Author’s response to reviews

Title: Predictors for independent external validation of cardiovascular risk clinical prediction rules: Cox proportional hazards regression analyses

Authors:

Jong-Wook Ban (jong-wook.ban@conted.ox.ac.uk)
Richard Stevens (richard.stevens@phc.ox.ac.uk)
Rafael Perera (rafael.perera@phc.ox.ac.uk)

Version: 1 Date: 08 Jan 2018

Author’s response to reviews:

6442 Turnberry Lane SE
Olympia, WA
USA
Jong-wook.ban@conted.ox.ac.uk

5 December 2017

Ioanna Tzoulaki
Associate Editor
BMC Diagnostic and Prognostic Research

Dear Dr. Tzoulaki,

Thank you very much for inviting us to revise our manuscript. We are pleased to submit the revised version of DAPR-D-17-00021, “Predictors for independent external validation of cardiovascular risk clinical prediction rules: Cox proportional hazards regression analyses”.

We would also like to thank the peer reviewers, Johanna Damen and Kuanrong Li, for evaluating our manuscript and providing very constructive feedback. We tried our best to address the reviewers’ concerns by describing the relevance for our research more clearly, reporting sufficient details of the methods, and discussing the implications of our results in context.

For each question raised by the reviewers, we provided our response and any revision(s) made to the manuscript (with page and line numbers from the PDF file of the revised manuscript created in Editorial Manager). We uploaded a marked-up copy of the manuscript (Revised manuscript) and the updated figures, tables, and supporting materials in Editorial Manager as instructed. We look forward to hearing back from you.
Response to Reviewers

Reviewer #1

This manuscript describes the analysis of predictors for independent external validation of prediction models in the field of cardiovascular disease. A systematic review describing prediction models for cardiovascular disease was used as a starting point to identify prediction models, and using citation tracking, all external validation studies for these models were identified. Key study characteristics and quality of reporting were extracted, and related to the chance of having an independent external validation (i.e. validation by authors who have no relation to the authors of the development study). The manuscript is well-written, methodology is up to date, and studies were identified in a systematic way. I have a few concerns, which I will describe below, split up in major and minor concerns.

Major

1. Although I do believe that a relevant research question is answered in this manuscript, the relevance of studying this is not described in a very convincing way. The authors describe that "CPRs that perform well in these independent external validation studies are much more likely to gain the trust of clinicians." (page 4, line 74) and "These cardiovascular CPRs with no independent external validation may represent research waste because clinicians are unlikely to trust or use CPRs with uncertain generalizability." (page 5, line 80) I believe a reference is necessary for these strong statements. Also, I believe there are other, much stronger, factors that influence the uptake of prediction models in clinical practice, such as the reputation of the researchers developing the model, the extensiveness of the data used to develop the model, and the uptake of the model by clinical guidelines. The pooled cohort equations for example have been introduced in clinical practice, before any independent external validation had been performed. Furthermore, the authors describe that "Understanding why some cardiovascular CPRs are externally and independently validated and others are not, may help researchers create cardiovascular CPRs that are more likely to get independent external validations." (page 5, line 82). I am not sure if this is what we should aim for, since there are already way to many prediction models out there.

Response: We agree with reviewer’s comment. Although we believe it is ideal that clinicians implement CPRs that performed well in independent external validations, we are unaware of any empirical evidence to support these statements yet. As the reviewer pointed out, the uptake of a CPR may be influenced by factors more immediate to clinicians such as recommendations from guidelines. We deleted the statement "CPRs that perform well in these independent external validation studies are much more likely to gain the trust of clinicians" because the key argument
of the paragraph was clearer without it. We replaced the statement "These cardiovascular CPRs with no independent external validation may represent research waste because clinicians are unlikely to trust or use CPRs with uncertain generalizability" with “Without reliable external validations, any use of a CPR in practice cannot be fully evidence-based”. We also revised the third paragraph of the background section to focus the attention to the main objectives of our study, as follows (page 4-5):

Line 79-91: Many CPRs for various cardiovascular conditions have been developed but most cardiovascular CPRs have not been externally validated [2, 8, 9]. CPRs that have been externally validated have often been done by researchers involved in deriving the CPR [2, 6]. Without reliable external validations, any use of a CPR in practice cannot be fully evidence-based. However, it is unknown how often and quickly cardiovascular CPRs are externally validated by independent researchers, or why some cardiovascular CPRs are validated by independent researchers and others are not.

