plot:
May S., Hosmer DW. A simplified method of calculating an overall Goodness-of-Fit test for the Cox proportional hazards model. *Lifetime Data Anal* 1998;4:109-120

Steyerberg EW. Clinical prediction models. 2009 Springer Science + Business Media, LLC. New York, US.

Model discrimination and reclassification:
Chambless LE, Cummiskey CP, and Cui G. Several methods to assess improvement in risk prediction models: extension to survival analysis. *Statist med* 2011;30:22–28

Internal validation analysis:
Harrel FE, Lee KL and Marck DB. Tutorial in biostatistics: multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statist med* 1996;15:361–387

Steyerberg EW. Clinical prediction models. 2009 Springer Science + Business Media, LLC. New York, US.

1A. Model calibration – Grønnesby-Bogan goodness-of-fit test

```
%MACRO CALIB_TEST(DATASET, OUTPUT_TEST, OUTPUT_EXP, EV, TIME, MIN_MOD, MAX_MOD,
COND=);
```

Description of macro variables:
DATASET: name of the input SAS dataset
OUTPUT_TEST: name of the output SAS dataset with the results of the Grønnesby-Bogan test (chi-square value, chi-square p-value)
OUTPUT_EXP: name of the output SAS dataset with the number of observed and expected events in each decile of predicted risk
EV: variable name for the event
TIME: variable name for the follow-up time
MIN_MOD: number corresponding to the first model to be tested in the MODEL_LIST dataset
MAX_MOD: number corresponding to the last model to be tested in the MODEL_LIST dataset
COND=: run the whole analysis (including model estimate) in a subgroup of subjects, i.e. men only (optional)

```
%MACRO CALIB_TEST(DATASET, OUTPUT_TEST, OUTPUT_EXP, EV, TIME, MIN_MOD, MAX_MOD,
COND=);

PROC DATASETS NOLIST; DELETE &OUTPUT_TEST &OUTPUT_EXP; QUIT;
%DO CI_MOD = &MIN_MOD %TO &MAX_MOD; *RUN WITHIN EACH MODEL TO EVALUATE;
DATA _NULL_;SET MODEL_LIST;
CALL SYMPUT("MODEL", MODEL);
CALL SYMPUT("LABEL", LABEL);
WHERE NUM = &CI_MOD; RUN;
************************************************************
1. RUN THE REGRESSION MODELS AND FIND PERCENTILES OF PRED RISK;
      PROC PHREG DATA=&DATASET NOPRINT;
      MODEL &TIME*&EV(0)=&MODEL;
      OUTPUT OUT = PRED XBETA = XBETA SURVIVAL = SURV;
      WHERE &COND;
      RUN;
      PROC UNIVARIATE DATA = PRED NOPRINT;
      VAR XBETA; OUTPUT OUT = TTT
          pctlpts  = 10 20 30 40 50 60 70 80 90
          pctlpre  = XBETA
          pctlname = P10 P20 P30 P40 P50 P60 P70 P80 P90; RUN;
DATA _NULL_; SET TTT;
CALL SYMPUT("P10", XBETAP10);CALL SYMPUT("P20", XBETAP20);CALL SYMPUT("P30",
XBETAP30);CALL SYMPUT("P40", XBETAP40);CALL SYMPUT("P50", XBETAP50);
CALL SYMPUT("P60", XBETAP60);CALL SYMPUT("P70", XBETAP70);CALL SYMPUT("P80",
XBETAP80);CALL SYMPUT("P90", XBETAP90);RUN;
      DATA PRED; SET PRED;
      IF XBETA LT &P10 THEN P_CLASS = 1;
      ELSE IF XBETA LT &P20 THEN P_CLASS = 2;
      ELSE IF XBETA LT &P30 THEN P_CLASS = 3;
      ELSE IF XBETA LT &P40 THEN P_CLASS = 4;
      ELSE IF XBETA LT &P50 THEN P_CLASS = 5;
      ELSE IF XBETA LT &P60 THEN P_CLASS = 6;
      ELSE IF XBETA LT &P70 THEN P_CLASS = 7;
      ELSE IF XBETA LT &P80 THEN P_CLASS = 8;
      ELSE IF XBETA LT &P90 THEN P_CLASS = 9;
      ELSE P_CLASS = 10;
      SURV_ZERO = SURV**EXP(-XBETA);
      EXP = -EXP(XBETA)*LOG(SURV_ZERO);
      RUN;
```

```
**********************************
2. GRONNESBY-BORGAN GOODNESS OF FIT;
      PROC PHREG DATA=PRED;
      CLASS P_CLASS;
      MODEL &TIME*&EV(0)=&MODEL P_CLASS;
      CONTRAST "GOF"  P_CLASS  1 0 0 0 0 0 0 0 0,
                               P_CLASS  0 1 0 0 0 0 0 0 0,
                               P_CLASS  0 0 1 0 0 0 0 0 0,
                               P_CLASS  0 0 0 1 0 0 0 0 0,
                               P_CLASS  0 0 0 0 1 0 0 0 0,
                               P_CLASS  0 0 0 0 0 1 0 0 0,
                               P_CLASS  0 0 0 0 0 0 1 0 0,
                               P_CLASS  0 0 0 0 0 0 0 1 0,
                               P_CLASS  0 0 0 0 0 0 0 0 1/ TEST(SCORE);
      ODS OUTPUT CONTRASTTEST = GOF;
      RUN;
      DATA GOF; SET GOF; RENAME ScoreChiSq = GB_TEST ProbScoreChiSq = GB_PVAL;
      DROP Contrast ScoreDF;
      LENGTH EVENT LABEL $25.;
      LABEL = "&LABEL";
      MODEL = &CI_MOD;
      EVENT = "&EV";
      RUN;
**************************************************
3. EXPECTED AND PREDICTED COUNTS BY DECILE OF XBETA;
      PROC MEANS DATA = PRED N SUM NOPRINT;
      CLASS P_CLASS; OUTPUT OUT = EXP SUM = OBS EXP;
      VAR &EV EXP;
      TYPES P_CLASS;
      RUN;
      DATA EXP;
      SET EXP;
      LENGTH EVENT LABEL $25.; LABEL = "&LABEL"; MODELLO = &CI_MOD; EVENT =
      "&EV";
      RUN;
PROC DATASETS NOLIST;
APPEND DATA = EXP  BASE = &OUTPUT_EXP  FORCE;
APPEND DATA = GOF  BASE = &OUTPUT_TEST FORCE;
DELETE TTT PRED GOF EXP; QUIT;
%END; *END RUN FOR A GIVEN MODEL;
%MEND;
```

1B. Model calibration – Calibration plot of expected vs. predicted risk of event at time t

```
%MACRO CALIB_PLOT(DATASET, OUTPUT_PLOT, EV, TIME, TIME_STOP, MIN_MOD, MAX_MOD,
COND=);
```

