Reviewer Comments

Reviewer 1 (Anonymous)

Basic reporting

Experimental design

Validity of the findings

Comments for the author

A methodology to construct a point system to predict cardiovascular disease considering repeated measures of risk factors

The paper describes the development of a prediction model for cardiovascular disease that aims to take into account the variability if risk factors over time. The paper does not provide enough information for the reader to understand what has been done and why.

Major comments:
Although this is a methodology paper, the model is not specified and each step of the model development and validation is not shown. This information is necessary.

We have now specified all the models used to construct this new method (together with their references) at the beginning of Methods, in addition to all the steps of the application of the method to the simulated data set.

It is mentioned that the model predicts risk of CVD within 2 years from baseline. However, most clinical guidelines on CVD prevention use 10y predicted risk. As the model is not specified it is not clear as to why this short risk prediction period has been chosen.

We have now explained this issue in *Simulation on a data set* (Methods).

Use after validation: A simulation is described in which the CVD risk of a patient is reduced from a series of interventions. It is not mentioned what these

interventions are and where the expected effect sizes of the risk factor reductions are taken from.

We have explained this issue in the supplementary material (Other S2).

Figure 1: It would be helpful if the information could be provided in a table instead of as screenshots, and shown with information on how the scores were assigned, possibly in combination with the corresponding table.

We have indicated why we have used a figure (Other S2) and the method to construct a scoring system through a multivariate model (*Points systems in the Framingham Heart Study* in Methods).

Reviewer 2 (Juan Carlos Bazo-Alvarez)

Basic reporting

Abstract & Title

• Methodology is not the best word for describing your purpose. I suggest "method" or "procedure".

We have changed methodology as suggested, both in the title and in the abstract, and elsewhere.

Comment [C1]: Okay. However, the new versi of this manuscript does not have title and abstrac so I did not can review them.

• Your study proposes a new method for building a new type of CVD risk score, but not a new CVD risk score itself (or something equivalent: point system). This is not clear enough until the end of the manuscript. I suggest you clarify it from the beginning (title/abstract). This helps reader to keep his (her) expectations correctly balanced.

We have clarified this issue in the relevant places applying the changes indicated below.

Comment [C2]: The new version of this manuscript does not have title and abstract, so I c not can review them.

• Validation and utilization are big words in the world of CVD risk scores. Usually, validation implies to contrast your new tool (the product of your new procedure, your point system) against real longitudinal data. Utilization implies the application of this "validated tool". You do not have results about any of them. I suggest reconsider the use of these words. Some options you have: internal

validation, statistical validation by simulation, explanation of potential utilization. The idea is to capture your target audience since the beginning: investigators who are searching new statistical procedures for optimizing existing CVD risk scores.

We thank you for the suggestion and have changed these terms throughout the text.

Comment [C3]: OK.

Introduction

• I suggest you show strength and weakness of statistical methods behind current CVD scores, referring also other relevant scores (e.g. Reynolds and WHO). Remark how, in any of them, information from time-dependent variability of risk factors has not been included in estimates. After that, explain that your new method fills this gap.

• In practice, final users of CVD risk scores (charts users) do not have problems with "the accuracy of the estimation of the probability of CVD…", because they take their decisions considering cut-off point recommended by guidelines (e.g. people under 10 years high-risk). Actually, this is a very thick estimate of risk. The present challenge is to ensure that people under this classification –and usually under preventive medication - are under real risk. In other words, the current problem does not end in the accuracy of statistical model; it really ends in the improvement of guidelines criteria and final clinical decisions on the field. I suggest that you analyze this idea for an edition of your second paragraph of Introduction.

We have written a new paragraph in the introduction to explain both the strengths and weaknesses of the current cardiovascular risk scores, and the relation between clinical guidelines and cardiovascular risk scores.

Comment [C4]: Please, read my new commen in the last paragraph of Introduction.

• At the end of Introduction, in the moment of aims exposition, I suggest to write the most clear you can about real scope of your project: to show viability and properties of a new methodological alternative for constructing CVD risk scores. As I said before, try to do not create over-expectations on readers; for example, promising validation of a final tool.

We have applied the indicated change, and in the abstract as well.

Comment [C5]: I did not see any abstract in th last version of your manuscript.

Experimental design

Methods

• Big absence: a detailed description of dataset features. It is completely simulated? Which risk factors and outcomes are you taking into account? Consider that CVD risk scores do not differ only in population baseline or statistical method, they differ in nature of risk factors and outcomes, and how they have been handled. I suggest include a Dataset section in the manuscript.

We have now included a section to explain how we simulated our data set and the mathematical equations have been placed in the supplementary material.

Comment [C6]: Correct, I have seen new information about it in a section called "Simulatio on a data set". However, I did not see exactly whi CVD outcomes are considering for simulation (exa definitions of chosen outcomes). For example, yo listed -in the Introduction - CVD outcomes used fo constructing previous CVD scores (e.g. morbidity and mortality with coronary heart disorders, mortality from coronary heart disorders, cardiovascular morbidity and mortality, or just cardiovascular mortality). I would expect to find a similar list about which of these CVD outcomes were finally included in your simulations.

• Main methodological issue: you argument that your new tool (point system) is better than traditional tools, but do not show evidence about that. In other words, the use of time-variant information is potentially an advantage of your final tool (point system) that needs to be confirmed via comparisons with other current tools. This part is extremely important, because there is not an empirical justification for using your new proposed procedure in order to create new CVD risk scores (or point systems). I suggest you include a table for comparing accuracy of your new point system against accuracy of some equivalent CVD risk scores (SCORE or WHO maybe), at least using a simulated data (in absence of desirable real data).

As we are constructing models for short-term predictions, we cannot compare the current cardiovascular risk scores with our model. Nevertheless, understanding that this is an important issue, we have explained it in the Discussion.