2. Flow of articles is not fully described in figure 1. For example, the number of hits from the citation search and the number of articles assessed based on full text have not been reported. Please also correct the typo in the middle lowest box.

Response: In Figure 1, we tried to illustrate the flow of CPRs rather than articles. However, the flow of articles is not described in the manuscript as the reviewer pointed out. Therefore, we updated the outcome section of methods to describe the screening and selection process more clearly. We also revised Figure 1 to explain the flow of articles and updated the results section, as follows (page 7-8, 11-12):

Line 137-138: In August of 2016, we conducted forward citation searches of all derivation studies of cardiovascular risk CPRs included in the systematic review using Scopus.

Line 143-148: For each cardiovascular risk CPR, one of the authors (JW) screened titles and abstracts of retrieved references in chronological order and full text articles of potentially eligible references were reviewed. This process was continued until the first independent external validation study for the cardiovascular risk CPR was identified. Cardiovascular risk CPRs that had no independent external validation by the time of the forward citation search (August of 2016) were right censored.

Line 225-236: Of 125 cardiovascular risk CPRs we examined, 29 had independent external validation and 96 had no independent external validation (Figure 1). For 33 cardiovascular CPRs, external validations that had no overlapping author with the derivation study were found. However, four of these CPRs only had external validation studies that included authors who had prior co-authorships with the authors of the derivation study.

3. Often, multiple models are derived from the same dataset, such as de Framingham Heart Study. These models are thus developed using a comparable sample size, and also reporting characteristics are comparable because the same authors have published these prediction models. This creates dependency in your data and I believe you should either account for this in your analyses, or discuss the possible impact on your results.
Response: We acknowledge that CPRs derived using the same cohorts might share common researchers and data structure which can lead to similar design, reporting and publication characteristics. In order to account for the influence of using the same cohorts, it is necessary to either use a more advanced design (e.g. multilevel model) or add “cohort” as a predictor in a multivariable analysis. However, we were not able to do either due to the small number of cardiovascular risk CPRs and independent external validations. We accept that this is one of the limitations of our study and revised the strengths and limitation section of discussion, as follows (page 19-20):

Line 397-409: Some predictor variables under study were correlated: for example, the studies published in higher impact journals generally had the larger sample sizes. Further, the observations in the data set may not be fully independent, since a number of derivation studies originated from the same research group (e.g. Framingham Heart Study). The number of available cardiovascular risk CPRs and independent external validations in our data precluded assessing the predictors in a multivariable analysis that could account for these correlations. Therefore, any positive findings in our exploratory analyses should be interpreted cautiously, as hypothesis-generating, until they can be confirmed in multivariable analyses of a future, larger data set.

4. Cox proportional hazards regression model are normally used in a different context (e.g. predicting patients' outcome). Because of terminology used in this manuscript can cause some confusion. How should terms like 'median follow-up time' and 'median event time' (page 11, line 210, 211 + in the abstract) be interpreted in this specific context?

Response: Typically, median (50th percentile) survival time is used to summarize the results of survival analysis [1]. We used the 25th percentile of event time to report our results because of the following two reasons. Firstly, the outcome of interest in our study was time to event (independent external validation) rather than time to survival (not having independent external validation). Secondly, the cumulative probability of event never reached 50% and median event time could not be calculated [2].

Median follow-up time indicates how mature the data are in survival analysis and readers can judge whether the follow-up of participants was sufficient to detect the outcome of interest [1]. For example, a 2-year median follow-up may be adequate for a survival analysis of pancreatic cancer but inadequate for a survival analysis of prostate cancer.

We revised the statistical analysis section of methods (page 10), and results (page 13) to describe these concepts in this specific context. To avoid confusion, we removed the 75th percentile of survival time and the censored survival time from the manuscript.

