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

Michail Belias , Maroeska M. Rovers , Jeroen Hoogland, Johannes B. Reitsma, Thomas P.A. Debray, Joanna IntHout

Health Evidence, Radboud university medical center, Mailbox 133, P.O. Box 9101, Nijmegen 6500 HB, The Netherlands

Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, PO Box 85500, 3508 GA Utrecht, The Netherlands

Cochrane Netherlands, University Medical Center Utrecht, Utrecht University, PO Box 85500, Utrecht, 3508 GA, The Netherlands

Corresponding author

# Introduction

One of the main goals of an individual participant data (IPD) meta-analysis (MA) is to investigate whether treatment effect differences are present, and how they are associated with patient characteristics [1]. Investigating a treatment effect conditional on a continuous variable (e.g. BMI or age) may be challenging, since often the association between the outcome and the continuous variable is not a-priori known.

A naïve but often used approach is to ignore the functional form of the outcome-covariable association by either categorisation of the continuous variables or by assuming a linear effect. Categorisation leads to loss of information, and may result in model misspecification, reduced power, inflation of the type I error rates, and biased results [2–5]. Linear modelling may be problematic, as it ignores potential non-linearities and may thus also lead to biased results. Modelling treatment effect differences whilst accounting for non-linear functional shapes may provide the opportunity to accurately make inferences whether a patient should be treated or not.

An alternative approach is to first estimate the functional shape of the associations and then investigate potential treatment effect differences [6]. So far, a variety of methods that account for non-linear functional shapes have been proposed [7–16]. In this manuscript we consider the use of splines to account for potential non-linearities since splines allow to capture both non-linear main effects and non-linear treatment- co-variable interaction effects, without the need to pre-specify their functional form, thus avoiding mis-specification. Furthermore, it is possible to choose between fully parametric, semi-parametric or even non-parametric approaches, and to allow for penalisation and clustering. Thus, splines are very flexible to address a great variety of fitting problems. The most relevant types of splines in order to investigate possible differences in treatment effect for clinical decision making are based on piece-wise polynomials. The most commonly used spline approaches are restricted splines [17], B-splines [18, 19], Smoothing splines [20] and P-splines[21].

Splines are being used in single studies, both in intervention and prediction studies. However, the use of splines in IPD-MA is less common. In IPD-MA, splines may be modelled in either one or two stages. In a one-stage approach, a generalised additive mixed effects model (GAMM) [22] is applied. GAMMs fit a generalised additive model using covariables with or without spline transformation, while adjusting for within-study clustering of the participants. In two-stage IPD-MA, at the first stage an appropriate statistical model is fitted per study including spline transformations. Subsequently we can either extract study specific estimated regression lines with their standard errors or the coefficients of the estimated functions along with their variance-covariance matrix. At the second stage, we may either pool the extracted regression curves through pointwise meta-analysis [14] or pool the extracted coefficients through multivariate meta-analysis [23].

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 relations, but used fractional polynomials instead of splines. Gasparrini et al. [23] have 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. The estimand we focus on is the absolute risk difference between interventions conditional to 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 to illustrate the spline approaches. We describe the various spline approaches mentioned above and their application in IPD-MA using pointwise meta-analysis [14], multivariate meta-analysis [23], and GAMMs [22], and we provide the corresponding R-code. We also describe the results of the aforementioned spline and pooling methods using an empirical individual participant data-set, investigating the effect of antibiotics in children with acute otitis media (AOM) [26].

# Illustrative examples

In order to illustrate the aforementioned spline and IPD-MA approaches we generated artificial data to mimic a previously reported nonlinear association between BMI and mortality [27, 28]. We considered the case where the outcome is binary, but note that splines may be used to other types of outcomes such as continuous and time-to-event outcomes. For the control group we generated a J-shaped association showing increased mortality for underweight and overweight participants, while obese participants show a BMI-dependent exponentially increasing mortality risk. For the experimental group we assume a levelled J-shaped association, where the underweight and overweight participants have approximately the same risk as the participants with normal weight without treatment, while the obese participants show again a BMI-dependent exponential increase in their risk, but less steep. The functional shape for the associations in the control and experimental group is quadratic and quartic respectively, see Figure 1.

Figure 1. approximately here

To illustrate the performance of splines in IPD-MA we generated three distinct IPD-MA scenarios, each consisting of 5 RCTs with 500 participants per study. In the first scenario, which we refer to as the heterogeneous data-set with equal BMI ranges, the association between BMI and mortality is different across studies, see Figure 2, but the distribution and ranges of BMI are the same. In the second scenario, which we refer to as non-heterogeneous data-set with different BMI ranges, the parameter values of the association for both the treated and control group are identical across all studies, but the range of available BMIs varies across studies (see Figure 3). In the third scenario, which we refer to as the combined data-set with different BMI ranges and between study differences in the mortality risks, both the ranges of BMI and the association of BMI with the mortality risk vary across studies, see Figure 4. Exact equations are given in the Appendix.

Figure 2. approximately here

Figure 3. approximately here

Figure 4. approximately here

# Treatment effect (measure) modification

“Treatment effect modification”, also called “treatment effect measure modification” [29, 30] is the phenomenon where the effect of a treatment varies across the levels or strata of a certain variable. We prefer the term “treatment effect-measure modification” since effect modification may be present for one measure (e.g. risk difference) but not for another (e.g. odds ratio, risk ratio) [29–33]. The scale in which the results are presented is therefore a vital first decision.

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 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. In case of a binary outcome like mortality, a logit link function can be used in the model. In order to calculate the absolute risk difference between the treatment arms, we back-transform the predicted outcome per treatment arm with the inverse logit function. To calculate the confidence intervals of the absolute risk difference we use the proposals of Newcombe [34].

# Spline approaches in a single study

In a setting where the association between an outcome and a continuous variable X is nonlinear, one of the options for curve fitting is to use splines. Splines represent a continuous variable as a linear additive combination of (often) local parts which each have a simple mathematical form and are known as basis functions. Numerous basis functions have been developed involving various mathematical forms, such as polynomials, radials and Fourier series. However, for our goal basis functions based on piecewise polynomials are most relevant. As the term piecewise implies, the range of X is divided into intervals, using cut-offs called knots. Within each interval a d-degree polynomial of X is used to model the association between the outcome Y and X. These polynomials are connected across adjacent intervals. This way, instead of estimating a global non-linear association over the full range of data, we estimate the linear association between the outcome and a local transformation of X. Third degree polynomials are the simplest basis functions with good smoothness properties, as they are twice differentiable with a non-constant second derivative. Therefore, they are commonly applied in splines, but polynomials of higher or lower degree can also be used. In practice, smoothing with polynomials of a higher than third degree will usually give similar results to the third degree [35].

Two important choices that have to be made, in addition to the degree of the basis functions, are: (1) the number and the position of the knots and (2) whether a penalty should be applied. Splines calculated with the use of knots and without penalties are often called regression splines. The most commonly used regression splines are restricted or natural splines [36] and B-splines [18, 37]. Splines where a penalty is applied are called penalised splines. The most commonly used penalised splines are P-splines [21] and Smoothing splines [20]. A short summary of these four types of splines is presented below. Details are presented in the Appendix. Figure 5 shows how the aforementioned splines methods are associated with each other.