Comment [C7]: I want to be honest: this is stil being a very big concern to me. Thinking clinically short-term prediction (2 years) is not an standard for primary care. Thinking methodologically, it is easier to perform more accurate predictions for shorter periods of time (2 years instead of 10 yea because less variability and confounding effects a handled. I suggest to think about it and: 1) propos an strong rationale for justifying the utility of you short-term predictions in clinical contexts (supported by literature), 2) find real data (not simulated data) for evaluating the accuracy of you predictions in real contexts. This last step (b) is desirable in order to avoid the circular style of the construction and validation of your "point system (both made by using two simulated datasets).

• Your simulations seems to be adjusted for a period of 2 years of follow up; however, standard CVD risk scores are adjusted for 10 years of follow up. I suggest clarify this explanation or justify why you have used 2 years instead of 10 years.

We have explained this in the section about the data set simulation.

Comment [C8]: Please read previous commen

• This is clearly a paper with a strong methodological spirit (statistical methods). However, there is almost nothing about verification of statistical assumptions. I

suggest including relevant information about it, summarized in the manuscript and more detailed in supplemental material.

All the assumptions have been indicated for the model (at the beginning of Methods) and tested in the supplementary material for our example.

Comment [C9]: OK.

Validity of the findings

Results

• I suggest using the same standard criteria for drafting all your tables. Tables in page 18 could be omitted, and only described in the manuscript.

• Figures are not referred into the manuscript.

The tables and figures have been redone in the new version of the manuscript and all of them are now in the supplementary material.

Comment [C10]: I did not have access to your new supplemental material, so I did not can revie this change.

Discussion

• About robustness: You should mention that your new procedure preserves the robust facet of classic statistical methods behind current CVD risk scores (e.g. you still are applying Cox models).

Although in the previous version we indicated this (*We have attempted to fuse all these techniques into one single algorithm, retaining the virtues of each*), we have highlighted the techniques: relative risks model, scoring systems, dynamic predictions...

Comment [C11]: OK.

• C-statistic has implicit limitations that have not been mentioned, especially considering the way you have used it (simulated data). I suggest you explain implications of to perform only an internal validation procedure (without real data) and how you handled these implications. I recommend to read this reference previously (page 1770): http://circ.ahajournals.org/content/121/15/1768.short

We thank you for the reference and have made a comment about this issue in the relevant place.

Comment [C12]: I did not find comments abo C-statistic in the new version of your manuscript. Please clarify where is “the relevant place”.

Comments for the author

None.

Reviewer 3 (David Prieto-Merino)

Basic reporting

1) The use of the English language is not very good. This can sometimes make the text a bit more difficult to interpret. It needs a good revision.

The English version of the text has been reviewed by the medical writer Ian Johnstone (Oxford, UK).

2) The first problem is that the theoretical structure of the data that the authors are considering is not clear. At the beginning I thought that they were going to use, as predictor variables, biomarkers that will change their value along the follow up period of the patient (from the baseline onwards). The fact that they will be using Cox regression models with time dependent variables suggests this data structure as this models are thought for when the variables change over the follow up period (between time 0 and t*) as they mention. But then in table 4 the measurements of the risk factors are taken before the baseline 0, and there is no data of how they will change on time 0. So what is the theoretical data structure of the problem that the authors are trying to tackle? A simple graph with time on the x-axis, he key points of initial time for the follow up of outcomes and and the points where different kinds of data are collected will help a lot to understand.

As we have included all the models used, it is now clearer why we have measures with t<0 (*Statistical validation by simulation* and *Explanation of potential utilization*).

Experimental design

3) What do the authors mean exactly by the "longitudinal parameters" that they simulate and are so crucial. What is their physical meaning? Are they just the predictions of the values of the biomarkers at time t*? or are they some parameter of the trend of the biomarkers over the period (0, t*)? This concept is so crucial that it needs to be explicitly defined in the paper (Giving the Rizopolous reference is not enough).

We have indicated all the models with their characteristics, in addition to how the dynamic predictions are calculated.

4) About the validation: Once they have the simulations of the of the longitudinal parameters at time t* they convert this into a score using the Cox model derived previously. But this is used to predict risk in the future from when the explanatory variables are set. So the points of the score should help to predict the risk in the next say K years (so form t* to t*+K)? do the authors have the validation data the outcome in those future years (who died and when?)

We have explained how our simulated data set was calculated, and obviously it could be calculated, but taking into account that we are explaining a new method to develop a scoring system and that this data set is simulated, we have not shown this information. Furthermore, in Other S1 the readers have the time-to-event distribution of the simulated data set.

5) Validation with the score system and the outcome the estimate the C statistic. This is ok if there are no censored in the data without the event. But if there are, should it not be better to calculate the Harrell's D-statistic rather than the C-statistic?

We have applied the indicated change.

Validity of the findings

6) What is the data set on which they develop the model? From what kind of study they got it from? are there any censored detain the follow up? How did they create the validation set? was it a random subset from the data? why that size? Why do they drop patients without values of some of the longitudinal parameters during the follow up? Is it not the aim of this method precisely to predict those values with the previous history of the parameters?

A new section has been developed dealing with this issue.

7) Figure-2 Is confusing because it shows the "medium" risk group with higher numbers than the "high" and "very high" risk groups. Please put it on a risk scale dividing by the appropriate denominators.

This figure has been removed from the paper.

Comments for the author

I think this paper addresses a very interesting problem that will become only more important with the coming of the big data era: How to build risk prediction models using changing values of biomarkers rather than just one baseline measure from the biomarker. Unfortunately I think the paper is not very well written and is very difficult to understand what the authors have done at each step (or even what they are trying to do).

We have adapted the paper to address all the comments of the reviewers to explain our idea better.