Line 199-208: The probability for a cardiovascular risk CPR to have an independent external validation was estimated using the Kaplan-Meier method. We reported the 25th percentile of event time because the cumulative probability of event (independent external validation) never reached 50%. The median time from publication of a cardiovascular risk CPR to date of our forward citation search (median follow-up time) was estimated according to the reverse Kaplan-
Meier method [20, 21] to show whether the cardiovascular CPRs were followed-up long enough after their derivations for the assessment of independent external validation.

Line 262-270: Kaplan-Meier estimates of the probability for a cardiovascular risk CPR to have an independent external validation is illustrated in Figure 2. The median time from publication of a cardiovascular risk CPR to date of our forward citation search (median follow-up time) was 118 months (95% CI, 99-130). We found that it took 122 months (95% CI, 91-299) before the probability of a CPR to have an independent external validation reached 25%.

5. The discussion section now contains a lot of repetition from the results, without putting this into a broader context. Please change this.

Response: We revised the summary of results section of discussion to minimize the overlap with results, as follows (page 15-17):

Line 308-356: In this study, we examined the probability of having an independent external validation of a newly developed cardiovascular CPR and explored whether 12 characteristics of derivation, reporting, and publication of cardiovascular risk CPRs are associated with independent external validation. We found most cardiovascular CPRs are not independently validated even 10 years after publication. This greatly limits the value of studies deriving new CPRs, because without strong evidence of validity, CPRs cannot make an evidence-based contribution to clinical practice. We found that CPRs derived in the US were four times more likely to be externally validated by independent researchers although this is heavily influenced by multiple CPRs from the Framingham study. Besides geographic location, larger sample size and publishing in journals with higher impact factor are associated with shorter time to independent validation, as are providing information for risk calculation and internal validation results. These latter two at least are within the control of the derivation study authors and may provide a route for authors to increase the likelihood that their published CPRs will progress further along the pathway to evidence-based practice.

Minor

6. Regarding the systematic review that was used as a starting point in this study: "They excluded a study if it updated an existing CPR" (page 6, line 102). This is not completely true (although I have to admit that this is not clearly reported in the review). Studies that updated an existing model were either categorized as developing a new model, or as external validation + updating. Only studies that added new predictors to a model, to assess the incremental value, were excluded. Please rephrase this in the manuscript.

Response: Thank you very much for clarifying this. We revised the manuscript (page 6) according to reviewer’s comment.

Line 110-113: They excluded a study if it only assessed the incremental value of adding new predictors to an existing CPR, developed a CPR for a venous cardiovascular disease event (e.g.
Wells’ criteria for deep vein thrombosis), or developed a CPR for a specific population such as patients with diabetes, HIV, or atrial fibrillation.

7. “we used a logarithmic transformation to create a continuous variable Log10(sample size)” (page 8, line 155). Strictly spoken, sample size is already a continuous variable, so the logarithmic transformation is not creating a continuous variable. Please rephrase.

Response: We revised the manuscript (page 10) according to reviewer’s comment.

Line 191-194: We applied a logarithmic transformation to the sample size of derivation studies because it had a very skewed distribution.

8. "A post hoc sensitivity analysis showed that the HR for cardiovascular risk CPRs derived in the US (United States) excluding the ones developed by Framingham Heart Study researchers" (page 11, line 228). How many prediction models developed from the Framingham Heart Study were excluded in this analysis?

Response: Of 43 cardiovascular risk CPRs derived in the US, 26 were developed by Framingham Heart Study researchers. We added this information in the results, as follows (page 14):

Line 286-289: A post hoc sensitivity analysis showed that the HR for cardiovascular risk CPRs derived in the US (United States) excluding 26 cardiovascular risk CPRs developed by Framingham Heart Study researchers was 2.46 (95% CI, 0.92-6.61, p=0.0842) compared to cardiovascular risk CPRs derived elsewhere.