Description of macro variables:
DATASET: name of the input SAS dataset
OUTPUT_PLOT: name of the output SAS dataset with the observed risk (Kaplan-Meier) at time t (varianle OBS_RISK) and the mean of predicted risk at time t (variable PRED_RISK) by deciles of predicted risk
EV: variable name for the event
TIME: variable name for the follow-up time
TIME_STOP: number corresponding to time t of prediction interval
MIN_MOD: number corresponding to the first model to be tested in the MODEL_LIST dataset
MAX_MOD: number corresponding to the last model to be tested in the MODEL_LIST dataset
COND=: run the whole analysis (including model estimate) in a subgroup of subjects, i.e. men only (optional)

```
%MACRO CALIB_PLOT(DATASET, OUTPUT_PLOT, EV, TIME, TIME_STOP, MIN_MOD, MAX_MOD,
COND=);
PROC DATASETS NOLIST; DELETE &OUTPUT_PLOT; QUIT;

%DO CI_MOD = &MIN_MOD %TO &MAX_MOD;   *RUN WITHIN EACH MODEL TO EVALUATE;
DATA _NULL_;SET MODEL_LIST;
CALL SYMPUT("MODEL", MODEL);
CALL SYMPUT("LABEL", LABEL);
WHERE NUM = &CI_MOD; RUN;

*****************************************
1. REGRESSION MODEL AND PREDICTED SURV(T);
      DATA SURV_ZERO; SET &DATASET;
      &EV = .; &TIME = &TIME_STOP; RUN;
      DATA BIS_&DATASET; SET &DATASET SURV_ZERO; RUN;

      PROC PHREG DATA=BIS_&DATASET NOPRINT;
      MODEL &TIME*&EV(0)=&MODEL;
      OUTPUT OUT = PRED XBETA = XBETA SURVIVAL=SURV;
      WHERE &COND;
      RUN;
*****************************************************
2. DECILES OF PREDICTED RISK - USING XBETA EQUIVALENTLY;
      PROC UNIVARIATE DATA = PRED NOPRINT;
      VAR XBETA; OUTPUT OUT = TTT pctlpts  = 10 20 30 40 50 60 70 80 90
                                  pctlpre  = XBETA
                                 pctlname = P10 P20 P30 P40 P50 P60 P70 P80 P90;
RUN;
      DATA _NULL_; SET TTT;
      CALL SYMPUT("P10", XBETAP10);CALL SYMPUT("P20", XBETAP20);CALL
SYMPUT("P30", XBETAP30);CALL SYMPUT("P40", XBETAP40);CALL SYMPUT("P50",
XBETAP50);
      CALL SYMPUT("P60", XBETAP60);CALL SYMPUT("P70", XBETAP70);CALL
SYMPUT("P80", XBETAP80);CALL SYMPUT("P90", XBETAP90);
      RUN;

      DATA PRED_PCT; SET PRED;
      IF XBETA LE &P10 THEN P_CLASS  = 1;
      ELSE IF XBETA LE &P20 THEN P_CLASS = 2;
      ELSE IF XBETA LE &P30 THEN P_CLASS = 3;
      ELSE IF XBETA LE &P40 THEN P_CLASS = 4;
```

```
      ELSE IF XBETA LE &P50 THEN P_CLASS = 5;
      ELSE IF XBETA LE &P60 THEN P_CLASS = 6;
      ELSE IF XBETA LE &P70 THEN P_CLASS = 7;
      ELSE IF XBETA LE &P80 THEN P_CLASS = 8;
      ELSE IF XBETA LE &P90 THEN P_CLASS = 9;
      ELSE P_CLASS = 10;
      RUN;
****************************************
3. MEAN(PRED RISK(T)) BY DECILES OF XBETA;
      PROC MEANS DATA = PRED_PCT MEAN NOPRINT;
      CLASS P_CLASS; VAR SURV;
      WHERE &EV = .; *IMPORTANT: USE PRED(T);
      OUTPUT OUT = PRED_AVR MEAN = PRED_MEAN; TYPES P_CLASS;RUN;
************************************************
4. KAPLAN-MEIER SURV(T) BY DECILES OF PREDICTED RISK;
    DATA KM_EST;
      SET PRED_PCT;
    WHERE &EV NE .;*IMPORTANT: USE ORIGINAL DATA ON EVENT;
    RUN;
      PROC SORT DATA = KM_EST; BY P_CLASS; RUN;
      PROC LIFETEST DATA = KM_EST OUTSURV=KM NOPRINT;
      BY P_CLASS;
      TIME &TIME*&EV(0);
      RUN;
      PROC SORT DATA = KM; BY P_CLASS &TIME; RUN;
      DATA LAST_KM;
      SET KM; BY P_CLASS &TIME; IF LAST.P_CLASS; WHERE _CENSOR_ = 0 AND &TIME LE
&TIME_STOP; RUN;
************************************************************
5. MERGE PREDICTED AND OBSERVED RISK(T) BY DECILES OF XBETA;
      DATA G_CALIB;
      MERGE PRED_AVR (DROP= _TYPE_ _FREQ_) LAST_KM (KEEP = P_CLASS SURVIVAL);
      BY P_CLASS; IF SURVIVAL = . THEN SURVIVAL = 1;
      PRED_RISK     = 1 - PRED_MEAN;
      OBS_RISK      = 1 - SURVIVAL;
      LENGTH EVENT LABEL $25.; LABEL = "&LABEL"; MODEL = &CI_MOD; EVENT = "&EV";
      RUN;

PROC DATASETS NOLIST; APPEND DATA = G_CALIB BASE = &OUTPUT_PLOT FORCE;
DELETE BIS_&DATASET TTT SURV_ZERO PRED_AVR PRED_PCT PRED KM KM_EST LAST_KM
G_CALIB; QUIT;
%END; *END RUN FOR A GIVEN MODEL;

*annotated dataset for 45° line;
data anno;
   function='move';
   xsys='1'; ysys='1';
   x=0; y=0;
   output;

   function='draw';
   xsys='1'; ysys='1';
   color='black';
   LINE = 3;
   x=100; y=100;
   output;
run;
DATA DDD; SET CALIB_PLOT; RENAME OBS_RISK=Y PRED_RISK=X; RUN;
goptions reset=global gunit=pct cback=white device=win colors=(black blue green
red)
ftitle=swissb ftext=swiss htitle=6 htext=4;
%DO UIUI = &MIN_MOD %TO &MAX_MOD;
symbol&UIUI interpol=join width=0.5 VALUE=DOT height=1.3;
```

```
%END;
TITLE1 ;
TITLE2 "Calibration Plot";
axis1  order=(0 to 0.5 by 0.1) label = ("Predicted Risk")     length=60;
axis2  order=(0 to 0.5 by 0.1) label = (a=90 "Observed Risk") length=69;
legend1 down=%EVAL(&MAX_MOD - &MIN_MOD) shape=symbol(2,1) position=(BOTTOM RIGHT
INSIDE) mode=share LABEL=("Model");
PROC GPLOT DATA = DDD;
PLOT Y*X=MODEL / ANNO=ANNO haxis=axis1 vaxis=axis2 legend=legend;
RUN; QUIT;
PROC DATASETS NOLIST; DELETE ANNO DDD;QUIT;
%MEND;
```

2A. Model discrimination – AUC(t), Δ- AUC(t), IDI(t)

```
%MACRO DISCR_ANALYSIS(DATASET, OUTPUT, EV, TIME, TIME_STOP, REF_MODEL=,
CONTRAST_MIN=, CONTRAST_MAX=, COND=, PRINT=);
```

Description of macro variables:

DATASET: name of the input SAS dataset