Authors: Antonio Palazón-Bru, PhD[1], Julio Antonio Carbayo-Herencia, MD, PhD[2], Isabel Vigo-Aguiar, PhD[3], Vicente Francisco Gil-Guillén, MD, PhD[1].

Institutions:

1. Department of Clinical Medicine, Miguel Hernández University, San Juan de Alicante, Alicante, Spain.
2. Chair of Cardiovascular Risk, San Antonio Catholic University, Murcia, Region of Murcia, Spain.
3. Department of Applied Mathematics, University of Alicante, San Vicente del Raspeig, Alicante, Spain.

Corresponding author: Prof. Antonio Palazón-Bru, PhD. Department of Clinical Medicine, Miguel Hernández University, Carretera de Valencia - Alicante S/N, 03550, San Juan de Alicante (Spain). Phone number: +34 965919449. Fax number: +34 965919450. E-mail: antonio.pb23@gmail.com

INTRODUCTION

Given that cardiovascular diseases (CVD) are one of the main causes of death in the world (WHO, 2014), prediction models are interesting in order to determine those risk factors that can be acted on to reduce the probability of CVD (Molinero, 2003). The simplest model to make predictions about a dichotomous event, such as CVD, is logistic regression (Hosmer & Lemeshow, 2000). This model produces an equation which, once the values for the various risk factors are known, can be used to evaluate the likelihood of the appearance of disease. However, this sort of model fails to consider exposure time. This is precisely what is done in survival models, which analyse the time of occurrence of a particular event. Although the best known of these models is Cox (Hosmer & Lemeshow, 2008), it is not the only alternative available. There exist other possible methods to analyse survival, called parametric models as they assume a concrete type of distribution, such as the Weibull model, used in the SCORE project (Conroy et al., 2003). Indeed, the Framingham study used both logistic regression models and survival models (parametric and non-parametric) (National Heart, Lung, and Blood Institute, 2015).

In conjunction with the Framingham and SCORE predictive models, others have been developed that are also used in clinical practice, though to a lesser extent, such as the Reynolds risk score and the WHO/ISH score (Cooney, Dudina & Graham, 2009). Common to all these is the making of predictions about CVD over a 10-year period, though they consider different outcomes (morbidity and mortality with coronary heart disorders, mortality from coronary heart disorders, cardiovascular morbidity and mortality, or just cardiovascular mortality) and use different mathematical models (Cox and Weibull). These models enable physicians to make long-term decisions for their patients. In addition, the clinical practice guidelines recommend using these predictive models to stratify the cardiovascular risk of patients. For example, in Europe, the European Guidelines on cardiovascular disease

prevention in clinical practice indicate “A risk estimation system such as SCORE can assist in making logical management decisions, and may help to avoid both under-and overtreatment” (Perk et al., 2012). In other words, clinicians follow the guidelines to improve the decision-making process in order to prevent CVD, and it is these very guidelines that indicate the use of these predictive models. Accordingly, these models are very relevant in daily clinical practice.

Given the complexity of these mathematical models an algorithm is used to enable the clinician to understand them more easily, though precision is lost in the estimation of the probability of CVD (Sullivan, Massaro & D'Agostino, 2004). To do this the mathematical models have been transformed into coloured risk tables that can be used systematically in clinical practice. However, these tables are based on models that manage clinical variables in the baseline situation of the patient (Conroy et al., 2003; National Heart, Lung, and Blood Institute, 2015), and do not therefore take into account the variability of the variables over time, as the biological parameters are being considered constant over the follow-up period when in fact they vary greatly and the physician can intervene using drugs to either reduce or increase their value (NCEP, 2002; American Diabetes Association , 2014; James et al., 2014; Stone et al., 2014).

Predictive models for survival in other diseases do consider the temporal variability of a single biological marker (as well as the baseline variables). These are known as Joint Models for Longitudinal and Time-to-Event Data and comprise two parts: 1) A mixed linear model to determine the path of a longitudinal parameter and 2) A survival model relating the baseline variables and the longitudinal parameter with the appearance of an event. These models can be used to make more precise predictions about the development of a disease (Rizopoulos, 2012). However, due to their complexity they are not used in general clinical practice. In addition, joint modelling when the survival part is formed by a linear function

with multiple longitudinal parameters (usual modelling in traditional survival analysis in the health sciences) has only been examined theoretically and currently remains a complete computational challenge. This has resulted in the development of algorithms to make predictions, as in the univariate case (Rizopoulos, 2011).

As CVD are the leading cause of death worldwide (WHO, 2014), a simple tool is required that can be applied systematically in usual clinical practice that determines the risk for CVD with greater accuracy (taking into account the temporal variability of all the cardiovascular risk factors). Here we aim to show the viability and properties of a new methodological alternative for constructing cardiovascular risk scores (construction, statistical validation by simulation and potential utilization with the new theoretical model) dealing with this problem and its application in a set of simulated data, with the purpose of helping readers understand how to apply it to a real data set with repeated measures of cardiovascular risk factors.

Comment [C1]: I suggest to erase this part.

Comment [C2]: I suggest: "the temporal variability of CVD risk factors"

Comment [C3]: Please clarify. Do you expect that researchers only confirm the validity of your "current point system" in real data? or Do you exp that researchers can follow all your method, estimating statistical parameters and adjusting equations in order to re-calibrate the point system fit on the particular characteristics of each populat (as usual)? As researcher, I really prefer the secon one.

MATERIALS AND METHODS

The basic models used to develop the new method were the Cox model with time-dependent variables, points system in the Framingham Heart Study, Joint Models for Longitudinal and Time-to-Event Data, and predictions of the longitudinal biomarkers using these Joint Models.

