Author’s response to reviews

Title: Repeatedly measured predictors: a comparison of methods for prediction modelling

Authors:

Marieke Welten (m.welten@vumc.nl)
Marlou de Kroon (m.dekroon@erasusmc.nl)
Carry Renders (carry.renders@vu.nl)
Ewout Steyerberg (e.steyerberg@erasusmc.nl)
Hein Raat (h.raat@erasusmc.nl)
Jos Twisk (jwr.twisk@vumc.nl)
Martijn Heymans (mw.heymans@vumc.nl)

Version: 1 Date: 27 Dec 2017

Author’s response to reviews:

December 2017

Dear dr. Richard J Stevens,

Thank you for reviewing our manuscript: ‘Repeatedly measured predictors: a comparison of methods for prediction modelling.’ (DAPR-D-17-00019). We are grateful for the detailed comments of the reviewers and have made changes in our manuscript in response to these comments.

In the ‘response to the reviewers’ document, you will find our specific responses to all comments from your reviewers, including corresponding changes in the manuscript. Textual changes, which are part of the revised manuscript, are underlined. We hope to have answered all raised points sufficiently for the acceptance of our revised manuscript for publication in Diagnostic and Prognostic Research.

Yours sincerely,

Marieke Welten
response to the reviewers:

Reviewer: 1

Although a somewhat specialist topic, this manuscript is likely to be relevant to many readers of the journal and I read it with interest. I have only a few comments, outlined below.

Comment 1

After reading this article, I found that I was left wondering what the computational "cost" differences were between the methods (in terms of analysis time, computational resources needed, and also data preparation time/complexity). Are the authors able to provide more information on this? The discussion in Section 4.4 is helpful for deciding on which method to use, but this would also be helpful information for aiding such decisions.

Response: We thank the reviewer for the comments. For our study we chose methods that are easily applicable in all standard statistical software programmes and therefore they should probably not be rejected based on computational cost. Computational costs for the different methods is very difficult to quantify as it will depend on numerical factors such as the structure (current and desired for analysis) and size (number of persons; number of measurements; and the number of imputed datasets if applicable) of the dataset. Things to bear in mind when thinking of computational costs are:

1.) Data structure; Method 1 to 5 require a wide structured dataset and method 6 a long structured dataset. In case the measurements are not structured according to age or follow up time per variable (like our example of the unsorted Child Health Care data), measurements will probably first need to be sorted appropriately for methods 1, 2, 4 and 5. This data preparation time will largely depend on the current and desired dataset structure and the number of repeated measurements.

2.) Computing predictor variables. Methods 3 to 6 require new predictors to be constructed from the measurements. The mean or maximum of all available measurements is easily computed with statistical programmes. Determining the change/difference between subsequent measurements may also be relatively easy when there are no missing values and not many variables present. However if there are missing values it will depend on how these will be handled (is the change then also missing or will it be determined with the next valid measurement). The conditional measurements are easily calculated by running the different regressions and saving of the residual values, depending on the number of measurements and persons, computational time might take a bit longer than
the previous methods. Method 6 will probably take the longest time to compute the predictors, as individual specific growth curves need to be calculated by running analyses separately for individuals in the first step and afterwards the results will have to be extracted and merged with the data for analysis of step 2. Computational time will depend strongly on the desired growth curve (linear, quadratic, cubic), the number of persons and measurements.

3.) Analysis time of the outcome-predictor model. Differences in computational time of running the models including the final predictors will largely be influenced by the number of available measurements, more measurements will lead to more predictors in the model and thus likely longer computational time for methods 1, 4 and 5.

We have addressed this topic in our manuscript in section 4.4 Practical implications as follows:

“4.4.5. Computational cost

As stated before we have focused on comparing prediction methods that are easily applicable for epidemiologists, as all the 6 methods are easily applied none of the methods should probably be selected or rejected based on computational costs: they are all very manageable for statistical software programmes if applied appropriately. Although there are differences in computational costs for the different methods these are very difficult to quantify as it will depend on different factors such as the structure (current and desired for analysis) and size (number of persons; number of measurements; and the number of imputed datasets if applicable) of the dataset.