9. I believe the UK is part of Europe as well. Maybe it is better to call it 'continental Europe'.

Response: We updated the manuscript (and Table 1) by replacing ‘Europe’ with ‘continental Europe’, as follows (page 10):

Line 194-196: Only a small number of cardiovascular risk CPRs from continental Europe, the UK, Asia, and other geographic areas had an independent validation and these categories were combined.

Table 1: 2. Geographic location, n (%), USA, Continental Europe, UK, Asia, Other, Multiple countries.

Reviewer #2

Many risk prediction rules (mathematical models) have already been developed for diagnosis and prognosis of chronic diseases, including cancer and cardiovascular disease. Instead of developing new models from scratch, researchers should focus more on validating, modifying, and updating the existing models. Before clinical adoption, risk prediction rules should be carefully validated in independent populations. However, a large majority of risk prediction rules have never been
revisited by other researchers for external validation since their debut. This article aims to explain this phenomenon in a quantitative way. The authors' objective seemingly can be understood as building a risk prediction rule to predict the possibility for a prediction rule to be externally validated (if this is the case, shall we assess its own reliability?). From a methodological perspective, one major concern is the study's univariate survival analyses due to the small sample size and the low number of events. However, a multivariable analysis is equally challenging because the selected features of the prediction rules are correlated. For example, it is not difficult to imagine that studies from the United States usually have a large study population and therefore are more likely to be accepted by high-ranking journals. The authors therefore should consider refining their selection of the features. A number of issues need to be clarified as well (see below).

Response: It was not our intention to build a prognostic model to predict time to independent external validation, but rather, to describe the rate of such validations to date, and to explore factors that may influence this. We have edited the background section (page 5) and the strengths and limitations section of discussion (page 19) as follows to clarify our aims.

Line 93-96: Therefore, we estimated the probability of having an independent external validation of a newly developed cardiovascular CPR and explored whether features of derivation, reporting and publication of cardiovascular CPRs are associated with an independent external validation.

Line 402-404: Therefore, any positive findings in our exploratory analyses should be interpreted cautiously, as hypothesis-generating, until they can be confirmed in multivariable analyses of a future, larger data set.

1. p2, line 29-30: it is not clear how the authors calculated the time variable for the CPRs that have never been externally validated (i.e. the definition of the time at censoring, or the survival time as the authors referred to in the next paragraph).

Response: Thank you for pointing this out. We revised the outcome section of methods to provide a clear definition of censoring, as follows (page 7-8):

Line 137-138: In August of 2016, we conducted forward citation searches of all derivation studies of cardiovascular risk CPRs included in the systematic review using Scopus.

Line 143-148: For each cardiovascular risk CPR, one of the authors (JW) screened titles and abstracts of retrieved references in chronological order and full text articles of potentially eligible references were reviewed. This process was continued until an independent external validation study for the cardiovascular risk CPR was identified. Cardiovascular risk CPRs that did not have an independent external validation by the time of the forward citation search (August of 2016) were right censored.

2. p2, line 32: the 12 characteristics should be listed here. The authors should also clearly state that the Cox regression was actually univariate.

Response: We updated the abstract according to the reviewer’s comment, as follows (page 2):
Line 32-36: Using univariable Cox regression, we explored whether characteristics of derivation (design, location, sample size, number of predictors, presentation format, validation in derivation), reporting (participants, predictors, outcomes, performance measure, information for risk calculation), and publication (journal impact factor) are associated with time to the first independent external validation.

Please also note that some words were deleted in the abstract to comply with the journal’s word count limit of 350.

3. p2, line 37-38: Is it a coincidence that the 25 percentile of the event time is equal to the 75th percentile of the survival time? Since there are only 29 event times, consider reporting the range (as well).

Response: Numerically, the 25th percentile of event time should be equal to the 75th percentile of survival time. We revised the statistical analysis section of methods to explain why we reported the 25th percentile of event time to readers and deleted the 75th percentile of survival time to avoid confusion. We provided a 95% confidence interval for the 25th percentile of event time which was 91-299 months. Please refer to the results section (page 13) where we narrated the range of event times.