OUTPUT: name of the output SAS dataset containing AUC, R, average sensitivity (ISt) and average 1-specificity (IPt) for each contrast model, as well as for the reference model; the difference in AUC between each contrast model and the reference; and the IDI.
EV: variable name for the event
TIME: variable name for the follow-up time
TIME_STOP: number corresponding to time t of prediction interval
REF_MODEL: reference model
CONTRAST_MIN: number corresponding to the first contrast model to be tested in the MODEL_LIST dataset
CONTRAST_MAX: number corresponding to the last contrast model to be tested in the MODEL_LIST dataset
COND=: run the whole analysis (including model estimate) in a subgroup of subjects, i.e. men only (optional)
PRINT=: specify YES to print the output from the PHREG procedure

```
%MACRO DISCR_ANALYSIS(DATASET, OUTPUT, EV, TIME, TIME_STOP, REF_MODEL=,
CONTRAST_MIN=, CONTRAST_MAX=, COND=, PRINT=);

PROC DATASETS NOLIST; DELETE &OUTPUT; QUIT;

DATA FAKE; SET &DATASET; &EV = .; &TIME = &TIME_STOP; RUN;
DATA BIS_&DATASET; SET &DATASET FAKE; RUN;

*******************************************************
1. REFERENCE MODEL: REGRESSION, AVR SENS, AVR 1-SPEC, R;
      PROC PHREG DATA=BIS_&DATASET %if %upcase(&PRINT)^=YES %then %do; noprint
      %end;;
      MODEL &TIME*&EV(0)=&REF_MODEL;
      OUTPUT OUT = PRED_BASE XBETA = XBETA SURVIVAL = SURV_BASE ;
      WHERE &COND;
      RUN;

      DATA PRED_BASE;
      SET PRED_BASE;
      S_1_S_BASE=SURV_BASE*(1-SURV_BASE);
      St_square_BASE=SURV_BASE**2;
      RUN;

      PROC MEANS DATA = PRED_BASE MEAN VAR NOPRINT;
      VAR SURV_BASE S_1_S_BASE St_square_BASE;
      WHERE &EV = .;
      OUTPUT OUT = IDI_BASE MEAN = MEAN_SURV_BASE ES_1_S_BASE E_St_Square_BASE
      VAR = VAR_SURV_BASE;
      RUN;

      DATA IDI_BASE;
      SET IDI_BASE;
      R_BASE = VAR_SURV_BASE/(MEAN_SURV_BASE*(1-MEAN_SURV_BASE));
      ISt_BASE=1-ES_1_S_BASE/(1-MEAN_SURV_BASE);
```

```
      IPt_BASE=1-E_St_Square_BASE/MEAN_SURV_BASE;
      RUN;
***********************
2. REFERENCE MODEL: AUC;
      DATA PRED_AUC_BASE; SET PRED_BASE; KEEP XBETA SURV_BASE; WHERE &EV = .;
*AND &COND2; RUN;
      PROC SORT DATA = PRED_AUC_BASE; BY XBETA; RUN;

      PROC IML;
      USE PRED_AUC_BASE;
      READ ALL INTO Z_BASE;
      CLOSE PRED_AUC_BASE;

      N=NROW(Z_BASE);
      E1_StU_StV_BASE=0;
      EStZ_BASE=0;

        do i=1 to n;
            EStZ_BASE=EStZ_BASE+z_BASE[i,2];
            l_survobs_BASE = j(n,1,(1-z_BASE[i,2]));
            smaller_BASE   = (z_BASE[,1]<z_BASE[i,1]);
            equal_BASE     = (z_BASE[,1]=z_BASE[i,1]);
            smaller_BASE[i:n]=0;
            equal_BASE[i:n]=0;

      E1_StU_StV_BASE=E1_StU_StV_BASE+sum(l_survobs_BASE#z_BASE[,2]#smaller_BASE
)+0.5*sum(l_survobs_BASE#z_BASE[,2]#equal_BASE);
        end;

            E1_StU_StV_BASE=  E1_StU_StV_BASE/(n**2);
        EStZ_BASE=EStZ_BASE/n;
        auct_BASE=E1_StU_StV_BASE/(EStZ_BASE*(1-EStZ_BASE));

create AUC_BASE from auct_BASE[colname={AUC_BASE}];  /*create dataset for
bootstrap*/
append from aucT_BASE;
close AUC_BASE;
QUIT;
PROC DATASETS NOLIST; DELETE PRED_BASE PRED_AUC_BASE; RUN;
*******************************************************
3. CONTRAST MODELS: REGRESSION, AVR SENS, AVR 1-SPEC, R;
%DO CI_MOD = &CONTRAST_MIN %TO &CONTRAST_MAX; *RUN WITHIN EACH MODEL TO
EVALUATE;
DATA _NULL_;SET MODEL_LIST;
CALL SYMPUT("MODEL", MODEL);
CALL SYMPUT("LABEL", LABEL);
WHERE NUM = &CI_MOD; RUN;

      PROC PHREG DATA=BIS_&DATASET %if %upcase(&PRINT)^=YES %then %do; noprint
      %end;;
      MODEL &TIME*&EV(0)=&MODEL;
      OUTPUT OUT = PRED XBETA = XBETA SURVIVAL=SURV;
      WHERE &COND;
      RUN;

      DATA PRED;
      SET PRED;
      S_1_S=SURV*(1-SURV);
      St_square=SURV**2;
      RUN;

      PROC MEANS DATA = PRED MEAN VAR NOPRINT;
      VAR SURV S_1_S St_square;
```

```
      WHERE &EV = .; *AND &COND2;
      OUTPUT OUT = IDI MEAN = MEAN_SURV ES_1_S E_St_Square VAR = VAR_SURV;
      RUN;

DATA IDI;
SET IDI;
R = VAR_SURV/(MEAN_SURV*(1-MEAN_SURV));
ISt=1-ES_1_S/(1-MEAN_SURV);
IPt=1-E_St_Square/MEAN_SURV;
RUN;
************************
3. CONTRAST MODELS: AUC;
      DATA PRED_FIN; SET PRED; KEEP XBETA SURV; WHERE &EV = .;*AND &COND2; RUN;
      PROC SORT DATA = PRED_FIN; BY XBETA; RUN;
      PROC IML;
      USE PRED_FIN;
      READ ALL INTO Z;
      CLOSE PRED_FIN;

      N=NROW(Z);
      E1_StU_StV=0;
    EStZ=0;

        do i=1 to n;
            EStZ=EStZ+z[i,2];
            l_survobs = j(n,1,(1-z[i,2]));
            smaller   = (z[,1]<z[i,1]);
            equal     = (z[,1]=z[i,1]);
            smaller[i:n]=0;
            equal[i:n]=0;
E1_StU_StV=    E1_StU_StV+    sum(l_survobs#z[,2]#smaller)
+0.5*sum(l_survobs#z[,2]#equal);
        end;
        E1_StU_StV=  E1_StU_StV/(n**2);
        EStZ=EStZ/n;
        auct=E1_StU_StV/(EStZ*(1-EStZ));

create AUC_&LABEL from auct[colname={AUC}];
append from aucT;
close AUC_&LABEL;
QUIT;

DATA RRR_FINALE;
MERGE AUC_&LABEL IDI IDI_BASE AUC_BASE;
DROP _TYPE_ _FREQ_ MEAN_SURV_BASE VAR_SURV_BASE MEAN_SURV VAR_SURV ES_1_S
E_ST_SQUARE ES_1_S_BASE E_ST_SQUARE_BASE;
DELTA_AUC = AUC - AUC_BASE;
IDI = R - R_BASE;
LENGTH EVENT LABEL $25.;
MODEL = &CI_MOD;
LABEL = "&LABEL";
EVENT = "&EV";
RUN;

PROC DATASETS NOLIST;
APPEND DATA = RRR_FINALE BASE = &OUTPUT FORCE;
DELETE RRR_FINALE PRED PRED_FIN AUC_&LABEL IDI; QUIT;
%END; *END RUN FOR A GIVEN MODEL;
PROC DATASETS NOLIST; DELETE BIS_&DATASET FAKE IDI_BASE AUC_BASE; QUIT;
%MEND;
```