”(Page 24.)

Comment 2

I thought the discussion of alternative methods in Section 4.1 was useful, but in a bit of an odd place (in the Discussion) - perhaps better located in the Introduction?

Response: We agree that it is odd for the information on the selection and exclusion of methods to be first mentioned at the end of the article. However, we considered the information (mainly the reasoning behind excluding specific methods from our comparison) to be too extensive to be moved completely to the introduction and have therefore decided to keep this text in the discussion section and to additionally add the main information on method selection to the 1. Background section:

“Our aim is to compare different methods that can be used to model a repeatedly measured independent variable (longitudinal predictor) and a fixed outcome variable into prediction models and to evaluate the predictive quality of these models. We focused on methods that could relatively easily be applied by epidemiologists. Thus, for the current study we selected available methods: that can handle longitudinal predictor information and a fixed outcome, are able to assess individual risk prediction, and are easily applied by epidemiologists. We excluded methods based on clustering and random effects as they are not appropriate for the use in individual prediction modelling. As an example we used data from the Terneuzen Birth Cohort[7] in which we applied various methods for the prediction of overweight and body mass
index standard deviation score (BMI-SDS) at the age of 10 years with BMI-SDS measured between 0 and 5.5 years as a longitudinal predictor. In the discussion the properties of the different methods are discussed with respect to the flexibility in handling missing data and differences in timing of measurements, the ability to incorporate all information of the longitudinal predictor and to handle small or large numbers of repeated measurements, and user-friendliness. We focused on methods that could relatively easily be applied by epidemiologists to develop prediction models. As an example we used data from the Terneuzen Birth Cohort[7] in which we applied various methods for the prediction of overweight and body mass index standard deviation score (BMI-SDS) at the age of 10 years with BMI-SDS measured between 0 and 5.5 years as a longitudinal predictor.” (Page 5)

Comment 3

At line 180, the authors say that "In all developed prediction models linear relationships between the predictor variables and the outcome were assumed". However, if I understood correctly, for Method 6 (growth curve) they did in fact evaluate quadratic and cubic functions. In the Results they say that the models predictive quality improved when a quadratic growth curve was used, so I was then a bit confused about the comparison between the growth curve model and the other models. Comparing the other (linear) models to the results of the quadratic model doesn't seem like a fair comparison. I appreciate the authors' desire to demonstrate that this is a better model, but perhaps it would be helpful to provide the results from the linear and quadratic growth curve models for comparison with the other models?

Response: For method 6 we first estimated new predictors, namely the growth curve parameters of BMI-SDS over time (birth to 5.5y) using linear, quadratic and cubic functions. However, the subsequent association between these growth curve parameters (the predictor variables) and the outcome overweight/BMI-SDS at the age of 10 years was indeed assumed to be linear as described in line 180. We understand however that there might be some confusion for readers and have tried to clarify this further in section 2.5. Methods to develop prediction models as follows:

“Secondly, the mean of all measurements and the child-specific growth parameters (i.e. slope coefficient(s)) are entered in a new model to predict the outcome overweight and BMI-SDS at 10 years, again assuming a linear association between the predictor(s) and the outcome as for the previous methods. Here the mean represents the individual’s average value of BMI-SDS from 0 to 5.5 years and the slope parameters[5, 15-18] represent the trend of the individual’s change in BMI-SDS over time. We also added the standard error of the slope parameter from the first step to the model as an indication of fluctuation of the actual measurements around the curve.[15-18]” (page 11.)

