Manuscript ID RSM-07-2020-0103

Response to Reviewers

Dear Mrs. Rucker,

Thank you for giving us the opportunity to submit a revised version of the manuscript “Modelling continuous variables and treatment effect differences in IPD-MA: an introduction to splines” for publication in the Research Synthesis Methods Journal. We appreciate the time and effort that you and the reviewers dedicated to providing feedback on our manuscript and are grateful for the insightful comments on and valuable improvements to our paper. We have incorporated most of the suggestions made by the reviewers. Those changes are highlighted within the manuscript. Please see below, in grey, for a point-by-point response to the reviewers’ comments and concerns. All page and line numbers refer to the revised manuscript file with tracked changes.

Reviewers' Comments to Author:  
  
**Reviewer: 1**  
Comments to the Author  
Modelling continuous variables and treatment effect differences in IPD-MA: an introduction to splines  
  
This is potentially a very useful paper, because the use of splines for estimating interactions and non-linear trends has not been considered in much detail in the IPD MA literature. Indeed, here the authors discuss and illustrate a few different approaches to modelling splines in this context, some of which (like p-splines) I have not seen discussed before. Relevant articles that have come in the IPD setting that use splines include Gasparrini et al.1 2 and, in particular for interactions, Riley et al. (2020).3 The latter uses restricted cubic splines, and a multivariate two-stage IPD MA framework, but does not consider p-splines or GAA as the authors do here.  Hence, the paper truly does add useful added value.  
However, I do have some comments for consideration. In particular, some important comments about knot locations and the one-stage methods, and how the examples actually inform the reader about the best methods.

1. It would be helpful to set the scene in the introduction for how this work builds on previous literature, specifically the Riley and Gasparinni work, and also the work of White.

As suggested by the reviewer, we now added how our work is building on previous literature.

“One reason why splines are not often used in IPD-MA may be that the available guidance is limited. White et al. [24] did compare pointwise meta-analysis and multivariate meta-analysis techniques in presence of non-linear associations, but used fractional polynomials instead of splines. Gasparrini et al. [23] described the use of B-splines in combination with multivariate meta-analysis. They mention that multivariate meta-analysis may be combined with other approaches to account for non-linearities but don’t provide details. Riley et al. [25] described multivariate meta-analysis and one-stage mixed effects modelling. However, most of the examples were limited to either linear associations or a combination of restricted cubic splines and multivariate meta-analysis.”   
See page 4 lines 12-18

1. In the discussion, the authors say: “Other researchers have also drawn the attention to the importance of modelling non-linear associations in IPD-MA [49, 50]. These studies focused on estimating relative treatment effect functions whereas we focused on estimating the absolute risk differences.” – this is an important point, but does not come across in the abstract or the title. Also, the Riley et al. article does discuss absolute risk prediction conditional on treatment.

As suggested by the reviewer, we have adapted the abstract and title to include absolute risk prediction conditional on treatment.

Changes in the abstract:

“To introduce modelling of nonlinear absolute treatment effects using restricted splines,

B-splines, P-splines and Smoothing splines and different pooling methods in IPD-MA.”

Also, we adapted the title. The new title is:

“Nonlinear treatment effects in IPD-MA: an introduction to modelling absolute risk differences using splines.”

1. In the example, the authors compare the methods. But how do we know which is ‘best’? We don’t know the truth, so we can’t make any judgments about this, and whether splines are smooth or not, does not make them more correct. However, the argument might be that they are more realistic.

We agree with the reviewer that since we don’t know the underlying truth in the empirical example, the discussion of the results should be limited to objective findings. The goal of the empirical example was to show “plausibility” results aka how realistic the pooled curves from different methods are. Therefore, we added the following sentence:

“Since this is an empirical example, the underlying true associations are not known and we cannot draw firm conclusions with respect to the appropriateness of the different approaches. However, we show the pooled curves and compare them with regard to their smoothness and width of confidence intervals, and report convergence issues if any.”   
See page 21 lines 25-28.