2B. Model discrimination – plot the ROC curve

```
%MACRO ROC_PLOT(DATASET, OUTPUT, EV, TIME, TIME_STOP, REF_MODEL=, CONTRAST_MIN=,
CONTRAST_MAX=, COND=);
```

Description of macro variables:

DATASET: name of the input SAS dataset
OUTPUT: name of the output SAS dataset containing sensitivity and specificity at any percentile of predicted risk.
EV: variable name for the event
TIME: variable name for the follow-up time
TIME_STOP: number corresponding to time t of prediction interval
REF_MODEL: reference model
CONTRAST_MIN: number corresponding to the first contrast model to be tested in the MODEL_LIST dataset
CONTRAST_MAX: number corresponding to the last contrast model to be tested in the MODEL_LIST dataset
COND=: run the whole analysis (including model estimate) in a subgroup of subjects, i.e. men only (optional)

```
%MACRO ROC_PLOT(DATASET, OUTPUT, EV, TIME, TIME_STOP, REF_MODEL=, CONTRAST_MIN=,
CONTRAST_MAX=, COND=);
PROC DATASETS NOLIST; DELETE &OUTPUT; QUIT;

DATA AR_DIST; SET &DATASET; &EV = .; &TIME = &TIME_STOP; RUN;
DATA AR_DIST; SET AR_DIST &DATASET; RUN;

***************
REFERENCE MODEL;
PROC PHREG DATA = AR_DIST NOPRINT;
WHERE &COND;
OUTPUT OUT = PRED_REF XBETA = XBETA SURVIVAL=SURV;
MODEL &TIME*&EV(0) = &REF_MODEL;
RUN;

DATA PRED_REF; SET PRED_REF;RISK = 1 - SURV; RUN;
*XBETA PCTS - ONLY WHERE &EV = . FOR THE REF MODEL;
PROC UNIVARIATE DATA = PRED_REF NOPRINT;
VAR XBETA RISK;
OUTPUT OUT = AR_PCT_REF PCTLPRE=QR_ RI_ PCTLPTS = (1 TO 99 BY 1);
WHERE &EV = .; RUN;

*OVERALL(DENOMINATOR) - ONLY WHERE &EV = . FOR REF MODEL;
PROC MEANS DATA = PRED_REF MEAN N NOPRINT;
VAR SURV RISK; OUTPUT OUT = SENS_OVR_REF MEAN = MEAN_SURV_OVR MEAN_RISK_OVR N =
NUM;
WHERE &EV = .; RUN;
OPTION NONOTES;
      %DO I = 1 %TO 99;*RUN WITHIN PCTS;
      DATA _NULL_; SET AR_PCT_REF;CALL SYMPUT("PCT", QR_&I);CALL SYMPUT("RISK",
RI_&I); RUN;
      DATA SENS_&I; SET PRED_REF; WHERE &EV = . AND XBETA GE &PCT; RUN;
      PROC MEANS DATA = SENS_&I SUM NOPRINT;
      VAR   SURV          RISK    ; OUTPUT OUT = SENS_M_&I
      SUM = SUM_SURV_SENS SUM_RISK;
      RUN;
      DATA SENS_M_&I; SET SENS_M_&I;
      KEEP SUM_SURV_SENS SUM_RISK QR PCT RISK;
```

```
      QR = &I; PCT=&PCT; RISK=&RISK; RUN;

      DATA SPEC_&I; SET PRED_REF; WHERE &EV = . AND XBETA LT &PCT; RUN;
      PROC MEANS DATA = SPEC_&I SUM NOPRINT;
      VAR SURV; OUTPUT OUT = SPEC_M_&I SUM = SUM_SURV;
      RUN;
      DATA SPEC_M_&I; SET SPEC_M_&I; KEEP SUM_SURV QR; QR = &I; RUN;

      DATA XXX_&I; MERGE SENS_M_&I SPEC_M_&I SENS_OVR_REF; RUN;

      PROC DATASETS NOLIST;
      APPEND DATA = XXX_&I BASE = ROC_PCT_REF FORCE;
      DELETE XXX_&I SPEC_M_&I SENS_M_&I SENS_&I SPEC_&I; QUIT;
      %END;*END CYCLE WITHIN PCTS;
      OPTION NOTES;

DATA ROC_REF_FINAL; SET ROC_PCT_REF;
FP   = SUM_SURV_SENS/NUM;
SENS = (SUM_RISK/NUM)*(1/MEAN_RISK_OVR);
SPEC = (SUM_SURV/NUM)*(1/MEAN_SURV_OVR);
I_SPEC = 1-SPEC;
KEEP QR SENS SPEC I_SPEC MODEL LABEL TIME PCT RISK mean_risk_ovr mean_surv_ovr
sum_risk num sum_surv FP SUM_SURV_SENS;
LENGTH LABEL $25.;
MODEL =  999999; TIME = &TIME_STOP;
LABEL = "REF_MODEL";RUN;

*EXPORT DATASET;
PROC DATASETS NOLIST;
APPEND DATA = ROC_REF_FINAL BASE = &OUTPUT FORCE;
DELETE ROC_PCT_REF SENS_OVR_REF AR_PCT_REF PRED_REF; QUIT;

*MODELS TO BE EVALUATED;
%DO HKAS = &CONTRAST_MIN %TO &CONTRAST_MAX;
DATA _NULL_;SET MODEL_LIST; CALL SYMPUT("MODEL", MODEL);
CALL SYMPUT("LABEL", COMPRESS(LABEL)); WHERE NUM = &HKAS; RUN;

PROC PHREG DATA = AR_DIST ;
WHERE &COND;
OUTPUT OUT = PRED XBETA = XBETA SURVIVAL=SURV;
MODEL &TIME*&EV(0) = &MODEL;
RUN;

DATA PRED; SET PRED; RISK = 1 - SURV; RUN;
*XBETA PCTS - ONLY WHERE &EV = .;
PROC UNIVARIATE DATA = PRED NORMAL PLOT NOPRINT;
VAR XBETA RISK;
OUTPUT OUT = AR_PCT PCTLPRE=QR_ RI_ PCTLPTS = (1 TO 99 BY 1);
WHERE &EV = .;
RUN;

*OVERALL(DENOMINATOR) - ONLY WHERE &EV = .;
PROC MEANS DATA = PRED MEAN N NOPRINT;
VAR SURV RISK;
OUTPUT OUT = SENS_OVR MEAN = MEAN_SURV_OVR MEAN_RISK_OVR N = NUM;
WHERE &EV = .;
RUN;
OPTION NONOTES;
      %DO I = 1 %TO 99;*RUN WITHIN PCTS;
      DATA _NULL_; SET AR_PCT;CALL SYMPUT("PCT", QR_&I);CALL SYMPUT("RISK",
RI_&I); RUN;
      DATA SENS_&I; SET PRED; WHERE &EV = . AND XBETA GE &PCT; RUN;
      PROC MEANS DATA = SENS_&I SUM NOPRINT;
```

```
      VAR   SURV          RISK    ; OUTPUT OUT = SENS_M_&I
      SUM = SUM_SURV_SENS SUM_RISK;
      RUN;
      DATA SENS_M_&I; SET SENS_M_&I;
      KEEP SUM_SURV_SENS SUM_RISK QR PCT RISK;
      QR = &I; PCT=&PCT; RISK=&RISK; RUN;

      DATA SPEC_&I; SET PRED; WHERE &EV = . AND XBETA LT &PCT; RUN;
      PROC MEANS DATA = SPEC_&I SUM NOPRINT;
      VAR SURV; OUTPUT OUT = SPEC_M_&I SUM = SUM_SURV;
      RUN;
      DATA SPEC_M_&I; SET SPEC_M_&I; KEEP SUM_SURV QR; QR = &I; RUN;

      DATA XXX_&I; MERGE SENS_M_&I SPEC_M_&I SENS_OVR; RUN;

      PROC DATASETS NOLIST;
      APPEND DATA = XXX_&I BASE = ROC_PCT FORCE;
      DELETE XXX_&I SPEC_M_&I SENS_M_&I SENS_&I SPEC_&I; QUIT;
      %END;*END CYCLE WITHIN PCTS;
      OPTION NOTES;

DATA ROC_PCT_FINAL; SET ROC_PCT;
FP   = SUM_SURV_SENS/NUM;
SENS = (SUM_RISK/NUM)*(1/MEAN_RISK_OVR);
SPEC = (SUM_SURV/NUM)*(1/MEAN_SURV_OVR);
I_SPEC = 1-SPEC;
KEEP QR SENS SPEC I_SPEC MODEL LABEL TIME PCT RISK mean_risk_ovr mean_surv_ovr
sum_risk num sum_surv FP SUM_SURV_SENS;
LENGTH LABEL $25.;
MODEL = &HKAS; TIME = &TIME_STOP;
LABEL = "&LABEL";RUN;

*PLOTTING DATASET;
PROC DATASETS NOLIST;
DELETE PLOT_TTT;
APPEND DATA = ROC_REF_FINAL BASE = PLOT_TTT FORCE;
APPEND DATA = ROC_PCT_FINAL BASE = PLOT_TTT FORCE;
QUIT;

**********************
plotting the roc curve;
*annotated dataset for 45° line;
data anno;
   function='move';
   xsys='1'; ysys='1';
   x=0; y=0;
   output;

   function='draw';
   xsys='1'; ysys='1';
   color='black';
   LINE = 3;
   x=100; y=100;
   output;
run;
DATA DDD; SET PLOT_TTT; RENAME SENS=Y I_SPEC=X; RUN;

goptions reset=global gunit=pct border cback=white device=win colors=(black blue
green red)
ftitle=swissb ftext=swiss htitle=6 htext=4;
TITLE1 ;
TITLE2 "ROC Curve for Reference and Extended Model";
symbol1 interpol=join color=green width=0.5 value=dot    height=0.5;
```

```
symbol2 interpol=join color=blue  width=0.5 value=circle height=0.5;
axis1 label=("1 - Specificity")  order=(0 to 1 by 0.2) length=60;
axis2 label=(a=90 "Sensitivity") order=(0 to 1 by 0.2) length=69;
legend1 ORDER=(999999 &HKAS) value=("Reference" "Extended:&LABEL") down=2
label=none shape=symbol(2,1) position=(BOTTOM RIGHT INSIDE) mode=share;
PROC GPLOT DATA = DDD;
PLOT y*x=MODEL / anno=anno haxis=axis1 vaxis=axis2 legend=legend;
WHERE MODEL IN (999999 &HKAS);
RUN; QUIT;
PROC DATASETS NOLIST; DELETE ANNO DDD;QUIT;

*EXPORT DATASET;
PROC DATASETS NOLIST;
APPEND DATA = ROC_PCT_FINAL BASE = &OUTPUT FORCE;
DELETE ROC_PCT ROC_PCT_FINAL SENS_OVR AR_PCT AR_PCT_FIN PRED PLOT_TTT; QUIT;
%END; *FINE CICLO ENTRO MODELLO;

PROC DATASETS NOLIST; DELETE AR_DIST ROC_REF_FINAL; QUIT;
%MEND;
```

