Dear Jose

Manuscript ID BMJ.2017.040627.R1 entitled "Development and validation of QDiabetes®-2017 risk prediction algorithm to estimate future risk of type 2 diabetes: cohort study."

Thank you for reviewing our revised paper and giving us the opportunity to respond to the comments from your statistical consultant.

1) "We used the regression coefficients for each variable from the final models as weights which we combined with the baseline survivor function estimated from the Cox model" – but the baseline survivor function is not obtainable following a Cox model; another method (non-parametric?) method must have been used. Often this is part of the Cox model package in software, but it would be wrong to say it was obtained from the Cox model itself.

Authors response: We have rephrased this "We used the regression coefficients for each variable from the final models as weights which we combined with a non-parametric estimate of the baseline survivor function 1".

2) It is great to see that the authors have included a decision curve analysis, as requested. When this is introduced they say: "The net benefit of a risk equation at a given risk threshold is given by calculating the difference between the proportion of true positives and the proportion of false positives multiplied by the odds of the risk threshold." I find this explanation slightly confusing, as it would benefit from explaining what the odds means (odds of what?). Perhaps say odds DEFINED by the risk threshold value (i.e. value/(1-value)). And add italics to make distinct the two terms such as: "The net benefit of a risk equation at a given risk threshold is given by calculating the *difference between the proportion of true positives* and *the proportion of false positives multiplied by the odds of the risk threshold*."
I could not actually see the decision curves in Figure 4, as it was not uploaded to the system, so I simply trust that the interpretation was correct.
Authors response: We have reworded this as suggested. We don't know what happened about the figure 4 as we were certain we had uploaded it and it is still on bench press when we have re-uploaded the revised paper. can you check bench press is working as intended?

3) thanks for clarifying that the non-linear trends were identified using the complete data. I think this should be noted as a potential limitation, as by ignoring missing data, the trends identified are more likely to be biased and/or have less power to detect genuine trends than the missing data analysis. This reference may be helpful from Morris et al.
http://onlinelibrary.wiley.com/doi/10.1002/sim.6553/abstract
Authors response: We have noted this as a potential limitation in section 5.3 and included the reference from Morris et al2. We also note that this was a practical consideration since the routine can take days to run given the size of the dataset.

4) "Although the new models are more complex than the existing models, this is unlikely to affect the take up of the new models as they are all designed to be calculated automatically based on information recorded in the electronic patient record" – can this really be argued when there was only 16% complete data in the registry for blood glucose measurements, smoking and body mass index? i.e. are these variables really routinely recorded?
Authors response: Whilst smoking and body mass index are routinely recorded, blood glucose is not. However, our statement is still true because (a) Model A does not include blood glucose and (b) Models B and C will only be calculated in people who already have a blood glucose or HBA1C recorded.

5) "additional external validation of Models B and C in datasets with more completely collected data would be valuable before the models are used in clinical practice." – thank you for acknowledging this limitation. I wonder if this could also be included in the abstract conclusion, as otherwise the abstract conclusion reads much stronger than this point would warrant.
Authors response: We agree and have added this to the abstract as suggested.

6) Thank you for adding the calibration slopes, which are reassuringly close to 1 as expected in the overall population. Indeed, they are only very slightly less than 1, which reflects the very small overfitting, a consequence of the huge sample size. Indeed, this results formally backs up the low overfitting hypothesised by the authors in their discussion.
Why do the authors think the calibration slopes are generally less than 1 (between 0.7 and 0.9) for subgroups, however? Why is the model not performing as well in these individuals, and does this suggest there are further improvements warranted in the future?

Authors response: We are not certain why calibration slopes are lower in some of the subgroups but it might reflect different strengths of associations with the predictor variables in these subgroups. Future models could examine interactions between predictor variables and ethnic group.

7) I see another reviewer, whose comments are indeed very insightful, criticised the use of funnel plots. In their defence, they are just used to display the distribution of C statistics; they are not for examining asymmetry as in a meta-analysis. A forest plot would be too long for showing all the C values for each practice. They are calculated on the logit-scale (more symmetric), and then transformed to c scale, and so should be skewed indeed. So I agree that these have been left in.

Authors response: Thank you for this clarification.

8) The full model is still not reported in full in the paper. The authors say that "information on estimates for the full models with baseline survival values is published on the qdiabetes.org website." – surely this is a key part of the TRIPOD guidelines, and should also be in the paper? Is the rationale that you may alter the model over time, and thus don't want the model to be set in stone here? If so, the model can still be provided but with a note of the latest version being on-line?

Authors response: TRIPOD does allow for publication of algorithms in this way (see pages 46 and 473) and the wording we have used has been suggested by Gary Collins in relation to others recent papers published this year in the BMJ. The full model is fully published on the website along with all the other models we have developed and the updates. We are strongly of the view that to include it in the paper would result in people not seeking the website and not finding the most up to date versions and further information. We think it is likely to result in people trying to implement old versions of the algorithms for clinical purposes which would be unacceptable.

9) I do not agree with competing risk models being difficult to interpret. After all, they again allow an equation to provide risk estimates. But I agree that competing risks is probably best considered in a different paper, specifically in the older age group. Indeed I hope, going forward, the authors look at refining their model in the frail populations where competing risks are likely, such that the risk estimates for diabetes are improved.

Authors response: Thank you for this suggestion which we will take on board. Also we have changed the sentence in the Discussion to say that competing risks models can give counterintuitive results rather than being difficult to interpret and added a further reference4.

References

1. Kalbfleisch J, Prentice R. The Statistical Analysis of Failure Time Data. Hoboken2002:114-8.
2. Morris TP, White IR, Carpenter JR, et al. Combining fractional polynomial model building with multiple imputation. Statistics in Medicine 2015;34(25):3298-317. doi: 10.1002/sim.6553
3. Collins GS, Reitsma JB, Altman DG, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD StatementThe TRIPOD Statement. Annals of Internal Medicine 2015;162(1):55-63. doi: 10.7326/M14-0697
4. Andersen PK, Geskus RB, de Witte T, et al. Competing risks in epidemiology: possibilities and pitfalls. Int J Epidemiol 2012;41(3):861-70. doi: 10.1093/ije/dyr213 [published Online First: 2012/01/19]