1. The examples are way too brief, and do not explain or illustrate the key benefits/limitations of the approaches as it stands. We need more guidance for the reader. I am not suggesting a simulation study, but illustration of key pros and cons needs to be clearer. Perhaps a series of key messages could form sub-heading in the results of the examples.

We would like to thank the reviewer for this comment. As suggested, we now emphasize the key benefits/limitations of the approaches in Section 5.4, “Properties of the pooling methods”. To give more structure to the section and also aid the reader, we highlighted the main advantages and disadvantages in bold.

**“Pointwise meta-analysis: robust and flexible but non-smoothness may occur”**See page 18 line 8.

**“Multivariate meta-analysis: efficient if specified “correctly”, but lacks robustness and flexibility”**See page 18 line 37.

**“GAMM can handle different study domains and sample sizes, whilst producing smooth pooled regression curves, but careful modelling is required”**

See page 19 line 9.

1. In the abstract, it is not clear whether 1-stage or 2-stage IPD MA methods are being used. This should be clarified.

As suggested by the reviewer we revised the abstract and we included the terms two-stage and one stage.

“We describe splines and illustrate their performance in an artificial single study. We describe two-stage methods based on pointwise and multivariate meta-analysis and a one-stage method based on generalised additive mixed effects models (GAMMs) to pool the results of multiple studies. “  
See abstract

1. ‘IPD-sets’ is an odd word

Thank you for pointing this out. We changed the term IPD-sets into terms more appropriate to the context:

* “examples” in heading page 4 line 27
* “scenarios” page 22 line 20 and page 28 line 3 and
* “data-set” page5 lines 4,6,8, page 11 lines 28, 29, 30,   
  page 22 line 32 and in the Appendix page 42 table.

1. In 5.1 the authors say “For the spline approaches, we positioned knots per study as follows. For the restricted cubic splines, we placed 5 knots, following Harrell’s suggestion to use the 5%, 27.5%, 50%, 72.5%and 95% quantiles of BMI” – I do not understand the rationale for this, because this makes the knot positions different in each study. However, for the pointwise method, perhaps this does not matter? The authors need to clarify this.

We would like to thank you for pointing this out. We added the following sentence to clarify that in pointwise meta-analysis we are allowed to fit different models per study:

“Since in pointwise meta-analysis we are pooling the predicted outcomes we can apply any good fitting model per study. As a consequence, different modelling techniques may be applied across studies, including linear models, fractional polynomials, and splines of different degrees and with different knot specifications. During this stage we may use criteria to the find the best fitting model per study, e.g. Aikaike information criterion, GCV or likelihood ratio tests (possibly with a nominal significance level larger than 0.05 as proposed by Sauerbrei and Royston [14].”  
See page 13 lines 26-31.

1. Penalised methods – more explanation for this is needed. I can understand why this is important for individual risk prediction, to reduce overfitting. However, this introduces bias in parameter estimates, and therefore is not welcome if we want to obtain unbiased estimates from our meta-analysis. The authors need to emphasise this, and it raises a point similar to before: is the aim to of the IPDMA to estimate the treatment-covariate interactions (i.e. relative treatment effects, and function) or is it for individualised risk prediction? For the latter, I can understand why penalised methods are important, but not the former. We need more clarity on this issue in the Introduction and throughout (see earlier comment about need to put in context of existing work and where this paper adds value).

We would like to thank the reviewer for this suggestion as we were aware that in penalized splines the beta coefficients may be biased (due to penalisation) and therefore we should avoid pooling them with multivariate meta-analysis. However, in the first version we tried to combine penalized splines with multivariate meta-analysis, which in all cases failed to converge and we reported it.

Now as suggested by the reviewer we removed the combination of penalized splines and multivariate meta-analysis from both the illustrative examples and the empirical example and stated that multivariate meta-analysis may not be used with penalised splines:  
  
“The coefficients of the penalised splines are biased due to penalisation and pooling them may be problematic and show biased results. However, we may use unpenalised spline approaches such as described in section 4.1.”  
See page 14 lines 36-38

We also make clear in the introduction which is the estimand we focus on:   
“The estimand we focus on is the absolute risk difference between interventions conditional on a continuous covariable, as we consider this measure the most relevant for clinical decision making. Our goal therefore is to explain and illustrate how to model conditional absolute treatment effects using the aforementioned spline approaches in scenarios with multiple studies, using artificial data-sets for illustration.”   
Page 4 lines 19-22.

