Author’s response to reviews

Title: The current application of the Royston-Parmar model for prognostic modelling in health research: a scoping review

Authors:

Ryan Ng (ry.ng@mail.utoronto.ca)

Kathy Kornas (kathy.kornas@utoronto.ca)

Rinku Sutradhar (rinku.sutradhar@ices.on.ca)

Walter Wodchis (walter.wodchis@utoronto.ca)

Laura Rosella (laura.rosella@utoronto.ca)

Version: 1 Date: 30 Oct 2017

Author’s response to reviews:

October 30, 2017

Dear Dr. Debray,

Re: DAPR-D-17-00017

We would like to thank the reviewers for their thoughtful review of our manuscript, “The current application of the Royston-Parmar model for prognostic modelling in health research: a scoping review.”

We have provided a response to each of the items identified by the reviewers, and have indicated corresponding changes in the manuscript in the highlighted text. There were some additional changes made to the manuscript to improve the flow of the scoping review after incorporating the feedback from the reviewers (not highlighted).

On behalf of my co-authors, I would like to thank you for considering our manuscript.

Sincerely,

Ryan Ng
Reviewer #1:

The authors present a 'scoping' review of flexible parametric models for prognostic modelling. This review should be of interest to readers and I'm not aware of any similar reviews having been published. However, I feel that the manuscript still needs considerable work before publication.

1. I felt the rationale for conducting the scoping review could be better described. The authors state 'the baseline hazard function is estimated smoothly and therefore these models are attractive for prognostic modelling'. Please discuss any potential benefits for prognostic modelling in more detail. Why are they attractive? Please also consider how prediction models are often presented and used, often for a particular time point e.g. mortality within 5 years. Would flexible parametric models still be of benefit?

- Thank you for the insight. We realized the need to provide further details about prognostic models and to clarify the rationale of the study. We have re-written the manuscript to incorporate this feedback, specifically:

  - Description of prediction model presentation (Background, paragraph 1 on pg.3, line 71): “While predictions are often presented for a specific time point, prognostic models based on survival analysis allow time-specific predictions for any time point during the follow-up period, which improves the usefulness of the model for different contexts”

  - Benefits of the RP model versus the parametric survival model and the Cox PH model (Background, paragraph 6 on page 8, line 187): “In the PH context, Royston-Parmar model can be thought of as a hybrid approach of the parametric survival model and the Cox PH model. Modelling the baseline log cumulative hazard function as a restricted cubic spline is similar to the parametric survival model, but the complexities of the baseline function can be modelled without sacrificing model fit. The restricted cubic spline can also be reported mathematically as differences in cubes via basis functions. This expression permits the continuous estimation absolute and relative measures of effect and their uncertainty, which is an advantage versus the Cox PH model combined with an estimator. Continuous estimation of absolute measures of effect (e.g. hazard rates, differences in survival probability, standardised survival function, population-averaged survival function) increases the applicability the prognostic model because time-specific predictions can be made by the user”

  - Clarification of the rationale of the study in the abstract: (Abstract - Background, on page 2, line 35)
Clarification of the rationale of the study in the main text (Background, paragraph 6 on page 9, line 195):” Despite these advantages of the Royston-Parmar over the Cox PH model (with an estimator) and parametric survival models for prognostic modelling, the use of the model in health research for prognostic models is unknown. Thus, a scoping review of the application of the Royston-Parmar model for prognostic models in health research was conducted to document its current use, to raise awareness of the model, to identify gaps in current reporting, and to offer recommendations for future reporting.”

2. Although I generally understood what the authors meant, I found the language throughout a bit loose. Please be more specific in the descriptions of statistical methods and concepts. Some examples are given from the background section, but much of the manuscript is written in this way:
   a. * Line 31: "The flexible parametric survival model was developed in 2002". The paper the authors refer to was published in 2002, although the authors also acknowledge an earlier paper in 2001.
      
      Thanks for noting this discrepancy. We have re-written to be specific that the RP model was published in 2002 in Statistics in Medicine and referenced accordingly.
      
      The change can be found on page 2 in the Background section of the Abstract starting on line 30.

   b. * Line 32: Authors state that the R-P model is defined by a RCS used to model the baseline log cumulative hazard function when proportional hazards are of interest. However, these models are broader than this, they can incorporate non-proportional hazards using time-dependent effects and can also model on the proportional odds scale etc.
      
      This is an excellent point. We have made changes to clarify this point, specifically, the abstract background has been re-written to mention other scales and that they can incorporate non-proportional hazards.
      