Line 270-274: A coronary heart disease risk score by Polonsky et al. [24] had the shortest interval of 6 months until the first independent external validation. All independent external validations were done before 142 months except for a coronary heart disease risk score by Wilson et al. [25] which took 299 months until the first independent external validation.

4. p2, line 45-47: consider replacing "incidence" with "likelihood", or simply use "xxx times more likely…".

Response: We updated the abstract according to the reviewer’s comment, as follows (page 3):

Line 47-49: Publishing a cardiovascular risk CPR in a journal that has one unit higher impact factor was associated with a 6% (95% CI, 3-9) higher likelihood of an independent external validation.

5. p2, line 49-50: consider deleting this sentence, since a similar sentence has already appeared in the Background paragraph.

Response: Thank you for the comment. We wish to keep this sentence in the abstract and conclusion as this was one of the main findings of the study. Instead, we revised the background section to make our aims clearer, as follows (page 4-5):

Line 79-91: Many CPRs for various cardiovascular conditions have been developed but most cardiovascular CPRs have not been externally validated [2, 8, 9]. CPRs that have been externally validated have often been done by researchers involved in deriving the CPR [2, 6]. Without reliable external validations, any use of a CPR in practice cannot be fully evidence-based. However, it is unknown how often and quickly cardiovascular CPRs are externally validated by
independent researchers, or why some cardiovascular CPRs are validated by independent researchers and others are not.

6. Among the 125 CPRs, some might be the modified version of the others and they are not independent. It is unclear whether and how the authors handled this situation (at least it should be discussed).

Response: It is true that some cardiovascular risk CPRs in our study are related (e.g. QRISK and QRISK2). We agree that these CPRs often share common researchers and data structure which can lead to similar design, reporting and publication characteristics. In order to account for the influence of modified CPRs, it is necessary to either use a more advanced design (e.g. multilevel model) or add “CPR group” as a predictor in a multivariable analysis. However, we were not able to do either due to a small number of cardiovascular risk CPRs and independent external validations. We acknowledge that this is one of the limitations of our study and revised the strengths and limitation section of discussion, as follows (page 19-20):

Line 397-409: Some predictor variables under study were correlated: for example, the studies published in higher impact journals generally had the larger sample sizes. Further, the observations in the data set may not be fully independent, since a number of derivation studies originated from the same research group (e.g. Framingham Heart Study). The number of available cardiovascular risk CPRs and independent external validations in our data precluded assessing the predictors in a multivariable analysis that could account for these correlations. Therefore, any positive findings in our exploratory analyses should be interpreted cautiously, as hypothesis-generating, until they can be confirmed in multivariable analyses of a future, larger data set.

7. Consider adjusting for the "age" of the CPRs.

Response: According to the reviewer’s suggestion, we assessed the age of cardiovascular risk CPR using a univariable Cox proportional hazards regression. The age of cardiovascular risk CPR does not seem to be associated with having an independent external validation as the HR for ‘age’ was 1.01 (95% CI, 0.97-1.04) per year with a corresponding p-value of 0.678. We were not able to assess this predictor in a multivariable analysis because we only had 125 cardiovascular risk models and 29 independent external validations.

8. Please add the number of studies excluded according to each of the criteria (e.g. p7 line 124-127, p8 line 156-159, etc.).

Response: A total of 46425 references were found in the forward citation searches of the derivation studies. Titles and abstracts of 9683 references were screened and full text articles of 463 potentially eligible references were reviewed. Of these, 435 articles were excluded: 393 were either irrelevant or did not meet one or more inclusion criteria, 20 assessed the risk of a different type of outcome, 15 compared risks estimated by one CPR with another CPR, and 7 used a modified version of a CPR. We revised Figure 1 to include this information.
9. p8 line 150-151: please list in this paragraph the predictor variables and their definitions. This information is important and should be presented in the main text.