1. There are lots of Section numbers, and I think this is off-putting, which some sections having only one or a few paragraphs. Consider making more cohesive.

In order to increase the cohesiveness of the manuscript we adopted a two-level hierarchy in the section numbering. We removed single subsections and 3rd level subsections.

Natural or restricted splines Page 8 line 1

B-splines Page 8 line 35

Properties of regression splines Page 9 line 35

P-splines Page 11 line 7

Smoothing splines Page 11 line 31

Properties of penalised splines Page 12 line 11

1. The one-stage model introduced in Section 5.3 is not actually given as an equation, and a crucial issue is not mentioned: that of separating within-trial and across-trial relationships. This is an important point made in the literature already, for example the articles by Fisher and Riley in the reference list. Amalgamating these relationships can often lead to bias, in terms of the actual relationships at the within-trial level.

We removed all formulas from the text. For one-stage modeling and described the difference between fixed (common), random and stratified effects. We included in the description of the one-stage method a clear notification that one-stage modeling may be prone to ecological bias and therefore we may consider the use of stratification or centering the potential effect modifier.

“Researchers may choose for the effects of each basis function to be either fixed (common), random or stratified [49]. The fixed (common) effect assumption is that the effect of the basis function is identical across all studies. The random effects assumption is that the effect of the basis function comes from a distribution of effects, while the stratified effects assumption is that for each study the effect of the basis function may be different and is estimated per study. The fixed (common) effect may be modelled straightforward by including the basis function as they are. Stratified effects can be modelled by including an interaction of the basis function with the (categorical) clustering variable (e.g. study). Random effects can be modelled by penalising the interaction of the basis function with the clustering variable, as Wood [22, 50, 51] and Kimeldorf and Wahba [52] have shown.

Depending on the estimand of choice and the assumptions researchers wish to make they may use any combination of the above assumptions for their model. Note that interaction terms included in one-stage mixed effect models may be prone to ecological bias and amalgamate the within and across study effects [45, 53–55]. To avoid this, two methods have been proposed. One approach is to stratify by study all or some of the main effects including at least the treatment effect [25]. Another approach is to center the covariate X about its study‐specific mean (X\_j ) ̅ creating a new variable Z=X-X ̅j. Subsequently, include Z, X ̅j, and the interaction of Z with the treatment in the one-stage model [25, 56].”  
See section 5.3 page 17 lines 5-19.

We also added in the discussion the lack of illustrating ecological bias in our examples as a limitation.

“Second, we did not illustrate the performance of the pooling methods in scenarios with ecological bias. Modelling choices that avoid ecological bias in presence of non-linear associations still require further research and were thus outside the scope of this article.”   
See page 23 lines 28-30

1. Again, this comes back to the aims of the paper, in terms of what is the estimand of interest and building on the literature.

We have included in the introduction a description of the estimand of interest, that is the absolute risk difference between interventions conditional on the continuous covariable.

“The estimand we focus on is the absolute risk difference between interventions conditional on a continuous covariable, as we consider this measure the most relevant for clinical decision making.”  
See page 4 lines 19-21

We changed the Title to:  
“Nonlinear treatment effects in IPD-MA: an introduction to modelling absolute risk differences using splines.”

Also, we describe in the objective of the abstract that our goal is   
“To introduce modelling of nonlinear absolute treatment effects using restricted splines, B-splines, P-splines and Smoothing splines and different pooling methods in IPD-MA.”

1. To be honest, I feel like the one-stage model is a distraction, as the main parts of the paper are about the 2-stage approach. I would suggest that the one-stage parts are mentioned in the discussion, but that the authors focus of the paper is about the use of splines in a two-stage setting.