Figure 5. approximately here

## Regression splines

## **Introduction**

In order to understand the rationale to use splines, we shortly describe polynomial regression and truncated power series and their limitations. Generalised linear models (GLMs) are often used to model the association between an independent variable X and an outcome Y. In case of non-linear associations, transformed versions of X can be used instead of X. For instance, the statistical model for a GLM with link function g and with a d-degree polynomial of X is:

However, global functions, over the full range of X, may have poor fit near the boundaries due to the instability of the estimated polynomial in this area. To avoid these issues, polynomials fitted within subsets of X, also called piecewise polynomials, may be preferred to global functions. However, piecewise polynomials when fitted in two consecutive intervals, will show different predicted values at the boundaries of the intervals (*i.e.* at the knots), thus their functional shape will be discontinuous. For this reason, we may use restrictions to “connect” interval-specific polynomials. One convenient solution is to fit a global polynomial, and model the deviations from this globally defined shape within truncated parts of X. A disadvantage of truncated power series is that they can still show erratic behaviour near the boundaries of X.

## **Natural or restricted splines**

A solution to the erratic behaviour of truncated power series is to fit a first-degree global polynomial (linear) over the full range of X and model the deviations from it. Note that as for all splines also restricted splines need to be continuous over the knots. Therefore, specific truncated transformations of X are used, which we present in the Appendix. The number and location of the knots may be based on clinical knowledge or on descriptive statistics. For instance, Harrell suggests the use of quantiles and advocates that four knots in most cases are adequate [17, 38], see Table 1. Depending on the available sample size and required complexity of the functional shape we may use a different number of knots.

In our single study example, we used restricted cubic spline transformations of X both as main effects and as interactions with the treatment. Following Harrell’s suggestion, we placed 5 knots at values corresponding to 5%, 27.5%, 50%, 72.5% and 95% quantiles of X. In Figure 6(a), we present the predicted mortality risks per treatment arm, conditional on BMI, along with the 95% confidence intervals. Subsequently, we calculated the effect of the treatment conditional on BMI, by calculating the conditional risk for the control minus the conditional risk for the treated, see Figure 7(a). To calculate the absolute risk difference’s 95% confidence intervals, we followed the proposal of Newcombe, see section 3. Note that in our artificial data the boundaries and distribution of BMI-values for the treated and control group are the same. In practice, this may not be true and knots may be placed at different positions for the main effects and the interaction terms..

## **B-splines**

B-splines are another commonly applied regression spline approach. They are based on a parametrisation of polynomial splines. Their basis functions are based on step functions called zero-degree basis functions equal to 1 within an interval and 0 otherwise. More details are given in the Appendix.

In Figure 6(b) we show the results of the B-splines approach for the simulated single study data. In order for B-splines and restricted cubic splines to be comparable in terms of the degrees of freedom, we used 2nd degree B-spline transformations of X both for the main effects and for the interactions with the treatment. We used 4 equidistant knots; 2 inner knots at BMI values 25.65 and 32.84 plus 2 at the boundaries 18.5 and 40. Subsequently, we calculated the effect of the treatment conditional on BMI, see Figure 7(b), similar as for the restricted cubic splines.

## **Properties of regression splines**

The main advantages of regression splines are their simplicity and the fact that they can be represented by a formula. As a consequence, the estimated regression coefficients can be reported and used in further analysis, e.g. meta-analysis. Also, both restricted splines and B-splines are straightforward extensions of GLMs, with low computational cost.

B-splines of any degree are calculated based on zero-degree basis functions defined within each interval. This provides great local support and numerical stability [35]. Restricted splines fit a “basic” linear model and estimate deviations from it, which provides limited local support, since some basis functions are defined over the whole range of the variables, while others are not. Restricted splines with κ knots require degrees of freedom, while dth degree B-splines require degrees of freedom [35]. Furthermore, the model fit of regression splines (e.g. as quantified using the log-likelihood) depends on the number and position of the knots, thus careful modelling is required to avoid overfitting. In some occasions clinical knowledge or descriptive statistics may be used to define the knots, but in others it is unclear how many knots should be used and where they should be placed. To avoid this disadvantage, penalised splines have been proposed, which we describe in section 4.2.

## Penalised splines

The two commonly applied penalised splines that we discuss, P-splines and Smoothing splines, increase the number of knots to a large set (usually, 10-40) or even to be equal to the number of observations. This way they circumvent the problem of choosing the number and positions of the knots. Since estimating one parameter for each observation would clearly lead to a perfect fit and thus generate functional shapes with extreme variability, penalised splines introduce in their optimisation functions a penalty term () multiplied by a non-negative , often called a tuning parameter. As the term “tuning” implies, changing the value of changes the magnitude of the penalisation.

Penalised splines may circumvent the problem of knot selection, but at a cost. By using a penalty in their optimisation function, they introduce bias in their estimate in order to obtain a more stable solution. Furthermore, in both P-splines and Smoothing splines the tuning parameter must be specified. Too high or too low values of may lead to over- or undersmoothing respectively. Several approaches have been proposed in order to determine the “optimal” , such as Akaike’s information criterion AIC [39], “leave one out” generalised cross-validation (GCV) [40] or mixed-effects modelling [22]. These processes are automated in most of the statistical packages. Briefly, when using the AIC, a series of models fitted with different values are compared and the one with the lowest AIC is selected. “Leave one out” GCV is an iterative process, the algorithm goes as follows: 1) one observation is omitted 2) a model is fitted 3) using the model a prediction of the omitted value is generated and 4) the distance between the observed and predicted value is calculated. This procedure is repeated for each observation and for a series of values. The that minimizes the GCV, also minimizes the sum of the squared distances, i.e. the GCV score, is selected. In Bayesian/mixed effects modelling approach the penalty term is estimated in a similar way as a random effects parameter [22]. More details are shown to the Appendix.

## **P-splines**

A specific type of penalised splines, P-splines, proposed by Eilers and Marx [21], is a penalised version of B-splines, using a specific penalty term based on the sum of p-order differences between the coefficients of two consecutive intervals . The first order differences are defined as follows: , but Eilers and Marx propose the use of second order differences, which are the first order differences of the first order differences. Note that the degree of the underlying B-splines may be different from the order of the differences. A common combination is that of a third-degree B-spline with a second order difference. Using a penalty based on a zero-degree order difference results in the ridge penalty [41].

P-splines are based on equidistant knots. It is possible to use a knot sequence that is not evenly spaced; but in this case, weights need to be introduced [22, 35]. As P-splines with non-equidistant knots are rarely used in practice we don’t consider them in this article. In our single study example, we used P-spline transformations of X for both main effect and interaction terms. We used 17 equidistant knots; 15 inner knots plus the boundaries, while the parameter was selected through a ‘leave one out’ GCV process as described above. In Figure 6(c), we present the resulting mortality risks per treatment arm conditional on BMI, along with the 95% confidence intervals. Subsequently, the effect of the treatment conditional on BMI, calculated as the difference between the two curves in Figure 6(c), is presented in Figure 7(c).

## **Smoothing splines**

Smoothing splinesare another member of the family of penalised spline methods. Similar to P-splines the idea is to increase the number of knots, but this time to be equal or approximately equal to the number of observations. O’ Sullivan [41] suggested that a penalty based on Reinsch’s integral of the second derivative of , where is a cubic spline, multiplied by a tuning parameter, has good smoothing properties. This results in the following penalty term for Smoothing splines: .

In our single study example, we use Smoothing spline transformations of X for both the main effect and the interaction term, while the parameter is selected through a ‘leave one out’ GCV process as described above. In Figure 6(d), we present the resulting mortality risks per treatment arm conditional on BMI, along with the 95% confidence intervals. The effect of the treatment conditional on BMI is presented in Figure 7(d).

Figure 6. approximately here

Figure 7. approximately here

## **Properties of penalised splines**

Penalised splines are penalised extensions of the regression splines based on many knots. For instance, P-splines are B-splines with an order difference penalty applied on the coefficients, while Smoothing splines are cubic splines with a derivative based penalty. The penalisation contributes to the selection between a complex and a simple model in a similar way as in other well-known penalised GLM approaches, for instance LASSO, ridge or elastic-net. Penalised splines reflect ones belief that the predicted regression lines are more likely to be smooth than not. Therefore, their main advantage is that they are more likely to show smoother functional shapes as compared to unpenalised splines. Another great advantage is that they circumvent the need to specify the positions and the number of knots, which in most cases are not known beforehand and may need to be estimated.