Response: We revised the predictors of an independent external validation section by listing the predictor variables and providing their definitions according to the reviewer’s suggestion, as follows (page 8-9):

Line 160-174: Firstly, we reviewed features of CPR derivation that might be important to researchers planning an external validation [13] and selected the following six characteristics of derivation: study design, geographic location, sample size, number of predictors, presentation format, and validation in derivation. A cohort study is an ideal design when deriving a CPR. We determined a case-control design was used when the development of an outcome was verified before the prediction was made [14]. We used the United Nation’s standard country or area codes for statistical use (M49) [15] to define geographical regions where CPRs were developed. Some derivation studies created more than one version of a CPR and we used the predictors included in the full model to define the number of predictors. We determined a user-friendly format was used when a CPR was presented with a simplified format for a risk calculation such as scoring system, chart, or online calculator. A derivation study may include internal or external validation. Internal validations assess a CPR’s reproducibility using techniques such as split sample, cross-validation, or bootstrapping and external validations assess a CPR’s performance in a new population different from that of derivation study [4, 16-18]. An external validation may be included in a derivation study with or without an internal validation.

Line 178-183: We assessed whether authors clearly described participants (eligibility criteria, settings and key characteristics), predictors (including how and when they were measured), outcomes (including how and when they were measured), performance measure (such as discrimination or calibration), and information for risk calculation (a constant and all regression coefficients or a scoring system with probabilities of an outcome needed for calculating individual risks was provided).

Line 185-188: Lastly, we hypothesized that the impact factor of the journal in which CPRs are published might influence the chance of having an independent external validation. We used the impact factor reported in 2015 Thompson Reuters Journal Citation Index. A list of potential predictors and their definitions are presented in Additional file 1.

10. p11 line 225-226: Figure 3 and 4 are not the results of Cox univariate analyses, please correct. Please explain why KM analyses are necessary in addition to Cox regression.

Response: We apologize for the confusion. Firstly, we used the Kaplan-Meier method to estimate the overall probability for a cardiovascular risk CPR to have an independent external validation. The results of this analysis are graphically presented in Figure 2. Then, we assessed whether potential predictor variables were associated with independent external validation using univariable Cox regressions. We estimated Hazard ratios using Cox regression analyses and the results are presented in Table 2. In addition, we compared exposure groups graphically in Figure 3 and 4 using Kaplan-Meier plots, with p-values calculated by log-rank test. To make it clear to the readers, we revised the statistical analysis section of methods, as follows (page 10-11):
We used Cox proportional hazards regression to evaluate the association between potential predictors and the time interval between a derivation of a cardiovascular CPR and the first independent external validation. Hazard ratios (HRs) and their 95% confidence intervals (CIs) were estimated in univariable Cox proportional hazards regression models. In addition, we graphically compared exposure groups by plotting Kaplan-Meier estimates for the probability of an independent external validation for each level of categorical variables and each tertile of continuous variables.

We also discovered that p-values presented in Figure 3 and 4 were from Cox regression analyses. We corrected the errors by replacing these p-values with ones from log-rank tests.

11. The authors mentioned that the proportional hazards assumption was not violated. However, there are obviously some crossing curves in the KM plots.

Response: We tested the proportional hazards assumption by plotting scaled Schoenfeld residuals. When there is little or no difference between groups (HR close to 1), it is easy for survival curves to cross by chance alone, especially if event rates are low. Figure 3-1 for example does not provide evidence of non-proportionality, which may be because proportionality holds or because there are too few events in one group to detect a deviation.

For the most part, the proportional hazards assumption appears to hold. But the p-value for the test of nonproportionality was 0.0343 for ‘description of outcome’. The interpretation of this result was complicated. It was difficult to conclude that the proportional hazards assumption was violated because 12 hypotheses were tested. When correcting for multiplicity (e.g. Bonferroni method), we would need to apply a much lower threshold for p-value for each hypothesis test. At the same time, it was difficult to firmly rule the violation out. Therefore, we concluded that ‘no clear violation was detected’.

12. p11 line 226-227, p12 line 233-235: please report the direction and the effect size of the associations.

Response: We revised the univariable analysis section of results according to the reviewer’s comment, as follows (page 14):

Line 283-286: Three of six features of CPR derivation studies assessed were associated with having an independent external validation: geographic location (HR for USA = 4.15, 95% CI 1.89-9.13), sample size (HR = 2.32, 95% CI 1.37-3.91), and validation in derivation (HR for internal validation = 1.73, 95% CI 0.77-3.90).