In the Results section and Table 2 we have reported the model including the cubic growth curve predictors of method 6 to be compared with the other methods as this produced the model with the highest predictive quality. As described above the linear, quadratic, and cubic associations were only used for developing the predictors themselves and linear associations were assumed between predictors and outcomes just as for the other methods. We considered it to be fair for
each method to compare the best model with those of other methods (we also used the best (5.5y) measurement of method 2 for comparison). We have reported the predictive quality of all the models containing growth curve parameters of different growth curves (linear, quadratic, and cubic; with and without se) in Table A4 of the document Additional file 2, to enable readers to compare the linear and quadratic growth predictors as well in case they desire to do so.

Reviewer: 2

The authors have presented several approaches to utilizing repeatedly measured predictors in prognostic models with binary and continuous outcomes. Using a real-life example based on routinely collected data from child health care services, the authors show that the choice of method depends on a number of practical and theoretical considerations, and therefore can vary in different situations. The main strength of this study is that it presents an important prognostic modelling issue in a systematic and clear manner, and using an easily understandable example.

A few minor points to consider:

Comment 1

The authors applied the broken stick method to deal with the irregular time intervals of the predictor measurements, a common issue in a lot of health-related datasets. Would it be possible to add some more information about the broken stick method: its advantages, limitations etc. - to help readers gain understanding of when this method is most suitable?

Response: We thank the reviewer for the comments. For this article we applied the broken stick method to compute a complete dataset without missing values and with measurements on fixed time points/ages as to be able to apply and compare all methods for modelling longitudinal predictors. The broken stick method itself is not considered a method for modelling repeated predictors for individual risk estimation and, therefore, we did not expand on its general application, advantages and limitations in this article as we considered it to be beyond its scope. The following adaptations have been made to better address the broken stick method in section 2.4.2. Longitudinal predictor:

“To create an example dataset with this longitudinal predictor the broken stick method was applied to all the available growth data from 0-5.5 years of the population for analysis to substitute missing BMI-SDS values while simultaneously generating BMI-SDS values at the specific ages.[16] The broken stick method uses a linear mixed model to describe subject specific growth trajectories at fixed times using a piecewise linear growth curve and is therefore a method that can be applied when the dataset contains irregular spacing of ages.[16] By the inclusion of random effects, subject specific BMI-SDS values are obtained.[14] This example dataset with predictor BMI-SDS at the exact ages of 0d, 3m, 6m, 14m, 2y, 3y, and 5.5y is referred to as the broken stick-data.” (Page 8)
Comment 2

Lines 121-122: The authors mentioned using the LMS-method to convert BMI into BMI-SDS. Would it be possible to briefly explain what the LMS method is about, as reference 9 is not an English language document?

Response: The LMS-method is a method developed by Cole to generate smoothed growth curves and allows for calculation of sds-scores according to these curves. The growth curves are summarized using LMS parameters: Lambda (L) a skewness parameter based on the Box-cox power; Mu (M) median; and Sigma (S) generalized coefficient of variation.[10-12] We have added references to documents giving information on the LMS-method in the English language and made the following adjustment to the text in section 2.3 Data: “Next, all available BMI measurements were converted into BMI-SDS values as described by van Dommelen & van Buuren[9] using the LMS-method. The LMS-method is a method developed by Cole to generate smoothed growth curves by summarising for each age the distribution of the growth data into LMS parameters (Lambda (L) a skewness parameter; Mu (M) median; and Sigma (S) generalized coefficient of variation) and to calculate exact sds-scores from these LMS parameters[10-12].” (Page 7)

Comment 3

Line 402, Table 3: I was not sure what the distinction was between "characteristic of the method" and "strong characteristic of the method", would it be possible to clarify?

Response: We agree that the distinction between the categories might be unclear, in Table 3 we mainly tried to show a quick overview of what each methods strongpoints and weaknesses were by comparing the methods to each other based on the text in paragraph 4.2 and distinguishing the methods capabilities into three categories. We have adjusted the categories in Table 3 into more appropriate categories as follows: the statement is a characteristic of the method: - = no; ± = undecided/neutral; + = yes.(page19-20)