3. Net Reclassification Improvement – NRI(t), clinical NRI(t), continuous NRI(t)

```
%MACRO RECLASS_ANALYSIS(DATASET, OUTPUT, ID, EV, TIME, TIME_STOP, WEIGHT, LOW_T,
HIGH_T, REF_MODEL=, CONTRAST_MIN=, CONTRAST_MAX=, COND=, COND_CLIN=, PRINT=);
```

Description of macro variables:

DATASET: name of the input SAS dataset
OUTPUT: name of the output SAS dataset containing the following quantities: the probability of being reclassified upward (EV_IMPROVED) and downward (EV_WORSENED) among the events; the probability of being reclassified downward (NONEV_IMPROVED) and upward (NONEV_WORSENED) among non-events; the overall probability of event at t-years (KM_OVR); the probability of event among those reclassified upward (RISK_UPWARDS); the probability of being reclassified upward (P_UP) and downward (P_DOWN); the overall NRI; a flag (TYPE) indicating *i)* the overall NRI (TYPE = 1); *ii)* the clinical NRI (TYPE = 2); and *iii)* the continuous NRI (TYPE = 3).
ID: variable name for ID of any given subject (numeric)
EV: variable name for the event
TIME: variable name for the follow-up time
TIME_STOP: number corresponding to time t of prediction interval
WEIGTH: number indicating how to weight NRI among cases and among non-cases (generally 0.5)
LOW_T: a number indicating the lower bound for the intermediate risk category
HIGH_T: a number indicating the upper bound for the intermediate risk category
REF_MODEL: reference model
CONTRAST_MIN: number corresponding to the first contrast model to be tested in the MODEL_LIST dataset
CONTRAST_MAX: number corresponding to the last contrast model to be tested in the MODEL_LIST dataset
COND=: run the whole analysis (including model estimate) in a subgroup of subjects, i.e. men only (optional)
PRINT=: specify YES to print the output from the PHREG procedure

```
%MACRO RECLASS_ANALYSIS(DATASET, OUTPUT, ID, EV, TIME, TIME_STOP, WEIGHT, LOW_T,
HIGH_T, REF_MODEL=, CONTRAST_MIN=, CONTRAST_MAX=, COND=, PRINT=);

PROC DATASETS NOLIST; DELETE &OUTPUT CONT_&OUTPUT CLASS_&OUTPUT CLIN_&OUTPUT;
QUIT;

DATA FAKE; SET &DATASET; &EV = .; &TIME = &TIME_STOP; RUN;
DATA BIS_&DATASET; SET &DATASET FAKE; RUN;

%LET COND_CLIN = &LOW_T<=RISK_BASE<=&HIGH_T;

*****************************
1. REFERENCE MODEL: REGRESSION;
      PROC PHREG DATA=BIS_&DATASET %if %upcase(&PRINT)^=YES %then %do; noprint
      %end;;
      MODEL &TIME*&EV(0)=&REF_MODEL;
      OUTPUT OUT = PRED_BASE SURVIVAL = SURV_BASE;
      WHERE &COND;
      RUN;
*****************************
2. KM SURV(20) OVERALL SAMPLE;
      PROC LIFETEST DATA = &DATASET OUTSURV=KM NOPRINT;
      TIME &TIME*&EV(0);
      RUN;
```

```
      PROC SORT DATA = KM; BY _CENSOR_ &TIME; RUN;
      DATA LAST_KM_OVR;
      SET KM; BY _CENSOR_ &TIME; IF LAST._CENSOR_; WHERE _CENSOR_ = 0 AND &TIME
LE &TIME_STOP;
      KEEP SURVIVAL; RENAME SURVIVAL = KM_OVR;RUN;
      PROC DELETE DATA = KM; RUN;

************************************************************
3. REGRESSION MODEL AND PREDICTED RISK IN EACH CONTRAST MODEL;
%DO CI_MOD = &CONTRAST_MIN %TO &CONTRAST_MAX;   *RUN WITHIN EACH MODEL TO
EVALUATE;
      DATA _NULL_;SET MODEL_LIST;
      CALL SYMPUT("MODEL", MODEL);
      CALL SYMPUT("LABEL", LABEL);
      WHERE NUM = &CI_MOD; RUN;

      PROC PHREG DATA=PRED_BASE %if %upcase(&PRINT)^=YES %then %do; noprint
      %end;; *REGRESSION ON PRED_BASE IS IMPORTANT TO HAVE BOTH SURVIVALS IN THE
      SAME DATASET;
      MODEL &TIME*&EV(0)=&MODEL;
      OUTPUT OUT = PRED SURVIVAL=SURV;
      WHERE &COND;
      RUN;

*******************************************************
COMPARING PREDICTED RISK FROM THE OLD AND THE NEW MODEL;
DATA PRED_FINALE;
SET PRED;
WHERE &EV = .; *IMPORTANT TO COMPARE PREDICTED RISK AT TIME T;

RISK =      1 - SURV;
RISK_BASE = 1 - SURV_BASE;

ARRAY R (*)   RISK    RISK_BASE;
ARRAY R_C (*) RISK_CL RISK_BASE_CL;

DO PPPPP = 1 TO DIM(R);
IF      R(PPPPP) LT &LOW_T  THEN R_C(PPPPP) = 1;
ELSE IF R(PPPPP) LT &HIGH_T THEN R_C(PPPPP) = 2;
ELSE                             R_C(PPPPP) = 3;
END;
*CLASS NRI;
IF RISK_CL = RISK_BASE_CL THEN MOVE = 0;
ELSE IF RISK_CL < RISK_BASE_CL THEN MOVE = -1;
ELSE IF RISK_CL > RISK_BASE_CL THEN MOVE =  1;
ELSE MOVE = 9;

MOVE_UP =   (MOVE =  1);
MOVE_DOWN = (MOVE = -1);
*CONTINUOUS NRI;

IF RISK GT RISK_BASE THEN MOVE_CONT = 1;
ELSE IF RISK LT RISK_BASE THEN MOVE_CONT = -1;
ELSE IF RISK = RISK_BASE THEN MOVE_CONT = 0;
MOVE_UP_CONT = (MOVE_CONT = 1);
MOVE_DOWN_CONT = (MOVE_CONT = -1);
RUN;

PROC SQL;
CREATE TABLE HHH AS
(SELECT T1.*, T2.MOVE, T2.MOVE_CONT, T2.RISK_BASE FROM
&DATASET AS T1 LEFT JOIN PRED_FINALE AS T2 ON
T1.&ID = T2.&ID); QUIT;
```

```
*******************************************************
A. CLASS NRI;
PROC MEANS DATA = PRED_FINALE MEAN NOPRINT;
VAR MOVE_UP MOVE_DOWN;
OUTPUT OUT = PREV_MOVE MEAN = PROP_UP PROP_DOWN;
RUN;
DATA PREV_MOVE; SET PREV_MOVE; KEEP PROP_UP PROP_DOWN; RUN;
DATA HHH_CLASS; SET HHH; WHERE &COND;RUN;

      PROC SORT DATA = HHH_CLASS; BY MOVE; RUN;
      PROC LIFETEST DATA = HHH_CLASS OUTSURV=KM NOPRINT;
      BY MOVE;
      TIME &TIME*&EV(0);
      RUN;
      PROC SORT DATA = KM; BY MOVE &TIME; RUN;
      DATA LAST_KM; KEEP MOVE SURVIVAL;
      SET KM; BY MOVE &TIME; IF LAST.MOVE;
      WHERE (_CENSOR_ = 0 AND &TIME LE &TIME_STOP) OR (SURVIVAL = 1 AND &TIME LE
&TIME_STOP);
      RUN;
      *CLASS NRI: THOSE WHO DID NOT MOVE ARE NOT OF INTEREST;
      DATA LAST_KM; SET LAST_KM; WHERE MOVE NE 0; RUN;
      PROC DELETE DATA = KM; RUN;

*******************************************************
B. CLINICAL NRI;
PROC MEANS DATA = PRED_FINALE MEAN NOPRINT;
VAR MOVE_UP MOVE_DOWN;
OUTPUT OUT = PREV_MOVE_CLIN MEAN = PROP_UP PROP_DOWN;
WHERE &COND_CLIN;
RUN;
DATA PREV_MOVE_CLIN; SET PREV_MOVE_CLIN; KEEP PROP_UP PROP_DOWN; RUN;
DATA HHH_CLINICAL; SET HHH_CLASS;
WHERE &COND_CLIN;
RUN;
*****************************
KM SURV(T) INTERMEDIATE RISK;
      PROC LIFETEST DATA = HHH_CLINICAL OUTSURV=KM NOPRINT;
      TIME &TIME*&EV(0);
      RUN;
      PROC SORT DATA = KM; BY _CENSOR_ &TIME; RUN;
      DATA LAST_KM_OVR_CLIN;
      SET KM; BY _CENSOR_ &TIME; IF LAST._CENSOR_; WHERE _CENSOR_ = 0 AND &TIME
LE &TIME_STOP;
      KEEP SURVIVAL; RENAME SURVIVAL = KM_OVR;RUN;
      PROC DELETE DATA = KM; RUN;
***********************
KM SURV(T) BY MOVE;
      PROC SORT DATA = HHH_CLINICAL; BY MOVE; RUN;
      PROC LIFETEST DATA = HHH_CLINICAL OUTSURV=KM NOPRINT;
      BY MOVE;
      TIME &TIME*&EV(0);
      RUN;
      PROC SORT DATA = KM; BY MOVE &TIME; RUN;
      DATA LAST_KM_CLIN; KEEP MOVE SURVIVAL;
      SET KM; BY MOVE &TIME; IF LAST.MOVE;
      WHERE (_CENSOR_ = 0 AND &TIME LE &TIME_STOP) OR (SURVIVAL = 1 AND &TIME LE
&TIME_STOP);
      RUN;
      *CLINICAL NRI: THOSE WHO DID NOT MOVE ARE NOT OF INTEREST;
      DATA LAST_KM_CLIN; SET LAST_KM_CLIN; WHERE MOVE NE 0; RUN;
    PROC DELETE DATA = KM; RUN;
```

```
******************
C. CONTINUOUS NRI;
PROC MEANS DATA = PRED_FINALE MEAN NOPRINT;
VAR MOVE_UP_CONT MOVE_DOWN_CONT;
OUTPUT OUT = PREV_MOVE_CONT MEAN = PROP_UP PROP_DOWN;
RUN;
DATA PREV_MOVE_CONT; SET PREV_MOVE_CONT; KEEP PROP_UP PROP_DOWN; RUN;

DATA HHH_CONT;
SET HHH;
WHERE &COND;
RUN;
***********************
KM SURV(T) BY MOVE;
      PROC SORT DATA = HHH_CONT; BY MOVE_CONT; RUN;
      PROC LIFETEST DATA = HHH_CONT OUTSURV=KM NOPRINT;
      BY MOVE_CONT;