      The change can be found on page 2 in the Background section of the Abstract starting on line 35.

   c. * Line 68-71: "When regression modelling is used for prognostic models, survival analysis methods are employed because this class of statistical procedures account for the relationship
between the predictor(s) and the outcome(s), as well as the time until the outcome(s) occur." This is rather long-winded and not strictly true as logistic regression is often used for outcomes occurring within a short time interval (e.g. if within a few hours, 30 day mortality etc.).

- We agree that this is not always the case and have re-written to say that survival analysis is a common regression model type for prognosis. In addition, we have shortened the sentence to enhance clarity (see page 3, line 68, first paragraph of Background):

“When regression modelling is used for prognostic models, survival analysis is the most common regression model type because it accounts for the relationship between the predictor(s) and the outcome(s), and the time until the occurrence of the outcome(s).”

d. * Lines 79-82: "The baseline hazard function is not estimated in the Cox PH model because the regression parameters are estimated by maximizing a partial likelihood function, thus treating the baseline hazard function as a nuisance parameter." Other way around, baseline hazard function was seen as a nuisance parameter, therefore the partial likelihood function was maximised as a way of removing this "nuisance" parameter.

- Thanks for the correction. The sentence has been clarified to say that the baseline hazard is treated as a nuisance parameter, so it cannot be estimated (Background, paragraph 2 on page 4 line 84).

e. * Lines 82-84: "Because the baseline hazard function is not estimated, absolute measures of effects can only be predicted at the time points of observed events." Again, I don't think this is quite right. As you go on to discuss, you can use a non-parametric estimate of the baseline hazard function. Also, it is not that you can only predict at the time points of observed events, but rather that the probability is assumed to remain constant until the next event time.

- We appreciate the nuance the reviewer picks up here. We have updated this description to clarify that the survival function is a step function, and is estimated conditional on a particular time point (Background, paragraph 2 on page 4, line 87).

f. * Line 85: "methods to estimate survival post hoc..." Please be more specific with 'survival' here, survival probabilities or survival function.
We have re-written to be specific that the survival function is being estimated post hoc (Background, paragraph 2 on page 4, line 90).

g. * Line 88: please be specific about the 'limitation' you are referring to here.

We agree with this comment. The word limitation has been replaced with a more detailed description that the distance between steps can be mitigated (Background, paragraph 2 on page 4, line 98).

h. * Line 94: not 'smoothed', just estimating a function.

Thanks for the suggestion. We have removed the “smoothed estimation” term and replaced with “direct estimation”. We have also revised the paper to avoid using the term smooth and instead used terms like continuous function and step function where appropriate (Background, paragraph 3 on page 5, line 100).

3. A definition of what is meant by 'prognostic model' for the review would be good. For some people, this includes models predicting future onset of disease in healthy individuals, while others might not consider this prognostic. Are you interested in models that predict outcome probabilities for individuals or additionally make inferences at a higher level e.g. net survival for groups of individuals compared to the general population? Models can have very different aims.

Thank you for the comment. We have clarified that our definition of prognostic models includes both individual risk prediction and population risk prediction (Background, paragraph 1 on page 2, line 63). We agree that the purpose of models applied to individuals versus groups of people may differ, but we did not distinguish between the two in our review. Further we have provided an example of a prognostic model used for population risk prediction (Background, paragraph 1 on page 4, line 77).

4. Also, do you only consider multivariable models or also include univariable models?

Univariable and multivariable were both considered. We have now specified this in the methods (Methods – Study inclusion Criteria, paragraph 1 on page 12 line 270).
5. It might help to split the background into two sections. One with the background and rationale for the study and a separate section detailing the flexible parametric models and extensions. Depending on the definition of prognostic models mentioned in the above point, the extensions may/may not be relevant.

- We appreciate this suggestion. We have revised the introduction to better clarify our study rational and model descriptions as per earlier comments and believe this helps address any concerns about the background section.

6. I am curious about the choice to exclude prognostic factor studies (associational studies as referred to in the paper) which may still be interested in absolute estimates of hazard or survival, while including studies investigating relative survival, net survival and cure models for which inferences are generally at a population level rather than at the individual level.