While we appreciate the reviewer’s suggestion, we respectfully disagree. We believe one-stage approaches are being used more and more frequently [9], and have some clear advantages. For instance, our own group (Debray et al. [10]) point out that one-stage approaches may be preferable as they use a more exact statistical approach and account for parameter correlation. Belias et al. [11] suggest the use of one-stage IPD-MA especially when the outcome is binary and the sample size is limited. Also, one of the results of the current manuscript is, that one-stage modeling turns out to be a valuable tool. To our knowledge there is no literature on the performance of one-stage models combined with splines. We believe that keeping GAMMs in our manuscript will draw the attention of researchers to further investigate their possibilities.

1. Indeed, another major concern is that the one-stage approach mentions using percentile again for the knot locations, but this will lead to uninterpretable curves, as the value of the knot position will mean different things in different studies.

“Note that we positioned the knots over the full domain of BMI and that no data-augmentation nor extrapolation was needed.”  
See page 15 lines 40-41

1. “The main advantage of multivariate meta-analysis is, if the fitted curves are correctly specified and the ranges of X are the same across studies, that multivariate meta-analysis appears to be more efficient with narrower confidence intervals than the other pooling methods” – this is not correct, as the multivariate approach does not need the same ranges in each study. Some ranges can be estimated in all studies, and others in a subset of studies. This is a major advantage of the multivariate approach, and allows borrowing of strength. Strategic knot position is essential, however. This was discussed in the recent Riley (2020) article.

We would like to thank the reviewer for this comment. We now performed multivariate meta-analysis also in scenarios with different BMI ranges, following the method described by Riley et al. To do so we performed data augmentation as a preliminary step in both pointwise and multivariate meta-analysis, so that the results will be comparable. GAMMs already use the full domain of X in all studies therefore it wasn’t necessary to perform data augmentation.

“Data augmentation as described by White et al. [47] and Riley et al. [25] refers to the generation of pseudo data beyond the per study boundaries of X, with minimal weight and arbitrary outcome. Note that in multivariate meta-analysis careful specification of the knots is required as convergence issues may occur during the second stage.”  
See page 15 lines 3-5

“We applied multivariate meta-analysis in combination with regression splines in all three scenarios. To do so we performed data augmentation as a preliminary step [25, 48] in the second and third scenario. This way all studies had curves estimated over the full range of BMI.”  
See page 15 lines 35-38

1. STATA should be Stata

Thank you for pointing this out. We changed STATA to Stata.

1. “Pointwise meta-analysis suffered from overfitting during the first stage.” – how do we know this? We don’t know the truth.

Thanks for pointing this out. We removed the sentence as it was indeed confusing for the reader, see page 17 lines 28-29.

1. Separate curves are produced for control and treatment groups in the figures. However, I am concerned that this breaks randomisation. Can the authors clarify this please? Are curves produced for each of control and treatment groups separately? In previous papers (e.g. Gasparinni and Riley) the (relative) treatment effects are combined, which preserves randomisation. However, this would not be the case if control and treatment groups are analysed separately (even if their correlation is accounted for, the borrowing of strength could allow randomisation to be broken, especially if there is imbalance in the numbers of patients in control and treatment groups across studies, and/or the baseline risk varies)

Thanks for pointing us to this unclarity. In section 3 we note that:

” A commonly applied approach to investigate treatment effect measure modification is to model the interaction of a potential effect modifier with the treatment. In case of non-linear associations, a spline transformed version of the continuous modifier can be used. Therefore, we model the association between the modifier and the outcome by including a spline transformed version of the modifier, both as main effect and in interaction with the treatment.”   
See page 5 lines 34- 35 and page 6 lines 1-3

In all cases our analyses were including interaction terms and not stratified per treatment group. Therefore, the randomisation was not broken. To make the text clearer we removed the word stratified from the description of the results and from the description of pointwise meta-analysis as it was confusing.   
See page 13 line 40, page 16 line 19 and 35, page 20 line 30.

I hope these comments are constructive for the authors going forward.

Reference List  
1. Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. Stat Med 2012;31:3821-39.   
2. Gasparrini A, Armstrong B. Multivariate meta-analysis: A method to summarize non-linear associations. Stat Med 2011;30:2504-06.  
3. Riley RD, Debray TPA, Fisher D, et al. Individual participant data meta-analysis to examine interactions between treatment effect and participant-level covariates: Statistical recommendations for conduct and planning. Stat Med 2020;39(15):2115-37.  
4. White IR, Kaptoge S, Royston P, et al. Meta-analysis of non-linear exposure-outcome relationships using individual participant data: A comparison of two methods. Stat Med 2019;38(3):326-38.  
  