Penalisation also affects the inference, due to the bias-variance trade-off. For instance, the coefficient estimates are subject to a smoothing bias, therefore their interpretation may be problematic. Note that this issue does not necessarily apply to the predicted outcomes. A related issue is that the degrees of freedom have to be modified to account for the penalisation. Wood [42] suggests the use of effective degrees of freedom of a model. Effective degrees of freedom are calculated using the Welch‑Satterthwaite approximation formula and can be used to compare models fitted with different types of splines. Note

# Individual participant data meta-analysis using splines

In the previous sections we focused on estimating nonlinear main effects and interactions with the treatment in a single study. As trials are rarely powered to investigate effect modifiers, exploring non-linear effects in a single study may often be problematic or yield very wide confidence intervals. Splines are especially useful in the context of an IPD-MA, where they enable the statistical modelling of complex relationships such as non-linear associations [44]. They can be applied in a two-stage or one-stage meta-analysis approach. We apply the methods on three IPD-MA scenarios of 5 studies each. In the first scenario, the regression lines are heterogeneous whilst the BMI ranges are the same across studies. In the second scenario, the regression lines are homogeneous but the BMI ranges are different across studies. In the third scenario, both the regression lines and the BMI ranges are different across studies.

## Two-stage pointwise meta-analysis

In the first stage of two-stage pointwise meta-analysis, as proposed by Royston and Sauerbrei [14], we may fit an appropriate model and estimate the predicted outcome per study, optionally controlling for individual‐level confounders. Note that instead of using fractional polynomials as Royston and Sauerbrei, we may use any of the spline approaches described in section 4. At the second stage, for each distinct value of X a meta-analysis is performed on the resulting predicted values and standard errors, using either a fixed or random effects approach [14]. Given a continuous variable X the algorithm proceeds as follows:

**Stage 1**

1. Select a spline approach and fit an appropriate model including interaction between X and the treatment.. 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. Since this is an introduction to splines, we fit only splines.
2. Using the model with interaction, estimate regression lines and for the control and treated group in study *j* respectively, along with their standard errors and 95% confidence intervals.
3. Depending on the outcome we wish to show and depending on the scale on which we wish to make inferences, we may choose to use a link function :
   * 1. If, in stage 2, we aim to show the predicted outcome per treatment arm and conditional on X, we calculate the predicted outcome per treatment arm, and
     2. If, in stage 2, we aim to show the effect of the treatment conditional on X, we first calculate per study the absolute treatment effect or the relative treatment effect and calculate the corresponding confidence interval, see section 3. Note that if the goal of our meta-analysis is to make inferences on the treatment effect, this approach is preferable to step 3.1, to avoid amalgamating the within and between study heterogeneity [45, 46].

**Stage 2**

1. For each we perform either a fixed or random effects meta-analysis to get the pooled outcome of choice as a function of X along with its pointwise 95% confidence interval. Note that if the available data across the studies vary over different regions of X, pooling of the predicted outcomes may produce unsmooth results, see Figures 8 and 9.

We applied all aforementioned spline approaches in all 3 IPD-MA scenarios. As described in step 3.1 we pooled the results to show mortality risks per treatment arm (Figure 8), and following step 3.2 we pooled the treatment effects conditional on BMI (Figure 9). 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, for B-splines 4 equidistant knots (2 inner knots plus the boundaries per study), and for P-splines 17 equidistant knots (15 inner knots plus the boundaries per study). For the penalised splines (P-splines and Smoothing splines) the tuning parameter λ was selected through a ‘leave one out’ GCV process. In all scenarios, knots are placed over the full domain of BMI.

Subsequently, we estimated per study and treatment arm the mortality risk conditional on BMI, and we estimated per study the risk difference as described in step 3.2 and their confidence intervals as described in section 3. In the second stage we pooled both the regression lines per treatment arm and their risk difference, using random effects meta-analysis with a REML estimator for τ2.

Figure 8. approximately here

Figure 9. approximately here

## Two-stage multivariate meta-analysis

Instead of using pointwise meta-analysis per distinct value of X, the functional shapes can also be pooled using multivariate meta-analysis. This approach, as proposed by Gasparrini et al. [23], pools the set of regression coefficients estimated in the first stage, accounting for their within- and (if applicable) between-study correlation, using a fixed or random effects multivariate meta-analysis approach. Hereby, we may use the regression splines approaches described in section 4.1. Note that in order to pool the results of the first stage, each study should provide the same set of coefficients, estimated in the same domain of X. In case of different domains 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.

The multivariate meta-analysis algorithm proceeds as follows:

**Stage 1**

1. As a preliminary step choose the knots corresponding to the optimal locations across the studies along with the degree of the spline method we wish to apply.
2. Per study j fit a model including interaction between X and the treatment with the chosen specifications of step 1.
3. With Q the total number of coefficients and q [1, 2, …, Q], extract per study the estimated coefficients along with their variance-covariance matrix.

**Stage 2**

1. Use either fixed or random effects multivariate meta-analysis to estimate the pooled
2. To calculate the predicted outcome given X and treatment T multiply the pooled estimates with the design (or model) matrix containing the values of X along with their spline transformed values.
3. To estimate the treatment effect conditional on X, subtract the pooled-per-treatment arm outcomes and calculate the confidence interval as described in section 3.

In contrast to pointwise meta-analysis, multivariate meta-analysis pools the coefficients of the basis functions estimated in the first stage. As described in section 4.2 the coefficients of the penalised splines are biased due to penalisation. Therefore, pooling them may be problematic and show biased results. Note that in multivariate meta-analysis careful specification of the knots is required as convergence issues may occur during the second stage.

We applied multivariate meta-analysis in combination with regression splines in all 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 values over the full range of BMI. In stage 1, per study we fitted restricted cubic spline and B-spline transformations of BMI both as main effects and as interactions with the treatment. For the restricted cubic spline transformations, we used 5 knots, following Harrell’s suggestion to use the 5%, 27.5%, 50%, 72.5% and 95% quantiles of BMI, for B-splines 4 equidistant knots (2 inner knots plus the boundaries per study). Note that we position the knots over the full domain of BMI. Subsequently, we pooled the estimated coefficients using a random-effects meta-analysis with the REML estimation method.  We calculated regression lines per treatment arm by multiplying the design (or model) matrix with the pooled coefficients. Absolute risk differences were calculated by subtracting the pooled mortality risks of the treated minus the control, while for the confidence intervals we used the proposal of Newcombe [34].

Figure 10. approximately here

Figure 11. approximately here

## One-stage generalised additive mixed effects model

Instead of using a two-stage meta-analysis, we may also conduct the analysis in one stage, using a mixed effect model with splines, i.e. a generalised additive mixed effect model (GAMM). In this context, researchers may choose for the effects of each variable to be either fixed (common), random or stratified. As in GLMMs, the common effect assumption is that the effect of X is identical across all studies. The random effects assumption is that the effect of X comes from a distribution of effects, while the stratified effects assumption is that for each study the effect of X may be different and is estimated per study. The common effect may be modelled straightforward by including X, and stratified effects can be modelled by including an interaction of X with the (categorical) clustering variable (e.g. studies).

The general statistical model for a one-stage approach is:

where is the splines transformation of X for the control group and the spline transformation for the interaction of X × T.