      TIME &TIME*&EV(0);
      RUN;
      PROC SORT DATA = KM; BY MOVE_CONT &TIME; RUN;
      DATA LAST_KM_CONT; KEEP MOVE_CONT SURVIVAL;
      SET KM; BY MOVE_CONT &TIME; IF LAST.MOVE_CONT;
      WHERE (_CENSOR_ = 0 AND &TIME LE &TIME_STOP) OR (SURVIVAL = 1 AND &TIME LE
&TIME_STOP);
      RUN;
      *CONTINUOUS NRI: THOSE WHO DID NOT MOVE ARE NOT OF INTEREST;
      DATA LAST_KM_CONT; SET LAST_KM_CONT; WHERE MOVE_CONT NE 0; RUN;
    PROC DELETE DATA = KM; RUN;

************************************
5. IML FOR NRI, CLIN NRI AND CONT NRI;
      PROC IML;
      *NRI;
      USE LAST_KM_OVR;
      READ ALL INTO KM_OVR;
      CLOSE LAST_KM_OVR;
      USE LAST_KM;
      READ ALL INTO KM;
      CLOSE LAST_KM;
      USE PREV_MOVE;
      READ ALL INTO PROP;
      CLOSE PREV_MOVE;

      IF PROP[1,2] = 0 THEN DO; *ONLY UPWARD MOVEMENTS;
            NRI_UP   = (1-KM[1,2])#PROP[1,1]#1/(1-KM_OVR[1,1]);
            NRI_DOWN = - KM[1,2]#PROP[1,1]#1/KM_OVR[1,1];
            NRI = &WEIGHT*NRI_UP + (1-&WEIGHT)*NRI_DOWN;
            EV_IMPROVED = (1-KM[2,2])#PROP[1,1]#1/(1-KM_OVR[1,1]);
          EV_WORSENED = .;
          NONEV_IMPROVED = .;
          NONEV_WORSENED = KM[2,2]#PROP[1,1]#1/KM_OVR[1,1];
          KM_OVR       = KM_OVR[1,1];
          RISK_UPWARDS = (1-KM[2,2]);
      END;
      ELSE IF PROP[1,1] = 0 THEN DO; *ONLY DOWNWARD MOVEMENTS;
            NRI_UP   = - (1-KM[1,2])#PROP[1,2]#1/(1-KM_OVR[1,1]);
            NRI_DOWN = KM[1,2]#PROP[1,2]#1/KM_OVR[1,1];
            NRI = &WEIGHT*NRI_UP + (1-&WEIGHT)*NRI_DOWN;
            EV_IMPROVED = .;
            EV_WORSENED = (1-KM[1,2])#PROP[1,2]#1/(1-KM_OVR[1,1]);
            NONEV_IMPROVED = KM[1,2]#PROP[1,2]#1/KM_OVR[1,1];
            NONEV_WORSENED = .;
            KM_OVR       = KM_OVR[1,1];
```

```
            RISK_UPWARDS = .;
      END;
      ELSE DO;*BOTH UPWARD AND DOWNWARD MOVEMENTS;
            NRI_UP   = (1-KM[2,2])#PROP[1,1]#1/(1-KM_OVR[1,1]) - (1-
KM[1,2])#PROP[1,2]#1/(1-KM_OVR[1,1]);
            NRI_DOWN = KM[1,2]#PROP[1,2]#1/KM_OVR[1,1] -
KM[2,2]#PROP[1,1]#1/KM_OVR[1,1];
            NRI = &WEIGHT*NRI_UP + (1-&WEIGHT)*NRI_DOWN;
            EV_IMPROVED = (1-KM[2,2])#PROP[1,1]#1/(1-KM_OVR[1,1]);
            EV_WORSENED = (1-KM[1,2])#PROP[1,2]#1/(1-KM_OVR[1,1]);
            NONEV_IMPROVED = KM[1,2]#PROP[1,2]#1/KM_OVR[1,1];
            NONEV_WORSENED = KM[2,2]#PROP[1,1]#1/KM_OVR[1,1];
            KM_OVR       = KM_OVR[1,1];
            RISK_UPWARDS = (1-KM[2,2]);
    END;
      TYPE = 1;
    jauct_fin =
shape((TYPE||NRI||EV_IMPROVED||EV_WORSENED||NONEV_IMPROVED||NONEV_WORSENED||KM_O
VR||RISK_UPWARDS||PROP[1,1]||PROP[1,2]),1,10);

      create NRI from jauct_fin[colname={TYPE NRI EV_IMPROVED EV_WORSENED
NONEV_IMPROVED NONEV_WORSENED KM_OVR RISK_UPWARDS P_UP P_DOWN}];
      append from jauct_fin;
      close NRI;
      *CLINICAL NRI;
      USE LAST_KM_OVR_CLIN;
      READ ALL INTO KM_OVR;
      CLOSE LAST_KM_OVR_CLIN;
      USE LAST_KM_CLIN;
      READ ALL INTO KM;
      CLOSE LAST_KM_CLIN;
      USE PREV_MOVE_CLIN;
      READ ALL INTO PROP;
      CLOSE PREV_MOVE_CLIN;

      IF PROP[1,2] = 0 THEN DO; *ONLY UPWARD MOVEMENTS;
            NRI_UP   = (1-KM[1,2])#PROP[1,1]#1/(1-KM_OVR[1,1]);
            NRI_DOWN = - KM[1,2]#PROP[1,1]#1/KM_OVR[1,1];
            NRI = &WEIGHT*NRI_UP + (1-&WEIGHT)*NRI_DOWN;
            EV_IMPROVED = (1-KM[2,2])#PROP[1,1]#1/(1-KM_OVR[1,1]);
          EV_WORSENED = .;
          NONEV_IMPROVED = .;
          NONEV_WORSENED = KM[2,2]#PROP[1,1]#1/KM_OVR[1,1];
          KM_OVR       = KM_OVR[1,1];
          RISK_UPWARDS = (1-KM[2,2]);
      END;
      ELSE IF PROP[1,1] = 0 THEN DO; *ONLY DOWNWARD MOVEMENTS;
            NRI_UP   = - (1-KM[1,2])#PROP[1,2]#1/(1-KM_OVR[1,1]);
            NRI_DOWN = KM[1,2]#PROP[1,2]#1/KM_OVR[1,1];
            NRI = &WEIGHT*NRI_UP + (1-&WEIGHT)*NRI_DOWN;
            EV_IMPROVED = .;
            EV_WORSENED = (1-KM[1,2])#PROP[1,2]#1/(1-KM_OVR[1,1]);
            NONEV_IMPROVED = KM[1,2]#PROP[1,2]#1/KM_OVR[1,1];
            NONEV_WORSENED = .;
            KM_OVR       = KM_OVR[1,1];
            RISK_UPWARDS = .;
      END;
      ELSE DO;*BOTH UPWARD AND DOWNWARD MOVEMENTS;
            NRI_UP   = (1-KM[2,2])#PROP[1,1]#1/(1-KM_OVR[1,1]) - (1-
KM[1,2])#PROP[1,2]#1/(1-KM_OVR[1,1]);
            NRI_DOWN = KM[1,2]#PROP[1,2]#1/KM_OVR[1,1] -
KM[2,2]#PROP[1,1]#1/KM_OVR[1,1];
            NRI = &WEIGHT*NRI_UP + (1-&WEIGHT)*NRI_DOWN;
```

```
            EV_IMPROVED = (1-KM[2,2])#PROP[1,1]#1/(1-KM_OVR[1,1]);
            EV_WORSENED = (1-KM[1,2])#PROP[1,2]#1/(1-KM_OVR[1,1]);
            NONEV_IMPROVED = KM[1,2]#PROP[1,2]#1/KM_OVR[1,1];
            NONEV_WORSENED = KM[2,2]#PROP[1,1]#1/KM_OVR[1,1];
            KM_OVR       = KM_OVR[1,1];
            RISK_UPWARDS = (1-KM[2,2]);
    END;
      TYPE = 2;
    CLIN_jauct_fin =
shape((TYPE||NRI||EV_IMPROVED||EV_WORSENED||NONEV_IMPROVED||NONEV_WORSENED||KM_O
VR||RISK_UPWARDS||PROP[1,1]||PROP[1,2]),1,10);
      create CLINICAL_NRI from CLIN_jauct_fin[colname={TYPE NRI EV_IMPROVED
EV_WORSENED NONEV_IMPROVED NONEV_WORSENED KM_OVR RISK_UPWARDS P_UP P_DOWN}];
      append from CLIN_jauct_fin;
      close CLINICAL_NRI;
      *CONTINUOUS NRI;
      USE LAST_KM_OVR;
      READ ALL INTO KM_OVR;
      CLOSE LAST_KM_OVR;
      USE LAST_KM_CONT;
      READ ALL INTO KM;
      CLOSE LAST_KM_CONT;
      USE PREV_MOVE_CONT;
      READ ALL INTO PROP;
      CLOSE PREV_MOVE_CONT;
      NRI_UP   = (1-KM[2,2])#PROP[1,1]#1/(1-KM_OVR[1,1]) - (1-
KM[1,2])#PROP[1,2]#1/(1-KM_OVR[1,1]);
      NRI_DOWN = KM[1,2]#PROP[1,2]#1/KM_OVR[1,1] -
KM[2,2]#PROP[1,1]#1/KM_OVR[1,1];
      NRI = &WEIGHT*NRI_UP + (1-&WEIGHT)*NRI_DOWN;
      EV_IMPROVED = (1-KM[2,2])#PROP[1,1]#1/(1-KM_OVR[1,1]);
      EV_WORSENED = (1-KM[1,2])#PROP[1,2]#1/(1-KM_OVR[1,1]);
      NONEV_IMPROVED = KM[1,2]#PROP[1,2]#1/KM_OVR[1,1];
      NONEV_WORSENED = KM[2,2]#PROP[1,1]#1/KM_OVR[1,1];
      KM_OVR       = KM_OVR[1,1];
      RISK_UPWARDS = (1-KM[2,2]);
      TYPE = 3;
    CONT_jauct_fin =
shape((TYPE||NRI||EV_IMPROVED||EV_WORSENED||NONEV_IMPROVED||NONEV_WORSENED||KM_O
VR||RISK_UPWARDS||PROP[1,1]||PROP[1,2]),1,10);
      create CONTINUOUS_NRI from CONT_jauct_fin[colname={TYPE NRI EV_IMPROVED
EV_WORSENED NONEV_IMPROVED NONEV_WORSENED KM_OVR RISK_UPWARDS P_UP P_DOWN}];
      append from CONT_jauct_fin;
      close CONTINUOUS_NRI;
      QUIT; *END IML;
*OUTPUT: DATASET WITH DETAILS FOR EACH INDICATOR;
      DATA NRI;SET NRI;                        NRI  = 2*NRI; LENGTH EVENT LABEL
$25.;MODEL = &CI_MOD; LABEL = "&LABEL"; EVENT = "&EV"; RUN;
      DATA CLINICAL_NRI;  SET CLINICAL_NRI;   NRI  = 2*NRI; LENGTH EVENT LABEL
$25.;MODEL = &CI_MOD; LABEL = "&LABEL"; EVENT = "&EV"; RUN;
      DATA CONTINUOUS_NRI; SET CONTINUOUS_NRI;NRI  = 2*NRI; LENGTH EVENT LABEL
$25.;MODEL = &CI_MOD; LABEL = "&LABEL"; EVENT = "&EV"; RUN;
PROC DATASETS NOLIST;
APPEND DATA = NRI BASE = &OUTPUT FORCE;
APPEND DATA = CLINICAL_NRI  BASE = &OUTPUT FORCE;
APPEND DATA = CONTINUOUS_NRI  BASE = &OUTPUT FORCE;
DELETE OUTPUT PRED PRED_FINALE PREV_MOVE PREV_MOVE_CLIN PREV_MOVE_CONT
HHH HHH_CLASS HHH_CONT HHH_CLINICAL
LAST_KM LAST_KM_CLIN LAST_KM_OVR_CLIN LAST_KM_CONT
NRI CLINICAL_NRI CONTINUOUS_NRI; QUIT;
%END; *END RUN FOR A GIVEN MODEL;
PROC DATASETS NOLIST; DELETE PRED_BASE LAST_KM_OVR FAKE BIS_&DATASET; QUIT;
%MEND;
```