Cox model with time-dependent variables

Let T be a non-negative random variable denoting the observed failure time, which is the minimum value of the true event time T^* and the censoring time C (non-informative right censoring). In other words, $T = \min(T^*, C)$. In addition, we define δ as the event indicator, which takes the value 1 if $T^* \leq C$ and 0 otherwise. On the other hand, let $\boldsymbol{W}$ be the vector of baseline covariates and $\boldsymbol{Y}(t)$ the vector of time-dependent covariates, assuming a defined

value for $t \geq 0$. With these data, the Cox model with time-dependent variables takes the following form (risk function):

$$h\big(t|\boldsymbol{w}, \boldsymbol{y}(t)\big) = h_0(t)\exp\{\boldsymbol{\gamma}^T\boldsymbol{w} + \boldsymbol{\alpha}^T\boldsymbol{y}(t)\},$$

where $h_0(t)$ is the baseline risk function, and $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$ are the vectors of the regression coefficients for the baseline and time-dependent covariates, respectively (Andersen & Hill, 1982).

The estimation of the model parameters is based on the partial likelihood function (Andersen & Hill, 1982). On the other hand, we have to corroborate whether the functional form of the covariates in the model is linear. This should be performed using graphical methods (Martingale residuals against the covariate of interest). Finally, we have to assess whether the model fits the data well, through the analysis of the Cox-Snell residuals (graphical test).

The classical Cox regression model (with no time-varying covariates), deletes $\boldsymbol{\alpha}$ and $\boldsymbol{y}(t)$ from the above expression. Furthermore, the model has to verify the following condition (proportional hazard assumption):

$$\log\left(\frac{h(t|\boldsymbol{w})}{h_0(t)}\right) = \boldsymbol{\gamma}^T\boldsymbol{w}.$$

Points system in the Framingham Heart Study

We summarize the steps of the method developed by the Framingham Heart Study to adapt a Cox regression model with p covariates to risk charts (Sullivan, Massaro & D'Agostino, 2004):

1) Estimate the parameters of the model: $\hat{\boldsymbol{\gamma}}$.

2) Organize the risk factors into categories and determine reference values:

a. Continuous risk factor (e.g., age): set up contiguous classes and determine reference values for each. Example for age: 18-30 [24], 30-39 [34.5], 40-49 [44.5], 50-59 [54.5], 60-69 [64.5] and ≥70 years [74.5]. In brackets is the reference value. The Framingham Heart Study researchers recommend mid-points as acceptable reference values, and for the first and last class the mean between the extreme value and 1st (first class) or 99th percentiles (last class).

b. Binary risk factors (e.g. gender, 0 for female and 1 for male): the reference value is again either 0 or 1.

Let W_{ij} denote the reference value for the category j and the risk factor i, where $i = 1, \ldots, p$ and $j = 1, \ldots, c_i$ (total number of categories for the risk factor i).

3) Determine the referent risk factor profile: the base category will have 0 points in the scoring system and it will be denoted as W_{iREF}, $i = 1, \ldots, p$.

4) Determine how far each category is from the base category in regression units: calculate $\hat{\boldsymbol{\gamma}}_i \cdot (W_{ij} - W_{iREF})$, $i = 1, \ldots, p$ and $j = 1, \ldots, c_i$.

5) Set the fixed multiplier or constant B: the number of regression units equivalent to 1 point in the points system. The Framingham Heart Study generally uses the increase in risk associated with a 5-year increase in age.

6) Determine the number of points for each of the categories of each risk factor: the closest integer number to $\hat{\boldsymbol{\gamma}}_i \cdot (W_{ij} - W_{iREF})/B$.

7) Determine risks associated with point totals: $1 - \hat{S}_0(t)^{\exp\{\sum_{i=1}^{p}(\hat{\gamma}_i \cdot W_{iREF}) + B \cdot Points - \sum_{i=1}^{p}\hat{\gamma}_i \cdot \widehat{w}_i\}}$, where $\hat{S}_0(t)$ is calculated through the Kaplan-Meier estimator.

Joint Models for Longitudinal and Time-to-Event Data

Using the former notation, we have the random variables vector $\{T, \boldsymbol{W}, Y(T)\}$, where $Y(T)$ is only a time-dependent variable (longitudinal outcome) which has its values defined intermittently for t. In other words, for a subject $(i = 1, \ldots, n)$, $y(t)$ is only defined for t_{ij} $(j = 1, \ldots, n_i)$, $y_i(t_{ij})$, where $0 \leq t_{i1} \leq t_{i2} \leq \cdots \leq t_{in_i}$. Now, we will denote as $m(t)$ the true and unobserved value of the longitudinal outcome at time t ($m_i(t)$ for the subject i). To assess the effect of $m(t)$ on the event risk, a standard option is to adjust a Cox regression model with one time-dependent covariate:

$$h(t|\mathcal{M}(t), \boldsymbol{w}) = h_0(t^*) \exp\{\boldsymbol{\gamma}^T \boldsymbol{w} + \alpha m(t)\},$$

where $\mathcal{M}(t)$ for a subject i is defined as $\mathcal{M}_i(t) = \{m_i(u); 0 \leq u < t\}$, which denotes the history of the true unobserved longitudinal process up to time t. The other parameters in the expression follow the structure of the Cox regression model with time-dependent variables (see former section). The baseline risk function can be unspecified or can be approximated with splines or step functions (Rizopoulos, 2012).

In the above expression, we have used $m(t)$ as the true unobserved longitudinal process. However, in our sample we have $y(t)$; therefore we will estimate $m(t)$ using $y(t)$ through a linear mixed effects model to describe the subject-specific longitudinal evolutions:

$$\begin{cases} y_i(t) = m_i(t) + \varepsilon_i(t) \\ m_i(t) = \boldsymbol{x}_i^T(t)\boldsymbol{\beta} + \boldsymbol{z}_i^T(t)\boldsymbol{b}_i \\ \boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D}) \\ \varepsilon_i(t) \sim N(0, \sigma^2) \end{cases},$$

where $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$ denote the vectors of regression coefficients for the unknown fixed-effects parameters and the random effects respectively, $\boldsymbol{x}_i(t)$ and $\boldsymbol{z}_i(t)$ denote row vectors of the design matrices for the fixed and random effects respectively, and $\varepsilon_i(t)$ is the error term with

variance σ^2. Finally, $\boldsymbol{b_i}$ follows a normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{D}$, and independent of $\varepsilon_i(t)$ (Rizopoulos, 2012).