For each coefficient described above the fixed (or common) effect model assumes that βq is common across studies (βq = βqj). Under the random effects assumption the beta coefficients βqj differ across studies and come from a normal distribution with a pooled βq as mean and standard deviation τq2 (N(βqj ~ N(βq, τq2)). Under the stratified effects assumption, the βqj are different across studies, but do not follow a specific distribution. 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, 49–51]. To avoid this, two methods have been proposed. One approach is to stratify per study all or some of the main effects including at least the treatment effect [25]. Another approach is to center the covariate Xij about its study‐specific mean creating a new variable Z =Xij -Xj. Subsequently, include Z, Xj, and the interaction of Z with the treatment in the one-stage model [25, 52]

Wood [22, 53, 54] and Kimeldorf and Wahba [55] have shown that random effects may be modelled by penalising the interactions of X with the studies. Since this penalisation term can be written as a random effects parameter in a Bayesian/mixed effect model, we may additively combine both the functional form modelled by splines and their corresponding random effects in a mixed effect model. Note that we may combine any type of splines, either penalised or regression splines, with the random effects spline. This way a common functional shape of the association between the outcome and the covariable X can be investigated, while accounting for the clustering of the participants within the studies.

In our examples, we used the 4 aforementioned spline transformations both as main effect and in interaction with the treatment. We used a random intercept and random slope for BMI in combination with a fixed spline part. For the restricted cubic splines, we used 5 knots (the 5%, 27.5%, 50%, 72.5% and 95% quantiles of BMI), for B-splines 4 equidistant knots (2 inner knots plus the boundaries), and for P-splines we used 17 equidistant inner knots (15 inner knots plus the boundaries). Note that we position the knots over the full domain of BMI and that no data-augmentation is needed in one-stage generalised additive mixed effects models.

Figure 12. approximately here

Figure 13. approximately here

## Properties of the pooling methods

We illustrated the association between BMI and mortality risk in three scenarios using 4 spline methods and 2 two-stage approaches (pointwise and multivariate meta-analysis) and 1 one-stage approach (GAMM). Table 2 summarises the properties of the aforementioned approaches. An advantage that both two-stage approaches share is that during their first stage they provide a better insight in the underlying associations per study. Furthermore, in two-stage methods we may use heterogeneity measures such as Cochran’s Q statistic, τ2, and prediction intervals per value of X to assess whether it is sensible to pool the per-study associations. Therefore, it is always informative to investigate the results per study similar to the first stage in two-stage methods.  **Pointwise meta-analysis: robust and flexible but non-smoothness may occur.**

The main advantages of pointwise meta-analysis are its flexibility, robustness and ease of use. In pointwise meta-analysis we are allowed to fit different models across the studies, as we are pooling the predicted outcomes rather than the coefficients. For instance, in one study we may apply a restricted cubic spline transformation of X, in another a 2nd degree B-spline and in another no spline transformation of X. Also, we are allowed to vary the number and position of knots per study. Pointwise meta-analysis is also robust to model mis-specification. When the ranges of X are different across studies, pointwise meta-analysis may use the whole domain of X even without data augmentation. The main disadvantage of pointwise meta-analysis is that when the ranges of X are not the same across studies the pooled curve may be unsmooth. Also, since we are performing a meta-analysis for each value of X pointwise meta-analysis may be more computationally intensive than multivariate meta-analysis and GAMMs.

**Multivariate meta-analysis: efficient in case of similar domains, but lacks robustness and flexibility**

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 [48]. The main limitation is that multivariate meta-analysis lacks robustness to model misspecification. Multivariate meta-analysis lacks flexibility compared to pointwise meta-analysis, since the models fitted per study should have the same parametrisation e.g. the type of spline, the number and positions of knots and the same range of X [23]. This restriction may be problematic in cases where a subset of studies included in the meta-analysis has limited number of participants for instance meta-analyses that include RCTs. In that case modelling the association between the outcome and spline transformations of X may fail to converge and only multivariate meta-analysis with linearities assumptions may be possible. Finally, since multivariate meta-analysis pools beta coefficients estimated during the first-stage it may not be compatible with approaches where penalisation to that terms is applied.

**GAMM: allows for different study domains and sample sizes, but careful modelling required**

The main advantage of GAMMs is that they can handle differences in the distributions of X across studies, include all studies regardless of the number of observations each one has, and result in smooth pooled curves and confidence intervals. The main disadvantage of GAMMs is that we may lose the insight in the underlying associations per study offered during the first stage of two-stage methods. Furthermore, GAMMs require careful modelling, especially when aggregation (ecological) bias might be present as discussed by Riley et al. and Belias et al. [25, 46, 52].

# Software

All analyses were performed in the statistical software R version 3.6.0. For data manipulation we used the **tidyverse** [56] package, for the splines and GAMMs we used the **mgcv** [53] package, including its predict function for the confidence intervals, and for pointwise and multivariate meta-analysing the estimates the **meta** [57] and **mvmeta** [23] packages respectively. It is also possible to estimate splines in other software such as Stata or SAS. However, since R is freely available for every researcher, we provide the scripts to apply splines in both single and multiple studies scenarios only in R.

# Empirical example

To illustrate the use of splines combined with the aforementioned pooling methods in a real example we consider a previously published IPD-MA investigating the effect of antibiotics in children with acute otitis media [26]. Rovers et al. collected IPD from six randomised clinical trials with a total of 1643 children, aged from 0 to 12 years old. The primary outcome was fever and/or ear-pain after 3-7 days (yes/no) after antibiotics or placebo treatment. Hereby, we investigate the effect of antibiotics across the values of age, in children with unilateral or bilateral acute otitis media (AOM).

## Methods

From a total of 6 studies, we used 5 studies and omitted one study with 315 participants from all subsequent analyses as the information for unilateral or bilateral AOM was not reported. From the remaining 5 studies, one study (Appelman et al.) had limited number of events (children with fever and/or ear pain) and at some age- bilateral AOM combinations no events at all. Therefore, we followed different strategies across the pooling methods for this study.

For pointwise meta-analysis, for the Appelman study mentioned above we fitted a logistic regression including the main effects of bilateral AOM, treatment and age and their interactions without any spline transformation for age. For the remaining studies we fitted per study a logistic regression including the main effects of bilateral AOM, treatment and age transformed with the aforementioned spline approaches, and included the interactions of the spline transformed age with bilateral AOM (yes/no) and treatment. For restricted cubic splines, we followed Harrell’s suggestion and used per study 3 knots at 10%, 50%, 90% quantiles of age, for B-splines we used 2nd degree basis functions and per study 3 equidistant knots (1 inner knot plus the per study boundaries), while for P-splines we used 3rd degree basis functions and per study 17 equidistant knots (15 inner knots plus the per study boundaries). For the penalised splines (P-splines and Smoothing splines) the tuning parameter λ was selected through a ‘leave one out’ GCV process. Subsequently, to show the pooled risk conditional to children’s age and bilateral AOM and intervention group, we extracted the predicted outcome for fever and/or ear pain after 3-7 days in logit scale. Subsequently, we back-transformed them into risks and pooled them using a random-effects meta-analysis approach with REML τ2 estimator. In addition, to show the treatment effect conditional to children’s age and bilateral AOM, we first back-transformed the predicted mortality risk per study. Subsequently, per study we estimated the risk difference conditional to age for children with and without bilateral AOM along with their confidence intervals and then pooled them using a random-effects meta-analysis approach with REML τ2 estimator.