Line 291-295: Of six reporting and publication related features analyzed, reporting information for risk calculation (HR = 2.65, 95% CI 1.01-6.96) and publishing the derivation study in a journal with higher impact factor (HR = 1.06, 95% CI 1.03-1.09) were associated with having an independent external validation.
13. p13 line 256-261: it should be kept in mind that some studies are more likely to be accepted by high-IF journals, for example those with a large sample size or from certain countries.

Response: Cardiovascular risk CPRs derived outside of the US and using a larger sample size had a tendency to be published in journals with higher impact factor in our sample. Some degree of correlations between predictor variables are inevitable in regression analyses. For example, a cardiovascular risk CPR form the American Heart Association and American College of Cardiology (2013 ASCVD Risk Calculator) includes Age, Diabetes, Sex, Race, Smoker, Cholesterol, Systolic Blood Pressure, and Treatment for Hypertension as predictors. It can be easily expected that older, diabetic patients, and smokers have higher systolic blood pressure. Multivariable analysis can assess the predictor variables simultaneously while controlling for the influence of each other (unless the predictors are highly correlated). Unfortunately, we were not able to conduct a multivariable analysis due to the constraint in the number of cardiovascular risk CPRs and independent external validations. We revised the strengths and limitation section of discussion, as follows (page 19-20):

Line 397-409: Some predictor variables under study were correlated: for example, the studies published in higher impact journals generally had the larger sample sizes. Further, the observations in the data set may not be fully independent, since a number of derivation studies originated from the same research group (e.g. Framingham Heart Study). The number of available cardiovascular risk CPRs and independent external validations in our data precluded assessing the predictors in a multivariable analysis that could account for these correlations. Therefore, any positive findings in our exploratory analyses should be interpreted cautiously, as hypothesis-generating, until they can be confirmed in multivariable analyses of a future, larger data set.

14. p14 289-291: again the censoring mechanism remains unclear: for CPRs received external validation, the follow-up ends at the date of validation, then when did the follow-up end for CPRs that have not been externally validated?

Response: We revised the outcome section of methods to provide a clear definition of censoring to readers, as follows (page 7-8)

Line 137-138: In August of 2016, we conducted forward citation searches of all derivation studies of cardiovascular risk CPRs included in the systematic review using Scopus.

Line 143-148: For each cardiovascular risk CPR, one of the authors (JW) screened titles and abstracts of retrieved references in chronological order and full text articles of potentially eligible references were reviewed. This process was continued until an independent external validation study for the cardiovascular risk CPR was identified. Cardiovascular risk CPRs that did not have an independent external validation by the time of the forward citation search (August of 2016) were right censored.

15. Page 12 line 248 to page 14 line 285: this part to some degree is a repetition to the Result section, consider deleting some words.
Response: We rewrote the summary of results section of discussion to minimize the overlap with results, as follows (page 15-17):

Line 308-356: In this study, we examined the probability of having an independent external validation of a newly developed cardiovascular CPR and explored whether 12 characteristics of derivation, reporting, and publication of cardiovascular risk CPRs are associated with independent external validation. We found most cardiovascular CPRs are not independently validated even 10 years after publication. This greatly limits the value of studies deriving new CPRs, because without strong evidence of validity, CPRs cannot make an evidence-based contribution to clinical practice. We found that CPRs derived in the US were four times more likely to be externally validated by independent researchers although this is heavily influenced by multiple CPRs from the Framingham study. Besides geographic location, larger sample size and publishing in journals with higher impact factor are associated with shorter time to independent validation, as are providing information for risk calculation and internal validation results. These latter two at least are within the control of the derivation study authors and may provide a route for authors to increase the likelihood that their published CPRs will progress further along the pathway to evidence-based practice.

16. Table 2 last row: replace median(IQR) with one unit increase.

Response: We updated Table 2 by deleting “median (IQR)” so that it is consistent with presentation of other continuous variables (sample size and number of predictors).

References