The estimation of the parameters of the joint models is based on a maximum likelihood approach that maximizes the log-likelihood function corresponding to the joint distribution of the time-to-event and longitudinal outcomes (Rizopoulos, 2012).

Regarding the assumptions of the model, we have to assess them for both submodels (longitudinal and survival) using the residual plots. For the longitudinal part, we will plot the subject-specific residuals versus the corresponding fitted values, the Q-Q plot of the subject-specific residuals, and the marginal residuals versus the fitted values. On the other hand, for the survival part, we will plot the subject-specific fitted values for the longitudinal outcome versus the martingale residuals, and finally we will determine graphically whether the Cox-Snell residuals is a censored sample from a unit exponential distribution (Rizopoulos, 2012). Regarding, the last component (random effects part) of the joint model for which we have indicated an assumption, other authors have showed that linear mixed-effects models are relatively robust to misspecification of this distribution (Verbeke & Lesaffre, 1997).

Predictions of the longitudinal biomarkers using these Joint Models for Longitudinal and Time-to-Event Data

Let $\{t_i, \delta_i, \boldsymbol{w}_i, y_i(t_{ij}), 0 \leq t_{ij} \leq t_i, j = 1, \ldots, n_i\}, i = 1, \ldots, n$ be a random sample of the random variables vector $\{T, \Delta, \boldsymbol{W}, Y\}$, using the former notation. A joint model has been fitted using this sample. Now, we are interested in predicting the expected value of the longitudinal outcome at time $u > t$ for a new subject i who has a history up to the time t of the observed longitudinal marker $\mathcal{Y}_i(t) = \{y_i(s); 0 \leq s < t\}$:

$$\omega_i(u|t) = E_Y\{y_i(u)|t_i^* > t, \mathcal{Y}_i(t), \boldsymbol{w_i}; \boldsymbol{\theta}\},$$

where $\boldsymbol{\theta}$ denotes the parameters' vector of the joint model (Rizopoulos, 2011).

Rizopoulos developed a Monte Carlo approach to perform this task, based on Bayesian formulation. He obtained the following simulation scheme (Rizopoulos, 2011):

Step 1: Draw $\boldsymbol{\theta}^{(l)} \sim \mathcal{N}\left(\widehat{\boldsymbol{\theta}}, \widehat{var}\left(\widehat{\boldsymbol{\theta}}\right)\right)$.

Step 2: Draw $\boldsymbol{b_i^{(l)}} \sim \left\{\boldsymbol{b_i} | t_i^* > t, \mathcal{Y}_i(t), \boldsymbol{w_i}; \boldsymbol{\theta}^{(l)}\right\}$.

Step 3: Compute $\omega_i{}^{(l)}(u|t) = \boldsymbol{x_i^T}(u)\boldsymbol{\beta}^{(l)} + \boldsymbol{z_i^T}(u)\boldsymbol{b_i^{(l)}}$

This scheme should be repeated L times. The estimation of the parameter is the mean (or median) of the calculated values ($\omega_i{}^{(l)}(u|t), l = 1, \dots, L$) and the confidence interval is formed by the percentiles (95%: 2.5% and 97.5% percentiles) (Rizopoulos, 2011).

We highlight that these predictions have a dynamic nature, that is, as time progresses additional information is recorded for the patient, so the predictions can be updated using this new information.

Construction

We wish to determine the probability of having CVD with effect from a baseline situation ($t = 0$) up to a fixed point in time ($\tilde{t}$), given a series of risk factors measured at baseline and during this follow-up. To do this requires the following steps:

1) Adjust a Cox regression model with time-dependent variables. As we are unable to estimate a joint model with multiple longitudinal parameters (Rizopoulos, 2012), we use the classic extended Cox model (with no shared structure), which requires knowing the values of all the longitudinal parameters at any value of t. As this is not known because the parameters are recorded intermittently, we take the last value in time as a reference.

2) Use the procedure of the Framingham study to adapt the coefficients of the model obtained to a points system and determine the probabilities of CVD for each score up to the moment $\tilde{t}$. We then use these probabilities to construct risk groups that are easy for the clinician to understand (for example, in multiples of 5%) (Sullivan, Massaro & D'Agostino, 2004).

3) Adjust a joint model for longitudinal and time-to-event data for each longitudinal parameter recorded during the follow-up. This will also include all the baseline variables. These models are constructed to make predictions about the longitudinal parameters in new patients (statistical validation by simulation and potential utilization).

Statistical validation by simulation

Once the points system has been constructed, we wish to see whether the model determines the onset of CVD accurately in a different set of subjects (validation sample). In this validation sample we know the longitudinal markers up to the point $t = 0$ [record of cardiovascular risk factors in the clinical history which were measured before the baseline situation ($t < 0$)] and the value of the variables at baseline. With this information we determine the probability each subject has of experiencing an event, and we then compare this with what actually occurred; i.e., determine whether the model is valid. To determine this validity we follow these steps:

1) Determine L simulations of the longitudinal parameters at the time point $\tilde{t}$ using the models mentioned in step 3) of construction, from the history $(t < 0)$ and the baseline variables $(t = 0)$ (Rizopoulos, 2011). We will use these simulated values to construct a distribution of the points for each sample subject. Thus, each subject will have L values for the points variable (evaluating the points system using the simulated values and the baseline variables is sufficient), and for each l^{th} simulation each patient will have a points score. In other words, each simulation will have a distribution of the points variable.