For multivariate meta-analysis, we omitted the Appelman study mentioned above. Therefore, multivariate meta-analysis was based on the four remaining studies with sufficient number of observations to fit splines. 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. In the first stage of the multivariate meta-analysis we fitted a logistic regression model including the main effects of treatment, bilateral AOM and spline transformed age and the interactions of spline transformed age with treatment and bilateral AOM. Since in multivariate meta-analysis the positions of knots need to be the same across the studies, for restricted cubic splines we used 3 knots at 10%, 50%, 90% quantiles of age calculated on the 2 studies combined, for B-splines we used 2nd degree basis functions and 3 equidistant knots (1 inner knot at 2.5 plus the boundaries of age [0.5 – 9]). Subsequently, we extracted the beta coefficients and their variance-covariance matrix and pooled them using a random-effects meta-analysis approach with REML estimator. Finally, to show the risk of developing fever after 3-7 conditional to age, treatment and bilateral AOM, we multiplied the pooled coefficients with the corresponding design matrix and back-transformed the pooled outcomes using the inverse logit function. To calculate the absolute risk differences and their confidence intervals, we followed the proposal of Newcombe [34], see section 3.

For GAMMs, we included all 5 studies. We fitted a logistic regression model including the main effects of treatment, bilateral AOM and spline transformed age and the interactions of spline transformed age with treatment and bilateral AOM. We used similar knot positioning and degrees of splines as in pointwise and multivariate meta-analysis but using the whole data-set. We followed Wood’s proposal and included random-effects intercept and age slope additively to account for the within study clustering of participants [42].

## Results

Figures 14,16 and 18 show the pooled regression curves of pointwise and multivariate meta-analysis, and GAMMs conditional on age and bilaterality of AOM. Figures 15,17,19 show the absolute risk difference between treated and control (treatment effect) conditional age and bilaterality of AOM. 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. Therefore, we compare the pooled curves with regard to their plausibility, specifically in terms of smoothness and report convergence issues if any. As we were investigating a three-way interaction (treatment, bilaterality of AOM and age) in some studies the combinations of theses variables created groups of patients where we had limited number of events. Consequently, in the second stage pointwise meta-analysis showed wide confidence intervals in sub-domains of age. Furthermore, as in the artificial data-sets, the predicted pooled regression lines were unsmooth due to differences in the overlap of age across the studies. Multivariate meta-analysis also showed wide confidence intervals for the restricted cubic splines approach and did not show any results for B-splines due to failure to converge during the second stage. GAMMs combined with penalised splines showed smooth pooled regression lines and confidence intervals, see Figure 15 and Figure 16, while using regression splines showed wide confidence intervals in sub-domains of age and non-smooth regression lines.

Figure 14. approximately here

Figure 15. approximately here

Figure 16. approximately here

Figure 17. approximately here

Figure 18. approximately here

Figure 19. approximately here

# Discussion

Our results, in which we illustrated 4 spline-based approaches (restricted splines, B-splines, P-splines and Smoothing splines), and three pooling methods (pointwise meta-analysis, multivariate meta-analysis and GAMMs) on three scenarios with artificial data, showed that all approaches performed equally well in modelling the underlying true association analysis in case of an heterogeneous data-set with similar ranges of the effect modifier. In the two scenarios with different ranges, only GAMMs showed smooth pooled regression lines and confidence intervals. When applying the aforementioned splines and pooling methods on an empirical example investigating the association between age and the effect of antibiotics in children from 0-9 years with unilateral and bilateral otitis media, we found that GAMMs, especially when combined with penalised splines, showed smooth pooled regression lines and reasonable confidence intervals for the whole range of the potential treatment effect measure modifier (age), while pointwise meta-analysis showed discontinuous and very wide confidence intervals. Multivariate meta-analysis was limited to a subset of studies, failed to converge for B-splines and P-splines, and showed results in only a limited range of ages for restricted cubic splines.

The major strength of our manuscript is that as far as we are aware, we are the first to provide an introduction on how to apply a variety of splines methods in both single and multiple studies, in order to investigate treatment effect differences when non-linearities are present.

Some potential limitations should also be mentioned. First, we did not illustrate the performance of the aforementioned approaches in a scenario with across studies homogeneous associations and similar ranges of the effect modifier. We considered that this scenario is rarely accounted in practice and that all approaches would produce similar results. Second, we did not illustrate the performance of the aforementioned approaches in scenarios with ecological bias. We considered that ecological bias in IPD-MA when non-linearities are present would be too complicated for an introduction and therefore is out of the scope of our paper. Third, the data generating mechanism of the illustrative examples was simple. The association of mortality risk with BMI was quadratic and quartic for the control and treated group, respectively, and the between-study heterogeneity was limited (I2 less than 40%). These settings generated realistic data that are appropriate for pooling. . Fourth, and corresponding to our main aim to provide an introduction to splines, we limited our study spline-based approaches whereas other techniques might also be able to deal with non-linear associations, e.g. tree-based approaches [7–11], meta-stepp [12, 13], locally (weighted) estimated scatter-plot smoothing (lo(w)ess), or fractional polynomials [14, 15, 36].

Other researchers have also drawn the attention to the importance of modelling non-linear associations in IPD-MA [25, 48]. These studies focused on estimating relative treatment effect functions whereas we focused on estimating the absolute risk differences. Our examples and results show that accounting for nonlinearities is also of great importance to investigate treatment effect differences on the absolute scale. Therefore, we believe that this illustrative introduction on how to apply splines in single and multiple studies scenarios will aid researchers to account for non-linearities. Doing so may provide better insight in the underlying associations and contribute to more evidence-based conclusions and thus better clinical decision making.

In conclusion, taking into account non-linear associations whilst combining multiple studies needs careful modelling. Depending on the spline and pooling method different results may be found. We showed that across 3 common IPD-MA scenarios and one empirical example, one-stage GAMM combined with penalised splines seems to account for differences across studies in the ranges of the effect modifier and the functional forms, while showing continuous and smooth regression lines. Splines provide a helpful tool to capture nonlinear treatment effect differences in IPD-MA.

# References

1. Simmonds M, Stewart G, Stewart L. A decade of individual participant data meta-analyses: A review of current practice. Contemp Clin Trials. 2015;45 Pt A:76–83.

2. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ. 2006;332:1080.1.

3. Austin PC, Brunner LJ. Inflation of the type I error rate when a continuous confounding variable is categorized in logistic regression analyses. Statist Med. 2004;23:1159–78.

4. Maxwell SE, Delaney HD. Bivariate median splits and spurious statistical significance. Psychological Bulletin. 1993;113:181–90.

5. Weinberg CR. How bad is categorization? Epidemiology. 1995;6:345–7.

6. Greenland S. Basic Problems in Interaction Assessment. Environmental Health Perspectives. 1993;101:59.

7. Zeileis A, Hothorn T, Hornik K. Model-Based Recursive Partitioning. Journal of Computational and Graphical Statistics. 2008;17:492–514.

8. Seibold H, Hothorn T, Zeileis A. Generalised linear model trees with global additive effects. Adv Data Anal Classif. 2019;13:703–25.

9. Su X, Tsai C-L, Wang H, Nickerson DM, Li B. Subgroup Analysis via Recursive Partitioning. SSRN Journal. 2009. doi:10.2139/ssrn.1341380.

10. Mistry D, Stallard N, Underwood M. A recursive partitioning approach for subgroup identification in individual patient data meta-analysis. Statistics in Medicine. 2018;37:1550–61.

11. Fokkema M, Smits N, Zeileis A, Hothorn T, Kelderman H. Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. Behav Res. 2018;50:2016–34.

12. Wang XV, Cole B, Bonetti M, Gelber RD. Meta-STEPP with random effects. Res Syn Meth. 2018;9:312–7.

13. Wang XV, Cole B, Bonetti M, Gelber RD. Meta-STEPP: subpopulation treatment effect pattern plot for individual patient data meta-analysis. Statist Med. 2016;35:3704–16.

14. Sauerbrei W, Royston P. A new strategy for meta-analysis of continuous covariates in observational studies. Statist Med. 2011;30:3341–60.