  
**Reviewer: 2**  
Comments to the Author  
Your topic is highly relevant and you propose a new combination of the three parts  
1.      Modelling continuous variables  
2.      Estimating treatment effect differences for a continuous variable  
3.      Conducting an IPD meta-analysis for functions

For the first part you propose to use splines and the paper concentrates on spline modelling. Approaches based on subpopulation treatment effect pattern plots (Wang et al., 2016) and fractional polynomials (Kasenda et al., 2016, DOI: 10.1136/bmjopen-2016-011148) are published. The latter paper is not mentioned.

Thank you for pointing this out. We added the Kasenda paper. See reference 16 on page 20.

1. The paper has several weaknesses. It is dominated by lengthy and detailed explanations of various spline approaches (chap 4 has more than 6 pages) whereas the other issues are brief. What do we gain from all the spline formulas and the text? ​ This information can be easily found in various spline papers and books.

Thank you for this suggestion. We moved a large part of the splines section to the appendix. In the current version, we kept two paragraphs that shortly describe each spline method and one paragraph describes the specifications we used in the illustrative examples. Now section 4 is only 3 pages long.

A table providing an overview of potential advantages and disadvantages in the context of treatment effect modification and meta-analysis would be helpful.

We added a flowchart to show the relation between the approaches and which problems each method is accounting.   
See Figure 5

* + Not all spline approaches can be used with the three meta-analysis approaches.

Thank you for this. Although pointwise and GAMMs can be easily applied with any combination of splines, indeed multivariate meta-analysis seems to be problematic when combined with penalized splines. In penalized splines the beta coefficients may be biased (due to penalisation) and therefore we should avoid pooling them with multivariate meta-analysis. However, in the first version we tried to combine penalized splines with multivariate meta-analysis, which in all cases failed to converge and we reported it.

Now as suggested by the reviewer we removed the combination of penalized splines and multivariate meta-analysis from both the illustrative examples and the empirical example and stated that multivariate meta-analysis may not be used with penalised splines:

“The coefficients of the penalised splines are biased due to penalisation and pooling them may be problematic and show biased results. However, we may use unpenalised spline approaches such as described in section 4.1.”  
See page 14 lines 36-39

“Finally, since multivariate meta-analysis pools the coefficients estimated during the first stage it may not be compatible with approaches where penalisation to those coefficients is applied.”   
See page 19 lines 6-7

* + It would be most helpful to provide an overview of the properties of the pooling methods in a table. What are the main issues of the pooling approaches and what does that mean in the context of complicated spline functions? Some of the approaches need (simple) formulas for the function in each study. Obviously, that is difficult with some of the spline approaches.

We have updated Table 2. Now we include along with the main advantage and disadvantage, information on which splines may be combined with each pooling method.   
See page 33 Table 2.

* + The design of the simulation study is too simple. A strong non-linear effect of the control and a relatively similar type of function with a large difference in the treated patients (Figure 1). The five functions are even very similar in the heterogeneous IPD-set (Figure 2).

Thank you for this suggestion. It would have been interesting to also explore scenarios where the functional shapes are different and highly heterogeneous. However, corresponding to our goal to provide an introduction to splines in IPD-MA, we considered simple but realistic illustrative examples, where the heterogeneity of the regression lines would be limited to I2 less than 40%.

We also added your comment in the Discussion, where we describe the features we used in our paper,

“In our illustrative examples we introduced three features into our generating mechanisms. First, the association of mortality risk with BMI was simple and realistic as it was based on previously published papers [27, 28]. Second, we generated IPD-MA scenarios suitable to pool as the between-study heterogeneity of the regression curves was limited to I2 less than 40%. Finally, in the second and third scenarios we generated per study different boundaries for BMI, in order to illustrate the performance of the pooling methods in scenarios where the regression curves have limited overlap.”  
See page 23 lines 17-20.