2) For each l^{th} simulation adjust a classic Cox model (without time-dependent variables), using just the score obtained as the only explanatory variable. Determine the Harrell's concordance statistic for each of these L models. These values will give us a distribution of values for this statistic, with which we calculate the mean (or the median) and the 2.5% and 97.5% percentiles (Rizopoulos, 2011). This way we construct a confidence interval for this statistic, which will indicate the discriminating capacity of the points system to determine which patients will develop CVD.

3) Calculate the median of the points distribution for each patient in the validation sample. Note that we do not use the mean as it could contain decimals and this has no sense when applying the scoring system. Using these medians, classify each patient in a risk group and compare the rate of events predicted by the points system in each group to the actual observed rate. The test used for this process will be Pearson χ^2 test.

The concordance statistic used has been reported to have various limitations (Lloyd-Jones, 2010). For example, it does not compare whether the estimated and observed risks are similar in the subjects. Accordingly we have added the analysis of the differences between the expected events and the observed events, which minimises this particular problem. In addition, it is very sensitive to large hazard ratio values (≥9). Nonetheless, we have to consider that as all the variables are quantitative (not categorized), the hazard ratio values do not surpass this threshold. Accordingly, the joint analysis of the concordance index of Harrell and the differences between the expected and the observed events enables us to validate statistically by simulation of the proposed model.

Explanation of potential utilization

Once the points system has been validated statistically the clinician can then apply the system to determine the cardiovascular risk in a new patient, and take any necessary measures to

reduce this risk. The healthcare professional will already have historical information about the longitudinal parameters $(t < 0)$ and information about the baseline situation $(t = 0)$ of the new patient. The steps to be followed by the clinician are:

1) Determine the value of each longitudinal parameter at the time $\tilde{t}$. To do this we apply the models obtained in step 3) of construction to the history and the baseline situation of the new patient, in order to determine L simulations for each longitudinal parameter, similar to what was done in the validation process. For each l^{th} simulation we determine the score corresponding to the profile of cardiovascular risk factors obtained (simulated and baseline information values). This will give us a points distribution for the new patient.

2) Determine the median and the 2.5% and 97.5% percentiles of the points vector constructed above. The median will be the estimation of the score for the new patient and the percentiles will define the confidence interval (Rizopoulos, 2011). As each score has an associated risk, the healthcare professional will be able to know the probability of CVD at time $\tilde{t}$, together with its confidence interval. Finally, the clinician will know the values of the biological parameters at $\tilde{t}$ of the median of the points system. This way the clinician will be able to see which of these parameters has a score above normal; i.e., see the possible areas of intervention to reduce the cardiovascular risk.

3) The clinician now knows the cardiovascular risk and which parameters have a score above normal, so he or she can then design the best intervention for that patient. This presents a problem, as we need to know the value of each biological parameter at time $\tilde{t}$; i.e., the clinician knows an approximation based on simulations constructed from the patient history but does not know how the interventions will affect the cardiovascular risk.

From the previous step the clinician knows the parameters on which to act and the history of these parameters as well as the baseline situation. From these measurements the clinician can

establish a realistic objective for the next patient visit at time $\tilde{\tilde{t}}$ ($0 < \tilde{\tilde{t}} < \tilde{t}$). The clinician now inserts the desired value of the biological parameter at $\tilde{\tilde{t}}$ and determines its value at time $\tilde{t}$; i.e., determine L simulations for each cardiovascular risk factor using the previous models (step 3 of construction), adding a new value to the history ($\tilde{\tilde{t}}$).

These calculations will give the benefit of the intervention (estimation [mean or median] of the biological parameter at $\tilde{t}$) and the clinician will be able to see from the points system how the patient's risk will be reduced.

Simulation on a data set

With the sole purpose of explaining how to use the method proposed here, we have simulated a data set upon which to apply each of the steps described above. Note that we are in fact going to simulate two data sets, one to construct the model and the other to validate it statistically via simulation. So that both sets are biologically plausible we have used estimations obtained in the Puras-GEVA cardiovascular study, which is in the process of publication in Medicine (Artigao-Ródenas et al., 2015).

Our data sets will include the following biological parameters: age (years), systolic blood pressure (SBP) (mmHg), HbA1c (%), atherogenic index, gender (male or female) and smoking (yes or no). Of these, the SBP, HbA1c and the atherogenic index will be present at baseline ($t = 0$) and in the follow-up for the construction sample ($t > 0$) or recorded in the clinical history for the statistical validation sample via simulation ($t < 0$). The choice to include these variables was based on the current cardiovascular risk scales (Conroy et al., 2003; National Heart, Lung, and Blood Institute, 2015), except for HbA1c, which is used instead of a diagnosis of diabetes mellitus in order to include another time-dependent parameter in the final model, in addition to which this way enables us to value the control of the diabetes mellitus (HbA1c <6.5%) when preventing CVD.

For the main variable (time-to-CVD) we shall suppose that our cohort is used to predict CVD with a follow-up of 2 years. Note that the traditional cardiovascular risk scales use a time of 10 years (Conroy et al., 2003; National Heart, Lung, and Blood Institute, 2015). We have used this lower value because we are going to make predictions for the longitudinal parameters with effect from the baseline situation ($t = 0$) up to the prediction time and if we take a prediction value of 10 years the predictions for the longitudinal parameters will vary greatly and not allow us to make precise predictions about which patient will develop CVD, which would negate the usefulness of the method proposed here. Nevertheless, the fact that the predictions for the longitudinal parameters have a dynamic character (see *Predictions of the longitudinal biomarkers using these Joint Models for Longitudinal and Time-to-Event Data*) enables us to determine the risk at 2 years with greater precision whenever the patient attends the office of the healthcare professional.