4. Bootstrapped CI for AUC(t), Δ-AUC(t), IDI(t), NRI(t)

```
%MACRO BOOT_ANALYSIS(DATASET, OUTPUT, EV, TIME, TIME_STOP, WEIGHT, LOW_T,
HIGH_T, REF_MODEL=, CONTRAST_MIN=, CONTRAST_MAX=, COND= , NUMBOOT=, SEED=);
```

Description of macro variables:

DATASET: name of the input SAS dataset
OUTPUT: name of the output SAS dataset containing the following quantities: AUC, R, average sensitivity (ISt) and average 1-specificity (IPt) for each contrast model, as well as for the reference model; the difference in AUC between each contrast model and the reference; IDI; NRI, clinical NRI and continuous NRI. The variable BOOT indicates the bootstrapped cycle; BOOT = 0 is the analysis for the original dataset
EV: variable name for the event
TIME: variable name for the follow-up time
TIME_STOP: number corresponding to time t of prediction interval
WEIGTH: number indicating how to weight NRI among cases and among non-cases (generally 0.5)
LOW_T: a number indicating the lower bound for the intermediate risk category
HIGH_T: a number indicating the upper bound for the intermediate risk category
REF_MODEL: reference model
CONTRAST_MIN: number corresponding to the first contrast model to be tested in the MODEL_LIST dataset
CONTRAST_MAX: number corresponding to the last contrast model to be tested in the MODEL_LIST dataset
COND=: run the whole analysis (including model estimate) in a subgroup of subjects, i.e. men only (optional)
NUMBOOT: the number of bootstrapped samples. If NUMBOOT = 0 the analysis is run on the original dataset only (no bootstrap). The output from the PHREG procedure will be printed in this case.
SEED: a number indicating the seed for bootstrapping (need to be changed in every run)

```
%MACRO BOOT_ANALYSIS(DATASET, OUTPUT, EV, TIME, TIME_STOP, WEIGHT, LOW_T,
HIGH_T, REF_MODEL=, CONTRAST_MIN=, CONTRAST_MAX=, COND= , NUMBOOT=, SEED=);
PROC DATASETS NOLIST; DELETE &OUTPUT; QUIT;

*ORIGINAL DATASET;
DATA FAKE_BIS;
SET &DATASET;
WHERE &COND;
_SAMPLE_ = 0;
RUN;

*NO BOOTSTRAPP;
%IF &NUMBOOT = 0 %THEN %DO;
DATA BOOT; SET FAKE_BIS; RUN;
PROC DATASETS NOLIST; DELETE FAKE_BIS; QUIT;
%LET PRINT = YES;
%END;

*BOOTSTRAPP;
%IF &NUMBOOT > 0 %THEN %DO;
%LET PRINT = NO;

DATA FAKE; SET &DATASET NOBS=NOBS; WHERE &COND;
BOOT_ID = _N_; CALL SYMPUT ("MAX", _N_);RUN;

DATA BOOT_J (DROP = HHH);
```

```
 DO _SAMPLE_ = 1 TO &NUMBOOT;
  *DO HHH = 1 TO &M;
  *ID = INT(&M*RANUNI(-1)+1);
   DO HHH = 1 TO &MAX;
   BOOT_ID2 = INT(&MAX*RANUNI(&SEED)+1);
   SET FAKE POINT = BOOT_ID2;
   IF _ERROR_ THEN ABORT;
   OUTPUT;
  END;
 END;
 STOP;
RUN;

DATA BOOT; SET FAKE_BIS BOOT_J; RUN;
PROC DATASETS NOLIST; DELETE BOOT_J FAKE FAKE_BIS;QUIT;
%END;

%PUT &PRINT;

%DO JFK = 0 %TO &NUMBOOT; *BOOTSTRAP CYCLE. 0 = ORIGINAL DATASET;
*BOOT_ID TO BE USED IN THE NRI MACRO;
DATA BOOT_&JFK; SET BOOT;
WHERE _SAMPLE_ = &JFK;
BOOT_ID = _N_; RUN;

************************
2. MODEL DISCRIMINATION;
%DISCR_ANALYSIS(BOOT_&JFK, DISCRIMINATION_&JFK, &EV, &TIME, &TIME_STOP,
REF_MODEL=&REF_MODEL, CONTRAST_MIN=&CONTRAST_MIN, CONTRAST_MAX=&CONTRAST_MAX,
COND=&COND, PRINT=&PRINT);
************************
3. RECLASSIFICATION;
%RECLASS_ANALYSIS(BOOT_&JFK, RECL_AN_&JFK, BOOT_ID, &EV, &TIME, &TIME_STOP,
&WEIGHT, &LOW_T, &HIGH_T, REF_MODEL=&REF_MODEL, CONTRAST_MIN=&CONTRAST_MIN,
CONTRAST_MAX=&CONTRAST_MAX, COND=&COND, PRINT=&PRINT);

DATA NRI_&JFK;  SET RECL_AN_&JFK; WHERE TYPE = 1; KEEP NRI EVENT LABEL MODEL;
RENAME NRI = CLASS_NRI; RUN;
DATA CLIN_&JFK; SET RECL_AN_&JFK; WHERE TYPE = 2; KEEP NRI EVENT LABEL MODEL;
RENAME NRI = CLIN_NRI; RUN;
DATA CONT_&JFK; SET RECL_AN_&JFK; WHERE TYPE = 3; KEEP NRI EVENT LABEL MODEL;
RENAME NRI = CONT_NRI; RUN;

DATA RRR_FINALE;
MERGE DISCRIMINATION_&JFK  NRI_&JFK CLIN_&JFK CONT_&JFK;
BOOT = &JFK;
RUN;

PROC DATASETS NOLIST;
APPEND DATA = RRR_FINALE BASE = &OUTPUT FORCE;
DELETE DISCRIMINATION_&JFK RRR_FINALE NRI_&JFK CLIN_&JFK CONT_&JFK BOOT_&JFK
RECL_AN_&JFK; QUIT;

%END; *END BOOTSTRAP;
PROC DATASETS NOLIST; DELETE FAKE BOOT; QUIT;
%MEND;
```

5A. Internal validation analysis: over-optimism.

```
%MACRO OPTIMISM_AUC(DATASET, OUTPUT, ID, EV, TIME, TIME_STOP, MIN_MOD, MAX_MOD,
NUMBOOT=, SEED=, COND=);
```

Description of macro variables:

DATASET: name of the input SAS dataset
OUTPUT: name of the output SAS dataset containing the following quantities: the AUC on the bootstrapped dataset, the AUC on the original dataset, and the optimism as the difference among the two values
ID: variable name for ID of any given subject (numeric)
EV: variable name for the event
TIME: variable name for the follow-up time
TIME_STOP: number corresponding to time t of prediction interval
MIN_MOD: number corresponding to the first model to be tested in the MODEL_LIST dataset
MAX_MOD: number corresponding to the last model to be tested in the MODEL_LIST dataset
NUMBOOT=: the number of bootstrapped samples.
SEED=: a number indicating the seed for bootstrapping (need to be changed in every run)
COND=: run the whole analysis (including model estimate) in a subgroup of subjects, i.e. men only (optional)

```
*******************************************
5A. INTERNAL VALIDATION: OVER-OPTIMISM IN AUC;
%MACRO OPTIMISM_AUC(DATASET, OUTPUT_OPT, ID, EV, TIME, TIME_STOP, MIN_MOD,
MAX_MOD, NUMBOOT=, SEED=, COND=);
PROC DATASETS NOLIST; DELETE &OUTPUT_OPT; QUIT;

DATA FAKE; SET &DATASET NOBS=NOBS; WHERE &COND;
BOOT_ID = _N_; CALL SYMPUT ("MAX", _N_);RUN;*REPLY OF THE ORIGINAL DATASET;

*********************************
BOOTSTRAPPING THE ORIGINAL DATASET;

DATA BOOT (DROP = HHH);
 DO BOOT = 1 TO &NUMBOOT;
   DO HHH = 1 TO &MAX;
   BOOT_ID = INT(&MAX*RANUNI(&SEED)+1);
   SET FAKE POINT = BOOT_ID;
   IF _ERROR_ THEN ABORT;
   OUTPUT;
  END;
 END;
 STOP;
RUN;

%DO CI_MOD = &MIN_MOD %TO &MAX_MOD;    *RUN WITHIN EACH MODEL TO EVALUATE;
DATA _NULL_;SET MODEL_LIST;
CALL SYMPUT("MODEL", MODEL);
CALL SYMPUT("LABEL", LABEL);
WHERE NUM = &CI_MOD; RUN;

      %DO JFK = 1 %TO &NUMBOOT; *RUN WITHIN EACH BOOTSTRAPPED DATASET;

      DATA JLKJ; SET BOOT; WHERE BOOT = &JFK; AUC_VVV = 1;RUN;
*BOOTSTRAP SAMPLE;
      DATA JLKJ; SET JLKJ; ID_BOOT = _N_; RUN;
```

```
      DATA FAKE_BIS; SET JLKJ; &EV = .; &TIME = &TIME_STOP; AUC_VVV = 2; RUN;
*BOOTSTRAP REPLICATE;
      DATA FAKE_TER; SET FAKE; &EV = .; &TIME = &TIME_STOP; AUC_VVV = 3; RUN;
*ORIGINAL REPLICATE;

      DATA BIS_BOOTSTRAPP; SET JLKJ FAKE_BIS FAKE_TER; RUN;

*********************************************************************
AUC IN THE BOOTSTRAPPED DATASET AND SAME MODEL IN THE ORIGINAL DATASET;
      PROC PHREG DATA=BIS_BOOTSTRAPP NOPRINT;
      MODEL &TIME*&EV(0)=&MODEL;
      OUTPUT OUT = PRED XBETA = XBETA SURVIVAL=SURV;
      RUN;
*******************************
AUC IN THE BOOTSTRAPPED DATASET;
      DATA PRED_BOOT; SET PRED; KEEP ID_BOOT XBETA SURV; WHERE &EV = . AND
AUC_VVV = 2; RUN;
      PROC SORT DATA = PRED_BOOT; BY ID_BOOT; RUN;
      DATA PRED_BOOT; SET PRED_BOOT; BY ID_BOOT;IF LAST.ID_BOOT; RUN;
      PROC SORT DATA = PRED_BOOT; BY XBETA; RUN;
      DATA PRED_BOOT; SET PRED_BOOT; DROP ID_BOOT; RUN;

      PROC IML;
      USE PRED_BOOT;
      READ ALL INTO Z;
      CLOSE PRED_BOOT;

      N=NROW(Z);
      E1_StU_StV=0;
    EStZ=0;

        do i=1 to n;
            EStZ=EStZ+z[i,2];

            l_survobs = j(n,1,(1-z[i,2]));
            smaller   = (z[,1]<z[i,1]);
            equal     = (z[,1]=z[i,1]);

            smaller[i:n]=0;
            equal[i:n]=0;


      E1_StU_StV=E1_StU_StV+sum(l_survobs#z[,2]#smaller)+0.5*sum(l_survobs#z[,2]
#equal);

        end;
            E1_StU_StV =  E1_StU_StV/(n**2);
        EStZ=EStZ/n;
        auct_boot=E1_StU_StV/(EStZ*(1-EStZ));

            create AUC_BOOTSTRAPP from auct_boot[colname={AUC_BOOT}];
            append from auct_boot;
            close AUC_BOOTSTRAPP;
            QUIT;
PROC DATASETS NOLIST; DELETE PRED_BOOT; QUIT;

*************************************
AUC FROZEN MODEL ON THE ORIGINAL DATASET;
      DATA PRED_ORIG; SET PRED; KEEP &ID XBETA; WHERE &EV = . AND AUC_VVV = 3;
RUN;

      PROC SQL;
      CREATE TABLE FF_&JFK AS (SELECT T1.XBETA, T2.* FROM
```

```
      PRED_ORIG AS T1 LEFT JOIN FAKE AS T2 ON
      T1.&ID = T2.&ID); QUIT;

      DATA FAKE_QUATER; SET FF_&JFK; &EV = .; &TIME = &TIME_STOP; RUN;
      DATA FF_BIS_&JFK; SET FF_&JFK FAKE_QUATER;RUN;

      PROC PHREG DATA=FF_BIS_&JFK NOPRINT;
      MODEL &TIME*&EV(0) = XBETA;
      OUTPUT OUT = PRED_ORIG_XBETA XBETA = XBETA_ SURVIVAL=SURV;
      RUN;
      DATA PRED_ORIG_XBETA; SET PRED_ORIG_XBETA; KEEP XBETA_ SURV; WHERE &EV =
.; RUN;
      PROC SORT DATA = PRED_ORIG_XBETA; BY XBETA_; RUN;

      PROC IML;
      USE PRED_ORIG_XBETA;
      READ ALL INTO Z;
      CLOSE PRED_ORIG_XBETA;

      N=NROW(Z);

      E1_StU_StV=0;
    EStZ=0;

        do i=1 to n;
                  EStZ=EStZ+z[i,2];

            l_survobs = j(n,1,(1-z[i,2]));
            smaller   = (z[,1]<z[i,1]);
            equal     = (z[,1]=z[i,1]);

            smaller[i:n]=0;
            equal[i:n]=0;


      E1_StU_StV=E1_StU_StV+sum(l_survobs#z[,2]#smaller)+0.5*sum(l_survobs#z[,2]
#equal);
        end;
            E1_StU_StV =  E1_StU_StV/(n**2);
        EStZ=EStZ/n;
        auct_orig=E1_StU_StV/(EStZ*(1-EStZ));

            create AUC_BOOTSTRAPP_ORIG from auct_orig[colname={AUC_ORIG}];
            append from auct_orig;
            close AUC_BOOTSTRAPP_ORIG;
            QUIT;

DATA AUC_BOOTSTRAPP; MERGE AUC_BOOTSTRAPP AUC_BOOTSTRAPP_ORIG;
LENGTH EVENT LABEL $25.;
LABEL = "&LABEL"; MODEL = &CI_MOD; EVENT = "&EV";
BOOT = &JFK; OPT = AUC_BOOT - AUC_ORIG; RUN;

PROC DATASETS NOLIST; APPEND DATA = AUC_BOOTSTRAPP BASE = &OUTPUT_OPT FORCE;
DELETE AUC_BOOT AUC_ORIG PRED PRED_ORIG_XBETA PRED_ORIG AUC_BOOTSTRAPP
AUC_BOOTSTRAPP_ORIG
FF_&JFK FF_BIS_&JFK FAKE_QUATER BIS_BOOTSTRAPP JLKJ FAKE_BIS FAKE_TER; QUIT;
      %END;      *END BOOTSTRAPPED DATASET;
%END;      *END MODEL;
PROC DATASETS NOLIST; DELETE FAKE BOOT; QUIT;
%MEND;
```

5B. Internal validation analysis: calibration slope.

```
%MACRO CAL_SLOPE(DATASET, OUTPUT, ID, EV, TIME, MIN_MOD, MAX_MOD,  NUMBOOT=,
SEED=, COND=);
```

Description of macro variables:

DATASET: name of the input SAS dataset
OUTPUT: name of the output SAS dataset containing the following quantities: the estimate of the model on the bootstrapped dataset (ESTIMATE) (the calibration slope is the mean of the variable ESTIMATE)
ID: variable name for ID of any given subject (numeric)
EV: variable name for the event
TIME: variable name for the follow-up time
MIN_MOD: number corresponding to the first model to be tested in the MODEL_LIST dataset
MAX_MOD: number corresponding to the last model to be tested in the MODEL_LIST dataset
NUMBOOT=: the number of bootstrapped samples.
SEED=: a number indicating the seed for bootstrapping (need to be changed in every run)
COND=: run the whole analysis (including model estimate) in a subgroup of subjects, i.e. men only (optional)

```
%MACRO CAL_SLOPE(DATASET, OUTPUT, ID, EV, TIME, MIN_MOD, MAX_MOD,  NUMBOOT=,
SEED=, COND=);
PROC DATASETS NOLIST; DELETE &OUTPUT; QUIT;

DATA FAKE; SET &DATASET NOBS=NOBS; WHERE &COND; BOOT_ID = _N_; CALL SYMPUT
("MAX", _N_);RUN;

**************************************
BOOTSTRAPPING FROM THE ORIGINAL DATASET;
DATA BOOT (DROP = HHH);
 DO BOOT = 1 TO &NUMBOOT;
   DO HHH = 1 TO &MAX;
   BOOT_ID = INT(&MAX*RANUNI(&SEED)+1);
   SET FAKE POINT = BOOT_ID;
   IF _ERROR_ THEN ABORT;
   OUTPUT;
  END;
 END;
 STOP;
RUN;

%DO CI_MOD = &MIN_MOD %TO &MAX_MOD;    *RUN WITHIN EACH MODEL TO EVALUATE;
DATA _NULL_;SET MODEL_LIST;
CALL SYMPUT("MODEL", MODEL);
CALL SYMPUT("LABEL", LABEL);
WHERE NUM = &CI_MOD; RUN;
      %DO JFK = 1 %TO &NUMBOOT; *RUN WITHIN EACH BOOTSTRAPPED SAMPLE;
*************************************************************************
REGRESSION ON BOOTSTRAPPED DATASET AND FREEZE THE MODEL FOR THE ORIGINAL
DATASET;
      DATA JLKJ; SET BOOT; WHERE BOOT = &JFK;                RUN;
      DATA FAKE_TER; SET FAKE; &EV = .; &TIME = .;           RUN;
      DATA BIS_BOOTSTRAPP; SET JLKJ FAKE_TER; RUN;
      PROC PHREG DATA=BIS_BOOTSTRAPP NOPRINT;
      MODEL &TIME*&EV(0)=&MODEL;
      OUTPUT OUT = PRED_BOOT XBETA = XBETA_BOOT SURVIVAL=SURV;
      RUN;
```

```
***************************
CALIBRATION SLOPE FOR XBETA;
      DATA SLOPE; SET PRED_BOOT; WHERE &EV = . ; KEEP &ID XBETA_BOOT; RUN;
      PROC SORT DATA = SLOPE; BY &ID; RUN;
      DATA PRED_TOTALE;
      MERGE FAKE (IN=AA) SLOPE (IN = BB); BY &ID;
      IF AA AND BB;
      RUN;
      PROC PHREG DATA=PRED_TOTALE;*NOPRINT;
      MODEL &TIME*&EV(0)= XBETA_BOOT;
      ODS OUTPUT PARAMETERESTIMATES = PAR;
      RUN;
      DATA PAR; SET PAR; DROP  Parameter  Label DF;
      LENGTH EVENT LABEL_EV $25.;LABEL_EV = "&LABEL"; MODEL = &CI_MOD; EVENT =
"&EV";
      BOOT = &JFK;
      RUN;
      PROC DATASETS NOLIST; APPEND DATA = PAR BASE = &OUTPUT FORCE;
      DELETE PAR SLOPE FAKE_TER JLKJ BIS_BOOTSTRAPP PRED_BOOT PRED_TOTALE; QUIT;
      %END;      *END BOOTSTRAPP;
%END; *END MODEL;
%MEND;
```