* + The issue with different BMI ranges needs more discussion. Why? What happens in a real study? Will you discard parts of the data? I can’t imagine any real situation without relevant differences in the range of the predictor.

We agree with the reviewer’s assessment. We changed our analysis and now we perform all pooling methods in the full range of BMI. We did the same for the empirical example. To do so we performed data-augmentation as suggested by Riley et al. and White et al.

“In case of different ranges of X across studies, the use of common positions for the knots may leave some coefficients inestimable in some studies and meta-analysing them may cause complications [23]. A solution is to conduct data augmentation as a preliminary step. Data augmentation as described by White et al. [47] and Riley et al. [25] refers to the generation of pseudo data beyond the per study boundaries of X, with minimal weight and arbitrary outcome.”   
See page 15 lines 1-3.

And reported that we used data-augmentation in our illustrative and empirical examples.

“We applied multivariate meta-analysis in combination with regression splines in all three scenarios. To do so we performed data augmentation as a preliminary step [25, 48] in the second and third scenario. This way all studies had curves estimated over the full range of BMI.”

See page 15 lines 35-37.

Also, since in multivariate meta-analysis the ranges of age across the studies need to be the same, we performed data-augmentation as a preliminary step.

See page 20 lines 42-43.

1. The empirical example is unhelpful. Not all of the five studies can be used for all approaches and you provided results based on a different number of studies. You considered situations with 5, 3 and 2 studies. How can you compare results? For one analysis you even need to truncate age to ‘approximately between 0.5 to 6 years old’ (p15, l 19).

We agree with the reviewer that it might be unhelpful to show results of the empirical example for different numbers of studies per pooling approach. Therefore, we adapted the analyses: we created pseudo data in order to be able to include all five studies in all three pooling methods.   
We also agree with the reviewer, that we don’t know the underlying truth in the empirical example. Therefore, we considered that we need to point out that the results are limited to objective findings. We added the following sentence:

“Since this is an empirical example, the underlying true associations are not known and we cannot draw firm conclusions with respect to the appropriateness of the different approaches. However, we show the pooled curves and compare them with regard to their smoothness and width of confidence intervals, and report convergence issues if any.”   
See page 21 lines 25-28.

* + A suitable table may help to better understand which combinations of your three parts are possible.

We thank you for pointing this out. We included in Table 2 a new row indicating whether the pooling methods can be performed with all or a subset of the splines methods.  
See Page 33 Table 2

* + Figures showing treatment effect functions in the example look horrible (Figs 14-15) and need further explanation (Fig 16, functional forms are very different in a and b). Improve the figures resolution and size.

Thank you for this observation. As the empirical example is meant to illustrate the advantages and disadvantages of the splines and pooling methods in a real example, we decided to change the analysis. We restricted the age range of the children to at most 9 years old, as acute otitis media mainly occurs in younger children. This biological phenomenon was also reflected in our study data, which had hardly any data for older children, and therefore resulted in extremely wide confidence intervals for the older children. In addition to that, we performed data augmentation as a preliminary step in pointwise and multivariate meta-analysis.

“We used data of children till 9 years old, as AOM above 9 years is seldom and we had a limited number of children over that age (only 15). From the remaining 5 studies, one study (Appelman et al.) had a limited number of events (children with fever/ear pain) and for some age-bilateral AOM combinations no events at all. Therefore, we followed different strategies across the pooling methods for this study.”

Also, we improved the resolution and size of all figures.

1. I doubt that any of the co-authors checked this version carefully. Based on a quick view it is obvious that author names are missing in refs 21 and 35; references 22 and 23 are identical. The abbreviation AOM is used in the example (p15, l35) but not introduced. This indicates that the writing of parts of the paper is also insufficient and I am uncertain whether formulas are correct.

All authors did read and check the paper several times, so we are very sorry that these mistakes did happen anyway. We have corrected the citations. The abbreviation of acute otitis media was already introduced in the introduction section, but we introduce it again in the empirical example section as we agree that after 16 pages a reader may have forgotten.  
See page 16 row 10

Kind regards,

On behalf of all authors,

Michail Belias