The longitudinal follow-up measurements (construction sample) assumed that the patient attends the physician's office once every 3 months for measurements of SBP, HbA1c and the atherogenic index. This is done until the end of the follow-up for each patient. The statistical validation sample using simulation supposes that there is a certain probability of having records in the clinical history of all the longitudinal parameters every 3 months for 5 years retrospectively ($t < 0$). The probability is different for each of the visits and will depend on each patient. In other words, we will have intermittent measurements of all these parameters from $t = -5\ years$ to $t = 0$.

The supplementary material (Other S1) details all the mathematical formulae used to construct our data sets, always based on the Puras-GEVA study (Artigao-Ródenas et al., 2015). The simulation was done using R 2.13.2 and IBMS SPSS Statistics 19.

Comment [C4]: Please clarify. Does your meth only can make accurate predictions up to 2 years? Please, take into account that for CVD primary ca the valuable predictions should be given very earl so that is why the 10 years cut-off point is a standard. Moreover, in these days CVD researche are very interested in lifetime predictions, implyin longer periods of time (http://jaha.ahajournals.org/content/4/8/e002112.s t).

Comment [C5]: Please, use the same style of references through all document.

RESULTS

Given the amount and extension of the results these are given in detail in the supplementary material (Other S2 and Other S3). As before, the analysis was done with R 2.13.2 and IBM SPSS Statistics 19.

Comment [C6]: It is understandable; however, is important to provide key results into the manuscript. For example, you can show the final equations that shape your point system (in a table a flowchart) and the summary of the results that show the accuracy of your proposed method again accuracy of other current method. This section is important, because gives to readers figures and fa before to start the rationale of discussion.

DISCUSSION

This paper describes a method to construct predictive models for CVD considering the variability of cardiovascular risk factors and at the same time having the simplicity of points systems, which are widely used in daily clinical practice worldwide (Conroy et al., 2003; Cooney, Dudina & Graham, 2009; National Heart, Lung, and Blood Institute, 2015).

The cardiovascular risk scales currently available do not value the temporal variability of the parameters controlling the risk factors, although a very positive aspect of these scales is that they take into account simplicity for immediate application by healthcare professionals, the persons who really have to apply these mathematical models (Conroy et al., 2003; Cooney, Dudina & Graham, 2009; National Heart, Lung, and Blood Institute, 2015). The joint models currently used do take into account variability over time of a single longitudinal parameter (Rizopoulos, 2011), but their interpretation is not as easy as a points system and they cannot be used with various longitudinal parameters, a key question in the multifactorial aetiology of CVD. We have attempted to fuse all these techniques into one single algorithm, retaining the virtues of each (relative risks model, scoring systems, dynamic predictions...).

Comparison between our proposed model and current cardiovascular risk scales is problematic. Our model is suitable to make short-term predictions, though the more time that

passes from the baseline situation ($t = 0$) when making a prediction, the variability of the predictions of the longitudinal parameters increases (Rizopoulos, 2012). This same situation can be found in other areas, such as the economy (stock exchange) or meteorology (weather forecast). This however does not weaken our model, since because the predictions for the longitudinal parameters are dynamic (Rizopoulos, 2011), any time that we update the clinical information about our patient the risk is immediately recalculated. This can be seen in the proposed example (Other S2), where when we introduce new values for the longitudinal parameters these are updated and a new score for the patient is calculated. In other words, the proposed method could be used to calculate the patient's risk every time the patient attends the office, whereas the traditional risk scales can be used with a longer time interval, as the prognosis is for 10 years. Thus, the two types of model could be used to assess the risk, for both the short term and the long term.

Comment [C7]: I can see the potential of this method, but there is not enough evidence for supporting it. I have to insist: find real dataset, the run your algorithm and compare your successive predictions against the classical 10-years predictio Show the advantages graphically. After seeing thi kind of evidence, researchers will feel motivated a confident to replicate your method in the near futu

Obtaining simulations from longitudinal parameters is not easy and implies a computational cost of about one minute with the statistical package R to implement a total of 100 using a normal computer. On the other hand, the historical values of the longitudinal parameters are recorded in the clinical history, which nowadays is usually electronic (Palazón-Bru et al., 2014). Given this situation, all the information needed to apply our models is already computerised, so the algorithms implemented in the statistical package R can be adapted to the underlying language of the database containing the values of the risk factors. Thus, all the calculations will be immediate for the healthcare professional. In other words, just pressing a key will be enough to bring up on the screen in a very short time the histogram shown in Other S2, the theoretical points system and the set of values of the risk factors determining the median score. In addition, when the physician decides to intervene he or she will indicate the duration of the intervention and the possible values for the new

patient. After introducing this new information the two histograms could be shown together (Other S2), which will enable the physician to see the benefit of the intervention.

As this algorithm was developed from a set of simulated data, we encourage others who have cardiovascular databases like that used here to implement a model with the characteristics described herein. Thus, if using real-life data achieves greater predictive precision we shall be able to apply this method~~ology~~ to obtain the best short-term prognosis and thus take the most appropriate decisions for the benefit of the patient.

CONCLUSIONS

We developed an algorithm to construct cardiovascular risk scales based on a points system that also takes into account the variability of the risk factors. These issues are important as the popularity of points systems in clinical practice and the improved predictive accuracy using all the information recorded in the clinical history will improve the currently used procedure. Nonetheless, we must be cautious as the algorithm has not yet been used with a real set of data. Cardiovascular cohort studies using this system are thus required in order to validate it.

ACKNOWLEDGEMENTS

The authors thank Ian Johnstone for help with the English language version of the text.

REFERENCES

American Diabetes Association. 2014. Standards of medical care in diabetes--2014. *Diabetes Care* 37 Suppl 1:S14-80.