15. Royston P, Sauerbrei W. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. Statist Med. 2004;23:2509–25.

16. Kasenda B, Sauerbrei W, Royston P, Mercat A, Slutsky AS, Cook D, et al. Multivariable fractional polynomial interaction to investigate continuous effect modifiers in a meta-analysis on higher versus lower PEEP for patients with ARDS. BMJ Open. 2016;6:e011148.

17. Harrell FE. General Aspects of Fitting Regression Models. In: Regression Modeling Strategies. Cham: Springer International Publishing; 2015. p. 24–6. doi:10.1007/978-3-319-19425-7\_2.

18. de Boor C. A Practical Guide to Splines. New York, NY: Springer New York; 1978. doi:10.1007/978-1-4612-6333-3.

19. de Boor C. Package for Calculating with B-Splines. SIAM J Numer Anal. 1977;14:441–72.

20. Best DJ, Green PJ, Silverman BW. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. Biometrics. 1994;50:1228.

21. Eilers PHC, Marx BD. Flexible smoothing with B -splines and penalties. Statist Sci. 1996;11:89–121.

22. Wood SN. Introducing GAMs. In: Generalized Additive Models. 2nd edition. Chapman and Hall/CRC; 2017. p. 161–94. doi:10.1201/9781315370279-4.

23. Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. Statist Med. 2012;31:3821–39.

24. White IR, Kaptoge S, Royston P, Sauerbrei W, The Emerging Risk Factors Collaboration. Meta-analysis of non-linear exposure-outcome relationships using individual participant data: A comparison of two methods. Statistics in Medicine. 2019;38:326–38.

25. Riley RD, Debray TPA, Fisher D, Hattle M, Marlin N, Hoogland J, et al. Individual participant data meta‐analysis to examine interactions between treatment effect and participant‐level covariates: Statistical recommendations for conduct and planning. Statistics in Medicine. 2020;39:2115–37.

26. Rovers MM, Glasziou P, Appelman CL, Burke P, McCormick DP, Damoiseaux RA, et al. Antibiotics for acute otitis media: a meta-analysis with individual patient data. The Lancet. 2006;368:1429–35.

27. Sun Y-Q, Burgess S, Staley JR, Wood AM, Bell S, Kaptoge SK, et al. Body mass index and all cause mortality in HUNT and UK Biobank studies: linear and non-linear mendelian randomisation analyses. BMJ. 2019;:l1042.

28. Bhaskaran K, dos-Santos-Silva I, Leon DA, Douglas IJ, Smeeth L. Association of BMI with overall and cause-specific mortality: a population-based cohort study of 3·6 million adults in the UK. The Lancet Diabetes & Endocrinology. 2018;6:944–53.

29. Rothman KJ. Epidemiology: an introduction. 2nd ed. New York, NY: Oxford University Press; 2012.

30. Brumback B, Berg A. On effect‐measure modification: Relationships among changes in the relative risk, odds ratio, and risk difference. Statist Med. 2008;27:3453–65.

31. Miettinen O. CONFOUNDING AND EFFECT-MODIFICATION. American Journal of Epidemiology. 1974;100:350–3.

32. VanderWeele TJ. Confounding and Effect Modification: Distribution and Measure. Epidemiologic Methods. 2012;1. doi:10.1515/2161-962X.1004.

33. Greenland S. Effect Modification and Interaction. In: Balakrishnan N, Colton T, Everitt B, Piegorsch W, Ruggeri F, Teugels JL, editors. Wiley StatsRef: Statistics Reference Online. Chichester, UK: John Wiley & Sons, Ltd; 2015. p. 1–5. doi:10.1002/9781118445112.stat03728.pub2.

34. Newcombe RG. MOVER-R confidence intervals for ratios and products of two independently estimated quantities. Stat Methods Med Res. 2016;25:1774–8.

35. Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M. A review of spline function procedures in R. BMC Med Res Methodol. 2019;19:46.

36. Harrell , FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Cham: Springer International Publishing; 2015. doi:10.1007/978-3-319-19425-7.

37. Dierckx P. Curve and surface fitting with splines. Choice Reviews Online. 1993;31:31-2162-31–2162.

38. Stone, C. J. Comment: Generalized additive models.

39. Akaike H. Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika. 1973;60:255–65.

40. Craven P, Wahba G. Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. Numer Math. 1978;31:377–403.

41. Ruppert D, Wand MP, Carroll RJ. Semiparametric Regression. 1st edition. Cambridge University Press; 2003. doi:10.1017/CBO9780511755453.

42. Wood SN. Generalized additive models: an introduction with R. Second edition. Boca Raton: CRC Press/Taylor & Francis Group; 2017.

43. Welch BL. The Generalization of `Student’s’ Problem when Several Different Population Variances are Involved. Biometrika. 1947;34:28.

44. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ. 2010;340 feb05 1:c221–c221.

45. Fisher DJ, Carpenter JR, Morris TP, Freeman SC, Tierney JF. Meta-analytical methods to identify who benefits most from treatments: daft, deluded, or deft approach? BMJ. 2017;:j573.

46. Belias M, Rovers MM, Reitsma JB, Debray TPA, IntHout J. Statistical approaches to identify subgroups in meta-analysis of individual participant data: a simulation study. BMC Med Res Methodol. 2019;19:183.

47. White IR. Multivariate Random-effects Meta-analysis. The Stata Journal. 2009;9:40–56.

48. White IR, Kaptoge S, Royston P, Sauerbrei W, Emerging Risk Factors Collaboration. Meta-analysis of non-linear exposure-outcome relationships using individual participant data: A comparison of two methods. Stat Med. 2019;38:326–38.

49. Riley RD, Steyerberg EW. Meta-analysis of a binary outcome using individual participant data and aggregate data. Res Synth Method. 2010;1:2–19.

50. Higgins JPT, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. Statist Med. 2001;20:2219–41.

51. Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijs L, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. Statist Med. 2008;27:1870–93.

52. Hua H, Burke DL, Crowther MJ, Ensor J, Tudur Smith C, Riley RD. One-stage individual participant data meta-analysis models: estimation of treatment-covariate interactions must avoid ecological bias by separating out within-trial and across-trial information: One-Stage IPD Meta-Analysis Models Must Avoid Ecological Bias. Statist Med. 2017;36:772–89.

53. Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models: Estimation of Semiparametric Generalized Linear Models. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2011;73:3–36.

54. Wood SN. Fast stable direct fitting and smoothness selection for generalized additive models. J Royal Statistical Soc B. 2008;70:495–518.

55. Kimeldorf GS, Wahba G. A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines. Ann Math Statist. 1970;41:495–502.

56. Wickham H. Tidyverse: Easily install and load the ’tidyverse’. 2017. https://CRAN.R-project.org/package=tidyverse.

57. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Mental Health. 2019;22:153–60.