- Thank you for noting this – and indeed we had these discussions when doing our review. We opted not to include associational studies because the aims of these differ significantly from prediction, and many aspects, such as model building and performance metrics accordingly differ. Thus, we decided to limit our review to prognostic studies in order to focus our discussion and provide more concrete guidance for application. Specifically, the aim of the study was to examine how flexible parametric survival models were used to construct prognostic models, either for individual risk prediction or population prediction. In the revised manuscript, we have made it explicit that we are interested in individual and population risk prediction (see comment 4). As long as the aim of the study was prediction, we included it in the review, and we did not exclude prognostic factor studies. For example, the included study by Csordas et al. (2016) had the objective of looking at the improvement in risk prediction of all-cause mortality in patients undergoing transcatheter aortic valve replacement when including Th1-type inflammatory markers (neopterin, KTR and Phe/Tyr ratio).

7. Why were studies looking at methodological developments in FP models excluded even if the methods were applied to real data, while studies providing an empirical comparison of methods were included (ref 27, 28)?

- We agree that the case could be made to include studies looking at the methodological developments in FP models, however we wanted to focus the review on how flexible parametric survival models were being applied to conduct prognostic model research in terms of model development/validation, and importantly, we aimed to identify gaps where improvements could be made for future reporting of FP models. Thus, to address these
aims, we excluded papers that focused on methodological developments, including those
that demonstrated model developments using real data.

8. If more than one prognostic model was presented in an article, were details summarised for
the article rather than for each model?

- Thanks for the question. Information for all prognostic models incorporating flexible
parametric survival models were included. Specifically, the main model was identified and
its information was abstracted; where relevant, the other models were presented as
sensitivity analysis in relation to the main model. We have clarified this point in our
description of abstraction items (Methods – Data abstraction and synthesis, paragraph 1 on
page 13, line 286).

9. I am aware of at least one study that used flexible parametric modelling to develop and
validate a prognostic model that would be relevant but is not included in this scoping
thromboembolism following treatment for a first unprovoked venous thromboembolism:
systematic review, prognostic model and clinical decision rule, and economic evaluation.
Health technology assessment. 2016; 20: 1-190. This raises concern as to how thorough the
search strategy was or how well the inclusion/exclusion criteria were applied.

- Thanks for raising this issue of missing articles with our search strategy. This was our
concern as well when we designed the search strategy. We consulted with a librarian to try
and resolve this issue because there are no MeSH terms, or subject headings for flexible
parametric survival models. To mitigate this problem, we tried to be as inclusive as possible
with our key word search, but a limitation of key word searches is that it only returns articles
that contain the term in the title or abstract. We also used a couple of search strategies not
commonly-used in scoping reviews, namely a citation search and looking at the personal
reference list of authors involved in the methodological development of the flexible
parametric survival model. A citation search looks for key article(s) in the reference list of
all other documents. In this review, we used the foundational articles related to the creation
and methodological development of flexible parametric survival models.

We checked against our original list of articles, and the VTE study was not picked up in any of
our search strategies. The article was missed unfortunately, even with our multiple search
strategies. The abstract did not mention flexible parametric survival models, Royston-Parmar
models, or any other potentially related key word (see Appendix), and the citation search did not
pick up the article even though it did cite the original RP paper. We have expanded on the
limitations of our search strategy in the discussion section, and mentioned the exclusion of this paper to illustrate this limitation (Discussion, paragraph 7 on page 23, line 537).

10. Lines 215-218: Why were searches in google scholar limited to the first 200 hits and other searches limited to the first 30 search results? Please provide reasoning for why this was thought to be appropriate.

- Thank you for raising this important point. The Google searches were limited because the hits are sorted by relevancy. The cut-off limit was different between the two searches because the relevancy of the hits was different between the two searches. Nonetheless, we acknowledge that important articles may have been missed, and have elaborated on the limitations of our search strategy in the discussion section (Methods – Search Strategy, paragraph 2 on page 10, lines 232 and 238).

11. Line 259: Do the authors mean 'abstraction' or rather 'extraction'?

- Thanks for the question. Abstraction and extraction are similar in that both are processes of retrieving information from the document, but there is a slight difference in how the data are retrieved. Extraction occurs when the information is copy in a word-for-word manner from the document, whereas abstraction occurs when summaries of the information are retrieved from the document (Methods – Data abstraction and synthesis, paragraph 1 on page 12, line 281).

12. Line 310 (and earlier): Might be better to refer to d.f. for splines or be more specific about whether you are referring to internal knots or including boundary knots.

- Thank you for the suggestion. The relationship between the degrees of freedom and internal knots has been clarified (Background, paragraph 4 on page 7, line 146), and we have been more specific throughout the manuscript about when we refer to internal knots; for example: (i) Background, paragraph 4 on page 6, line 13); (ii) Table 2.

13. Page 13: Flexible parametric survival model specifications section. Number of studies given but not clear out of how many as not always out of 12.
This section has been re-written to clarify when the results were for ten studies or all twelve studies (Results – Royston-Parmar model specifications, paragraph 1 on page 14, line 329).