Andersen PK, Gill RD. 1982. Cox's Regression Model for Counting Processes: A Large Sample Study. *Annals of Statistics* 10: 1100-1120.

Artigao-Ródenas LM, Carbayo-Herencia JA, Palazón-Bru A, Divisón-Garrote JA, Sanchis-Domènech C, Vigo-Aguiar I, Gil-Guillén VF; on behalf of GEVA. 2015. Construction and ~~validation~~ utilization of a 14-year cardiovascular risk score for use in the general population: the PURAS-GEVA chart. *Medicine (Baltimore)* 2015.

Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, De Backer G, De Bacquer D, Ducimetière P, Jousilahti P, Keil U, Njølstad I, Oganov RG, Thomsen T, Tunstall-Pedoe H, Tverdal A, Wedel H, Whincup P, Wilhelmsen L, Graham IM; SCORE project group. 2003. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *European Heart Journal* 24: 987-1003.

Cooney MT, Dudina AL, Graham IM. 2009. Value and limitations of existing scores for the assessment of cardiovascular risk: a review for clinicians. *Journal of the American College of Cardiology* 54: 1209-1227.

Hosmer DW, Lemeshow S. 2000. Applied Logistic Regression. New York: Wiley.

399

400 Hosmer DW, Lemeshow S, May S. 2008. Applied Survival Analysis: Regression Modeling
401 of Time-to-Event Data. New York: Wiley.

402

403 James PA, Oparil S, Carter BL, Cushman WC, Dennison-Himmelfarb C, Handler J, Lackland
404 DT, LeFevre ML, MacKenzie TD, Ogedegbe O, Smith SC Jr, Svetkey LP, Taler SJ,
405 Townsend RR, Wright JT Jr, Narva AS, Ortiz E. 2014. 2014 evidence-based guideline for the
406 management of high blood pressure in adults: report from the panel members appointed to
407 the Eighth Joint National Committee (JNC 8). *JAMA* 311: 507-520. Erratum in: *JAMA* 311:
408 1809.

409

410 Molinero LM. 2003. Modelos de riesgo cardiovascular. Estudio de Framingham. Proyecto
411 SCORE. Available at http://www.seh-lelha.org/pdf/modelries.pdf (accessed July 2015).

412

413 National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and
414 Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). 2002. Third
415 Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection,
416 Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III)
417 final report. *Circulation* 106: 3143-3421.

418

419 National Heart, Lung, and Blood Institute (Boston University). 2015. The Framingham Heart
420 Study. Available at http://www.framinghamheartstudy.org/ (accessed July 2015).

421

422 Lloyd-Jones DM. 2010. Cardiovascular risk prediction: basic concepts, current status, and
423 future directions. *Circulation* 121: 1768-1777.

424

425 Palazón-Bru A, Gil-Guillén VF, Orozco-Beltrán D, Pallarés-Carratalá V, Valls-Roca F,
426 Sanchís-Domenech C, Martín-Moreno JM, Redón J, Navarro-Pérez J, Fernández-Giménez A,
427 Pérez-Navarro AM, Trillo JL, Usó R, Ruiz E. 2014. Is the physician's behavior in
428 dyslipidemia diagnosis in accordance with guidelines? Cross-sectional ESCARVAL study.
429 *PLoS One* 9:e91567.

430

431 Perk J, De Backer G, Gohlke H, Graham I, Reiner Z, Verschuren M, Albus C, Benlian P,
432 Boysen G, Cifkova R, Deaton C, Ebrahim S, Fisher M, Germano G, Hobbs R, Hoes A,
433 Karadeniz S, Mezzani A, Prescott E, Ryden L, Scherer M, Syvänne M, Scholte op Reimer
434 WJ, Vrints C, Wood D, Zamorano JL, Zannad F; European Association for Cardiovascular
435 Prevention & Rehabilitation (EACPR); ESC Committee for Practice Guidelines (CPG). 2012.
436 European Guidelines on cardiovascular disease prevention in clinical practice (version 2012).
437 The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on
438 Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine
439 societies and by invited experts). *European Heart Journal* 33: 1635-1701. Erratum in:
440 *European Heart Journal* 33: 2126.

441

442 Rizopoulos D. 2011. Dynamic predictions and prospective accuracy in joint models for
443 longitudinal and time-to-event data. *Biometrics* 67: 819-829.

444

445 Rizopoulos D. 2012. Joint Models for Longitudinal and Time-to-Event Data With
446 Applications in R. Boca Raton: CRC Press.

447

448 Stone NJ, Robinson JG, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, Goldberg
449 AC, Gordon D, Levy D, Lloyd-Jones DM, McBride P, Schwartz JS, Shero ST, Smith SC Jr,
450 Watson K, Wilson PW; American College of Cardiology/American Heart Association Task
451 Force on Practice Guidelines. 2014. 2013 ACC/AHA guideline on the treatment of blood
452 cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American
453 College of Cardiology/American Heart Association Task Force on Practice Guidelines.
454 *Journal of the American College of Cardiology* 63(25 Pt B):2889-2934. Erratum in: *Journal*
455 *of the American College of Cardiology* 63(25 Pt B): 3024-3025.

456

457 Sullivan LM, Massaro JM, D'Agostino RB Sr. 2004. Presentation of multivariate data for
458 clinical use: The Framingham Study risk score functions. *Statistics in Medicine* 23: 1631-
459 1660. Review.

460

461 Verbeke G, Lesaffre E. 1997. The effect of misspecifying the random effects distribution in
462 linear mixed models for longitudinal data. *Computational Statistics and Data Analysis* 23:
463 541-56.

464

465 World Health Organization. 2014. The top 10 causes of death. Available at
466 http://www.who.int/mediacentre/factsheets/fs310/en/ (accessed July 2015).