58. Hastie T, Friedman J, Tibshirani R. Basis Expansions and Regularization. In: The Elements of Statistical Learning. New York, NY: Springer New York; 2001. p. 115–63. doi:10.1007/978-0-387-21606-5\_5.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Number of knots | 1st | 2nd | 3rd | 4th | 5th | 6th | 7th |
| 3 | 10% | 50% | 90% |  |  |  |  |
| 4 | 5% | 35% | 65% | 95% |  |  |  |
| 5 | 5% | 27.5% | 50% | 72.5% | 95% |  |  |
| 6 | 5% | 23% | 41% | 59% | 77% | 95% |  |
| 7 | 2.5% | 18.33% | 34.17% | 50% | 65.83% | 81.67% | 97.5% |

Table 1: Number of knots and their quantile location based on Harrell’s proposal for the restricted cubic splines approach.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Characteristic | Pointwise meta-analysis | | Multivariate meta-analysis | | GAMM | |
| Base for pooling | Uses the predicted outcome per value of X | | Uses coefficients of fitted curves | | Uses a common curve with random effects to account for across studies differences | |
| Allows study specific fitting strategies (different model specifications) with regard to the effect modifier | Yes | | No | | No | |
| Difficulty to perform | Easy | | Difficult | | Intermediate | |
| Performance | Same distribution of X across studies: wider confidence intervals than multivariate-meta and GAMM | Different distributions of X across studies:  more flexible than multivariate meta-analysis | Same distribution of X across studies:  narrower  confidence intervals than pointwise meta-analysis  and GAMM | Different distributions of X across studies:  lacks robustness. | Same distribution of X across studies: wider confidence intervals than multivariate meta-analysis | Different distributions of X across studies: wider confidence intervals than multivariate meta-analysis |
| Main advantage | Flexible | | Efficient | | Smooth | |
| Main disadvantage | May show discontinuities in the pooled curves | | Lack of robustness, very sensitive to modelling choices  Needs all parameters to be the same across studies.  Data augmentation might be needed as a preliminary step. | | Could lead researcher to lose insight into the underlying associations per study, if done without the first stage of the two-stage methods | |
| Can be performed using all types of splines | Yes.  Plus: different models may be applied across studies | | No  Not suitable in combination with penalised splines | | Yes. | |

**Table 2. Comparison of the pooling methods**

# Highlights

## What is already known

* Analysing associations between outcomes and continuous patient characteristics in IPD-MA may be challenging when non-linear associations are present.
* Splines offer great flexibility but are rarely used.

## What is new?

* We provide an introduction and guidance on how to model non-linear treatment effects using restricted splines, B-splines, P-splines and Smoothing splines and three IPD-MA methods to pool the results of multiple studies: pointwise meta-analysis, multivariate meta-analysis, and generalised additive mixed effects models (GAMMs).
* We illustrate the performance of the splines and pooling methods on three common IPD-MA scenarios and provide R code.
* We illustrate their performance of the splines and pooling methods on real acute otitis media data

## Potential impact for other fields

* Splines provide a helpful tool to capture nonlinear treatment effects in IPD-MA.
* Modelling nonlinear associations may provide personalised treatment effects with more accuracy, thus aiding to more precise clinical decision making.

# Appendix

For readability we adopt the following notation throughout the manuscript:

* The trials as j = 1,2, …, n
* Trial participants as i = 1,2, …,,
* Continuous effect modifier: X
* Binary treatment indicator: T with value 0 for the control group and 1 for the experimental group.
* The true association of X with the outcome: f(X)
* Smoothed estimated function:
* the boundaries of X and the boundaries of X per trial j
* The number of (inner) knots:
* w є [1,…, κ +1] the intervals defined by the knots
* g: a link function and its inverse function
* B(X;d) a basis function of dth degree

## Polynomial regression and truncated power series

In general, to model the association between an independent variable X and an outcome Y, generalised linear models (GLMs) are used. In case of non-linear associations, transformed versions of X can be used instead of X. For instance, the statistical model for a GLM with link function g and with a d-degree polynomial of X is:

However, a global function over the full range of X may have poor fit near the boundaries due to instability of the estimated polynomial in this area. To avoid these issues, piecewise polynomials may be preferred to global functions. The model for a d-degree polynomial for interval *w*, between knot *tw* and *tw+1*, would be:

These piece-wise polynomials, when fitted in two consecutive intervals, will show different predicted values at the boundaries of the intervals (*i.e.* at the knots), thus their functional shape will be discontinuous. For this reason, we may use restrictions to “connect” interval-specific polynomials. One convenient solution is to fit a global polynomial, and model the deviations from this globally defined shape within truncated parts of X. Thereto, each basis function is a polynomial with one term. Given a non-decreasing sequence of knots a truncated power series basis is defined by the following basis functions:

and the statistical model for the association between X and Y is:

The + subscript denotes that for a given z

The first term in equation (4) generates the global polynomial, often called the “basic” polynomial, whereas the second term, often called the “secondary” polynomial, is modelling the deviations from it. The resulting splines, using truncated power series basis functions, are often called polynomial ‘regression’ splines [58]. The term truncated reflects the fact that the intervals for the power series in the secondary polynomial are shortened to produce estimates only in sub-domains of X. A disadvantage of truncated power series is that they can still show erratic behaviour near the ranges of X.

# Regression splines

## Natural or restricted splines

A solution to this erratic behavior near the boundaries is to restrict the truncated power series to be linear near the boundaries of X [17]. These splines are often called natural or restricted (polynomial) splines. Given a non-decreasing sequence of knots the statistical model is given as:

where

and for *w* є [2, κ-1]

Harrell shows that restricted cubic splines can also be written as truncated power series with a linear “basic” polynomial, by dividing the basis functions by [36]. Therefore, an equivalent statistical model to (5) may be written as follows:

The number and location of the knots may be based on clinical knowledge or on descriptive statistics. For instance, Harrell suggests the use of quantiles and advocates that four knots in most cases are adequate [17, 38], see Table 1. Depending on the available sample size and required complexity of the functional shape we may use a different number of knots.

In Figure 5(a) we show the basis functions - scaled by for the restricted cubic splines approach with 5 knots placed at the 5%, 27.5%,50%, 72.5%, 95% quantiles. In our single study example, we used restricted cubic spline transformations of X both as main effects and as interactions with the treatment. Following Harrell’s suggestion, we placed 5 knots at values corresponding to 5%, 27.5%, 50%, 72.5% and 95% quantiles of X.

.

## B-splines

B-splines are another commonly applied spline approach. They are based on a parametrisation of polynomial cubic splines. Given a non-decreasing knot sequence and X є , the dth degree B-splines basis functions are calculated by the following algorithm proposed by De Boor [19].

First, d additional knots are generated before and d additional knots after . These are often called outer knots and their choice is arbitrary. We can set them to be equidistant or even equal to the boundary values and of X. A new knot sequence is generated, where:

are the left outer knots or endpoints,

,

, , …, the inner knots

,

the right outer knots

Within each interval *w* a zero-degree B-spline is calculated. Zero-degree B-splines are step functions equal to 1 within an interval and 0 otherwise.

All succeeding basis functions, with degree >1, are calculated using the following formula:

where [1, 2, …, d]. For example, the first degree is calculated using the zero degree B-splines, and the second degree from and so on, using formula 7.

Three variations of B-splines based on the inner knot positionings have been proposed. B-splines with a uniform knot vector use equidistant knots and are the most typically applied B-splines [35]. B-splines with an open uniform knot vector also use equidistant knots but they allow analysis of closed curves. Non-uniform B-splines use non-equidistant knots, placed at positions of the researcher’s choice. To our topic, uniform and non-uniform B-splines are the most relevant. Non-uniform B-splines may reflect the a-priori knowledge of a researcher over the underlying complexity of the functional form and/or distribution of the continuous variable.

In Figure 5(b) we show the basis functions of a 2nd degree B-spline with 4 equidistant knots; 2 inner knots plus the boundaries [α, β], placed at values α=0, 0.33, 0.66 and 1=β.