14. Page 14: When defining calibration and discrimination statistics, please be clearer in your definitions. E.g. discrimination (yates) slopes - separation in what?

• We have clarified that the discrimination slopes looks for separation in the average predictions between subjects with and without the outcome (Results – Application of the Royston-Parmar model…, paragraph 1, page 15, line 359).

15. Line 352: Say that ref 36 provided full prediction from the prognostic model including the restricted cubic spline function. Having seen the article, they report coefficients for each of the spline terms, however this doesn't make it useable as these correspond to spline terms created in the dataset. Please discuss this more fully.

• Thank you for the insightful comment. We have revised the results accordingly to say that no studies provided enough details for full prediction or to validate the model (Results – Application of the Royston-Parmar model…, paragraph 2, page 16, line 378)

• We have re-written the introduction and discussion to address this important suggestion. In the introduction, we added a section about the basis functions of the restricted cubic splines, including derived variables (Background, paragraph 5, page 6, line 143).

• We also included a section in the discussion describing how no studies reported enough information to allow other researchers to predict using the restricted cubic spline. We go on to discuss that the exact placement of the knots are needed to calculate the derived variables of the restricted cubic spline and that the coefficients of the spline are also needed (Discussion, paragraph 3 on page 18, line 435).

• We also added in some reporting guidelines for RP models that discusses how complete information should be reported (Discussion, paragraph 5 on page 21, line 497).

16. Line 385: Studies claimed that FP models improved model accuracy. It would be good to know if this was in terms of apparent, internal or external model performance.

• We have added these details into the manuscript, specifically that 2 studies used apparent validation and 1 study used internal validation (Discussion, paragraph 2, page 19, line 418).
17. Line 390-392: Authors suggest reporting comparisons of performance measures between flexible parametric models. How do you suggest this is done considering it is the external validity of a model that is of most interest?

- We agree that emphasis needs to be placed on external validation as well. We have included a statement in the discussion that further research is needed to understand how flexible parametric survival models perform during external validation because this is important to understand generalizability (Discussion, paragraph 2, page 19, line 426).

18. Line 400: see earlier comment on reporting RCS coefficients. Still need spline terms created in data. This is one of the remaining difficulties with RP models for prediction. It is very difficult to report the full model in a way that predictions can be made over time. Perhaps this should be discussed.

- We have addressed this valid point in a previous comment (#16).

19. Line 414: Authors mention that more emphasis should be placed on the ability to model time-dependent effects, however these are also difficult to report (as well as interpret). Please discuss this more fully.

- Thank you for the suggestion. We have discussed that modelling time-dependent effects may be more important for studies with longer follow-up where violations of the PH assumption are more likely to occur (Discussion, paragraph 6 on page 21, line 508).

- As well, we have listed that both splines need to be fully reported in order to obtain estimations (Discussion, paragraph 6 on page 21, line 511).

20. Paragraph beginning on line 426, discusses reporting of number and position of knots. However, simply having this information does not make the model any more usable. See earlier comment.

- We have addressed this valid point in a previous comment (#16).
Reviewer #2:

This paper conducts a systematic review of the use of the flexible parametric Royston-Parmar survival model, specifically within a prognostic modelling context. The paper is well written, with a good explanation of the fundamental model, and the core methods papers captured, which will serve as a good reference for future prognostic model development. I only have minor comments:

1. More emphasis is needed on the useful measures that can be obtained following a fully parametric model, such as survival differences, standardised survival etc. The usefulness of the stpm2 command is mainly due to its extensive post-estimation prediction tools.

   • We have revised the manuscript to include examples of absolute measures of effect that can be estimated from the flexible parametric survival model in the introduction (Discussion, paragraph 8 on page 22, line 842).

2. P4 line 99, add that lambda and gamma must both be > 0

   • Thanks. This phrase has been added (Background, paragraph 3 on page 5, line 105).

3. P5 l119, please add linear with respect to log time

   • Thanks. This detail has been added (Background, paragraph 4 on page 6, line 126).

4. P6 spline description. Please add more details, i.e. that the default knot positions imply that the function is linear beyond the boundary knots, and the spline function assumes continuous 0,1,2 derivatives. Please also make it clear when talking about total number of knots, and the difference between internal and boundary knots.

   • We have added more details about the restricted cubic spline including: boundary and internal knots, assumes continuous 0,1,2 derivatives and the functions are linear beyond the boundary knots (Background, paragraph 4 on page 6, line 133). The relationship between the degrees of freedom and relationship to number of knots is also included (Background, paragraph 4 on page 7, line 146).
5. P6 1147, "correctly specified". We can never know if it is correctly specified. Please change to something like "well specified".

- Thanks. We have changed to well specified (Background, paragraph 4 on page 7, line 164).

6. There are 2 R implementations that I know of, within flexsurv and Rstpm2. Rstpm2 also allows penalised splines, which should also be mentioned.

- Thank you for mentioning these implementations. We have included a section in the discussion that describes how flexible parametric survival models can be modelled in Stata, R and SAS (Discussion, paragraph 8 on page 22, line 842).

7. P15 ;357, how was the PH assumption tested?

- We have provided more details about the PH assumption in the manuscript, including the type of tests, and how non-proportionality was dealt with (Results – Application of the Royston-Parmar model…, paragraph 2 on page 16, line 382).

8. More emphasis should be placed on encouraging the reporting of knot locations, spline coefficients, whether the splines were orthogonalised (please mention this as a separate aspect of the model in the paper). Authors should also be encouraged to provide code used to fit the reported model.

- Thank you for these suggestions. We wholeheartedly agree, and have included a section in the manuscript providing recommendations on how the flexible parametric survival model should be reported (Discussion, paragraph 4 on page 20, line 462; and Discussion, paragraph 5 on page 21, line 497).

- We mention orthogonalization as a way to remove the correlation between derived variables in the spline function (Background, paragraph 4 on page 7, line 144).

9. Table 2 reports number of knots (df-1). Please clarify how you define df, are you including the intercept term?
Thanks for the questions. The degrees of freedom have been defined in the main text (Background, paragraph 4 on page 7, line 147) and the table 2 header has been re-labeled to describe it as the number of interior knots (Table 2, page 30).

Reviewer #3:

This is a nice paper which will help the Royston-Parmar (RP) model to become much more popular in practice. I have a minor point with the title and one major point with each of the two main parts of the paper.

1. The title is too general. You consider only the Royston-Parmar model and that should become obvious from the title. Other approaches can be considered as 'flexible parametric survival models'.

   Thank you for the suggestion. The title has been changed to: The current application of the Royston-Parmar model for prognostic modelling in health research: a scoping review. We have also re-worded the manuscript to reflect that we are interested in the Royston-Parmar model, and that it is a type of flexible parametric survival model.

2. The RP model is suitable for flexible modelling but your description is extremely positive without pointing to any potential drawback (what about necessary assumptions, numbers and placement of knots), and without providing any empirical evidence (eg by simulations) for these statements. In your review only 2 of 12 papers mentioned a minor limitation - but that is the summary from a highly selected group of modelers who decided to use the RP approach. Is there any paper which provides more evidence that RP is much better than Cox and other competing approaches? Always or in more specific situations? In the review you excluded such studies (see (ii) on page 10)).

   At least you should modify some of your statements to better reflect current knowledge. Better would be a section about the studies comparing RP with other models.

   Thank you for the suggestion. We agree with your suggestion and have included a section summarizing current knowledge about the number and placement of knots. We framed this information in terms of recommended reporting guidelines for transparent modeling practices and mitigating overfitting (Discussion, paragraph 4 on page 20, line 474).
3. Your review provides some ideas of the weaknesses of using RP models. Number of knots vary strongly without giving any reason for it and the placement is often not given (see Table 2). The latter may only be a problem of reporting or the analysts may have played around with numbers and placement of knots and give results from the best fitting model. Overfitting may be a (severe) disadvantage of the RP model. In times of increasing awareness of transparency, good reporting is an important part of each paper. Do you have ideas how to report an RP model in a suitable way? You were surprised that number and placements of knots were often not reported (426). I am surprised seeing a summary of survival analyses without seeing the effective sample size (number of events) and the number of variables in the study (univariable analyses, 10, 20 or more variables in a study? Any variable selection method used? (see Table 2)). Some (not all are relevant for your summary) of the items of the REMARK and TRIPOD guidelines provide help for better reporting.


• Thank you for the helpful suggestion. We have added two paragraphs in the discussion regarding recommendations about how to approach modelling flexible parametric models and how to report the models so they can be used for validation by other researchers. (Discussion, paragraph 4 on page 20, line 462; and Discussion, paragraph 5 on page 21, line 497).

• We’ve also updated Tables 1 (page 29) and Table 2 (page 30) to include additional details that were abstracted from each study, including the modelling process, such as the number of events (i.e. effective sample size, model selection, and number of variables in the final model).