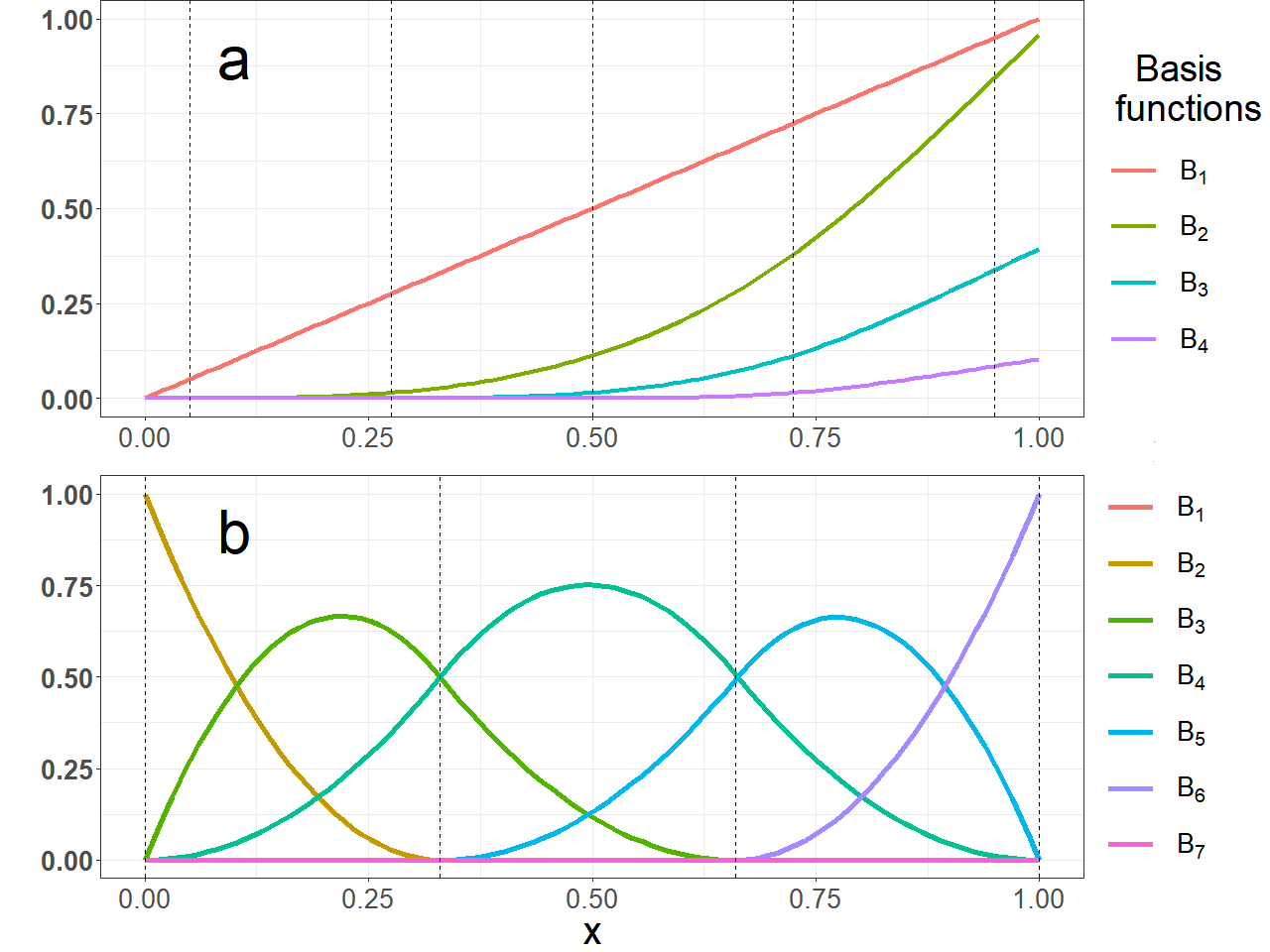


Figure 5. approximately here

# Penalised splines

The two commonly applied penalised splines that we discuss, P-splines and Smoothing splines, increase the number of knots to a large set (usually, 10-40) or even to be equal to the number of observations. This way they circumvent the problem of choosing the number and positions of the knots. Since estimating one parameter for each observation would clearly lead to a perfect fit and thus generate functional shapes with extreme variability, penalised splines introduce in their optimisation functions a penalty term () multiplied by a non-negative , often called a tuning parameter. As the term “tuning” implies, changing the value of changes the magnitude of the penalisation.

In GLMs the estimation of the regression coefficients is accomplished through optimisation of functions of . For Gaussian outcomes the least squares optimisation is estimating the that minimise the squared distance of the predicted and the observed values of the outcome, while for outcomes belonging to the exponential family (Gaussian, Binary, Poisson etc) we estimate the maximising the likelihood function of . Adding a penalty term () results in the following optimisation equations:

* Least squares approach
* Maximum Likelihood approach

Penalised splines circumvent the problem of knot selection, but at a cost. By using a penalty in their optimisation function, they introduce bias in their estimate in order to obtain a more stable solution. Further, in both P-splines and Smoothing splines the tuning parameter must be specified. Too high or too low values of may lead to over- or undersmoothing respectively. Several approaches have been proposed in order to determine the “optimal” , such as Akaike’s information criterion AIC [39], “leave one out” generalised cross-validation (GCV) [40] or mixed-effects modelling [22]. These processes are automated in most of the statistical packages. Briefly, when using the AIC, a series of models fitted with different values is compared and the one with the lowest AIC is selected. “Leave one out” GCV is an iterative process, the algorithm goes as follows: 1) one observation is omitted 2) a model is fitted 3) using the model a prediction of the omitted value is generated and 4) the distance between the observed and predicted value is calculated. This procedure is repeated for each observation and for a series of values. The that minimizes the GCV minimizes the sum of the squared distances, i.e. the GCV score, is selected. In Bayesian/mixed effects modelling approach the penalty term is estimated in a similar way as random effects parameters.

## P-splines

A specific type of penalised splines, P-splines, proposed by Eilers and Marx [21], is a penalised version of B-splines, using a specific penalty term based on the sum of p-order differences between the coefficients of two consecutive intervals . The first order differences are defined as follows: , but Eilers and Marx propose the use of second order differences, which are the first order differences of the first order differences

.

Note that the degree of the underlying B-splines may be different from the order of the differences. A common combination is that of a third-degree B-spline with a second order difference. Using a penalty based on a zero-degree order difference results in the ridge penalty [41]. Note that in some occasions penalised splines and penalised B-splines are misinterpreted as P-splines, but not all penalised B-splines or penalised splines are P-splines. For instance, ordinary B-splines may be fitted using a Smoothing splines approach, but this does not make them P-splines, unless they are penalised using the approach suggested by Eilers and Marx.

P-splines are based on equidistant knots. It is possible to use a knot sequence that is not evenly spaced; but in this case, weights need to be introduced [22, 35]. As P-splines with non-equidistant knots are rarely used in practice we don’t consider them in this article.

## Smoothing splines

Smoothing splinesare another member of the family of penalised spline methods. Similar to P-splines the idea is to increase the number of knots, but this time to be equal or approximately equal to the number of observations. O’ Sullivan [41] suggested that a penalty based on Reinsch’s integral of the second derivative of , where is a cubic spline, multiplied by a tuning parameter, has good smoothing properties. This results in the following penalty term for Smoothing splines: .

## Single study artificial data-set simulation functions

The risk of mortality per participant in the single study data-set was generated using the following formulas:

* For the control group

* For the treated group

Equivalently equations (1) and (2) can also be combined into a single equation:

where would be the association of BMI with mortality risk for the control and the additive effect of the treatment.

## Multiple studies artificial data-set simulation functions

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | True underlying function forms | | BMI Ranges | | | | |
| Scenarios | Control | Treated | 1st Study | 2nd Study | 3rd Study | 4th Study | 5th Study |
| Heterogeneous data-set with equal BMI ranges |  |  | [18.5,40] | [18.5,40] | [18.5,40] | [18.5,40] | [18.5,40] |
| Non-heterogeneous data-set with different BMI ranges |  |  | [18.5,27] | [21.2,30.2] | [24.5,33.5] | [27.8,36.7] | [31.40] |
| Combined data-set with different BMI ranges and between study differences in the mortality risks. |  |  | [18.5,27] | [21.2,30.2] | [24.5,33.5] | [27.8,36.7] | [31.40] |

The risk of mortality per participant and per study j in the three multiple studies scenarios was generated using the following formulas: