Overview and illustration of methods to investigate effect modification by a continuous covariate

Michail Belias

06 November, 2018

# Abstract (116 out of 200 words)

## Objective

To provide an overview and illustrate a variety of tree-based and regression-based approaches to detect and model effect-modification in meta-analysis(MA) of individual participant data(IPD). For instance, covariate-centred one-stage IPD-MA, mixed effects fractional polynomials, splines, meta-stepp and GLMM-trees, using both two and one stage approach when possible.

## Study Design and Setting

We applied the approaches on two empirical examples. In the first we investigate possible modification of the effect of Somatostatin on liver reduction in participants with polycystic liver disease, in the second effect modification of antibiotics on fever/ear-pain reduction in children with acute otitis media.

## Results

We show that GLMM-trees is a helpful exploratory tool, providing evidence for three-way interactions. On the regression-based methods two-stage approaches showed increased risk of non-convergance and dropping whole trials due to missing data. On the other hand one-stage methods produced results using all information available. Nevertheless, they produced varying results over the power to detect the effect modification, the modeling efficiency and the produced cut-points.

## Conclusion

We conclude that subgroup detection in IPD-MA requires knowing the underlying assumptions and careful modelling. Effect modification may be distorted by unadjusted non-linear associations. We propose the use of more flexible apporaches such GLMM-trees and splines to explore and investigate effect modification.

##### 

# 1. Introduction

Individual participant data meta-analysis (IPD-MA) is a type of systematic review where data gathered from multiple studies are combined and analysed centrally. The capability to standardise subgroup definitions and outcomes across studies, the increased power to investigate other than linear associations, the increased validity and reliability of the subgroups and the flexibility to search for subgroups based on combinations of patient and/or disease characteristics are some of the benefits of using IPD of multiple trials rather than traditional (aggregate) meta-analysis. A vivid field of research towards personalised healthcare is the investigation of effect modification. For this, IPD-MA is considered the gold standard as single trials rarely have sufficient power to identify relevant effect modification. Effect modification may be present in both categorical and/or continuous covariates. For instance, differences in the treatment effect may be present between smokers and non-smokers, or across the levels of the age of the patients. If subgroups are already predefined it may be sufficient to only conduct hypothesis testing, using statistical tools such as generalised linear models or generalised linear mixed-effects models with interaction terms included, combined with meta-analytical tools. However, effect modification across a continuous variable is more challenging, because if the possible subgroups still have to be detected. In that situation, we may need to evaluate effect modification and at the same time want to define a threshold from which point the treatment effect is relevantly different. A common technique to explore subgroups is by starting with the categorization of the continuous covariate. If subgroups are created based on good prior knowledge from literature then this approach is valid. Then, we can proceed as if we had a categorical variable. Nevertheless, categorization has been criticised for misspecification, loss of information and power, inflation of the type I error rate and even biased results[1–5].Another common practice is using the covariable as it is, but assuming linearity over the link function, a method that may also lead to deterioration of power, misspecification, and even spurious results [6]. Therefore, it is important to explore the functional form of the association between outcome and effect modifier. Various approaches to account for non-linear associations have been developed, such as splines and fractional polynomials (FP) [7].

Regression based approaches such as linear models, piecewise polynomials, FPs and smoothing splines may be performed either in one or two stages. In a two-stage approach, first each trial is modelled separately using an appropriate statistical model. Subsequently, we pool either the extracted coefficients if shared across the trials or their fitted functions, using standard meta-analytical tools. In contrast, in one-stage IPD-MA the IPD from all trials are analysed simultaneously whilst accounting for the clustering of participants within studies. Hereto, we model interactions between treatment and patient-level variables while accounting also for the shape of the associations with the outcome. Recent recommendations suggest mean-centring the potential effect modifiers per trial in order to account for potential ecological bias due to unadjusted confounding. In such a one-stage model, within-trial clustering can be accounted for using either a fixed effect (common intercept/slope), fixed effects (stratified intercept/slope), or random effects [9]. Other methods to explore effect modification are plot- and tree-based methods such as the generalised linear mixed-effects model tree (GLMM-tree) method [10]or meta-stepp, a plot based moving average (sliding window) method. Clinical decision making is often based on subgroup categories, separated by cut-points where the treatment effect is changing. The full association between the outcome and the continuous effect modifier is most informative and should be the start for the definition of these cut-points. Still, the locations of these cut-points are not straightforward. They may be based on point estimates or confidence intervals of the treatment effect function, i.e. the difference between the two treatments over the range of the covariable [11]. Interaction term analysis [12] may also be a suitable alternative.

Given all these methods, it is often unclear which method should be preferred in which situation. We aim to describe and illustrate the aforementioned methods by applying them on two empirical examples, while discussing the pros and cons.

# 2. Empirical examples

We use 2 IPD-sets to illustrate our methods. The first data-set [13] was investigating the effect of antibiotics in acute otitis media on children aged from 0 to 12 years old. Rovers et al. collected IPD from 6 randomised clinical trials with a total of 1643 children, aged from 0-12 years old. The primary outcome was fever and/or ear-pain after 3-7 days (yes/no). They concluded that antibiotics were more beneficial in younger children (less than 2 years old) with bilateral acute otitis media. Bilateral acute otitis media (yes/no), age, otorrhea were investigated also separately for potential effect modification and only bilateral acute otitis media showed a significant result. The second data-set [14] was used to investigate the effect of Somatostatin on liver volume reduction. Gevers et al. collected IPD from 3 randomised placebo-controlled trials with a total of 107 participants. In this example, the outcome was continuous (liver volume reduction), and age, sex, baseline liver volume, and diagnosis of either autosomal dominant polycystic liver or kidney disease were investigated for effect modification. They concluded that use of Somatostatin was more beneficial for younger (<47) female patients. One of the 3 trials had a cross-over design, therefore participants were treated both with the active and the control treatment in different time periods. In order to use these data for our illustrative purposes, we removed the cross-over design and used all patients only once, by selecting half of the patients from the active period and the other half (sex and age-matched) from the control period. Therefore, differences between our results and those reported in the original article will occur.

# 3. Methods

## 3.1 Notation

As described in section 2, both our datasets are composed of multiple trials. Therefore, we introduce some notation that will be used hereto. We are denoting:

* The studies as j = 1,2, …,J,
* Individuals as i = 1,2, …,I,
* The per trial mean of age as
* The per trial centred age as =
* Cut-point (knot) as
* The degree of a fractional polynomial as m
* The degree of a polynomial as p

## 3.2. Recursive-partitioning (tree-based) methods

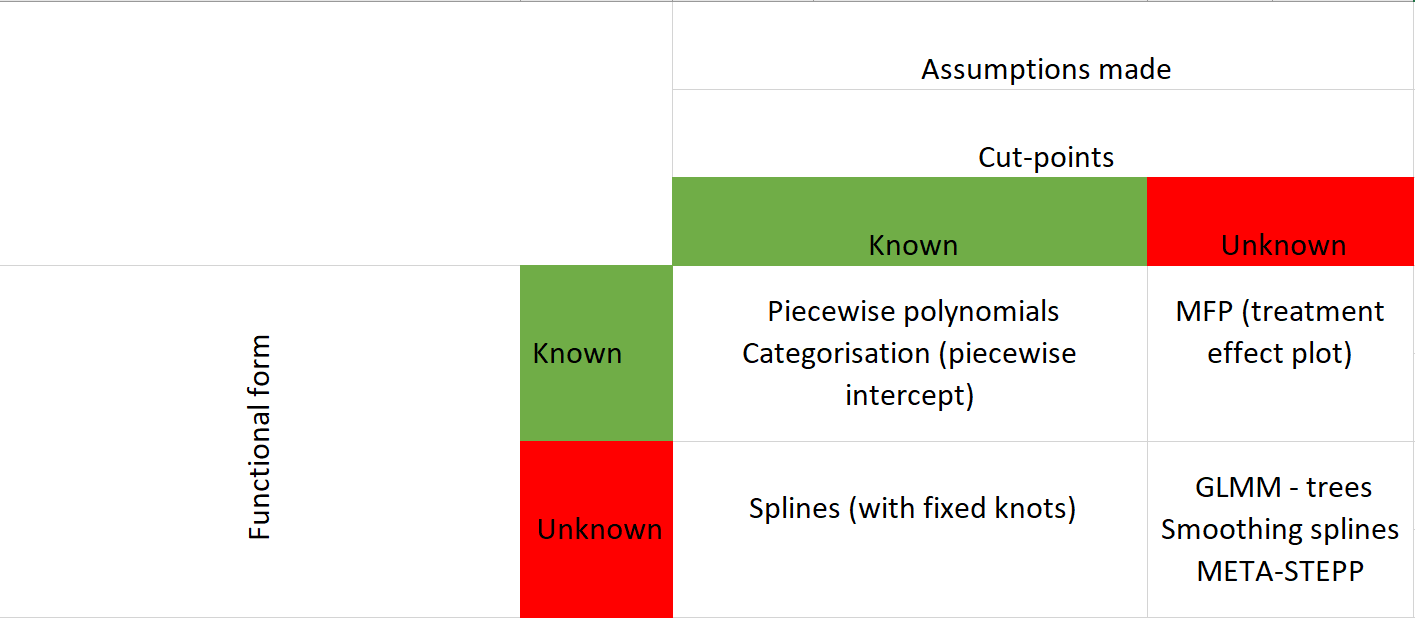
Recursive partitioning can be a first step to explore the underlying structure of the data, such as whether there are outcome differences across the levels of a continuous or categorical variable. It is a statistical method typically used in multivariable analysis [15]. A decision tree is generated by dichotomising the variable by cut-points creating subgroups. These subgroups are generated such that they have the highest difference in outcome. Recursive partitioning techniques can handle non-linear associations as they make no functional form assumptions. However, splitting the data-set into groups is a process that is sensitive to minor changes in the data, which may lead to completely different trees.

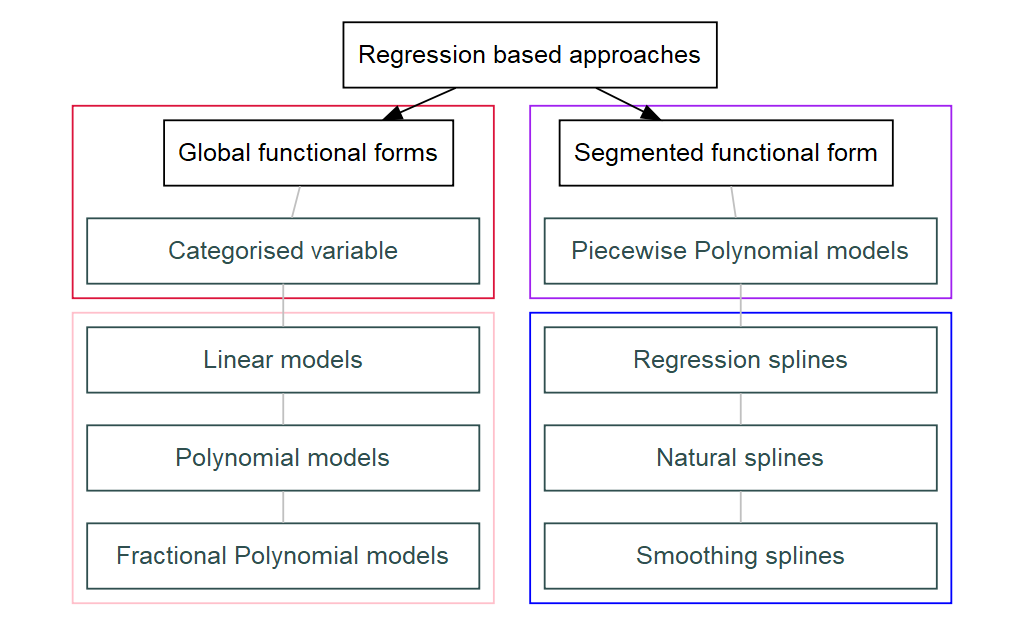
The Generalised linear mixed model (GLMM) tree approach is a state-of-the-art technique using a model-based recursive partitioning [16,17] algorithm. It was proposed by Fokkema et al. [18] to detect treatment-effect modifier interactions when we have clustered data, as in IPD-MA, where the participants are clustered within the trials. Since this approach is exploratory, we can play with the significance level for detecting subgroups, to be more or less liberal. Furthermore, we can use GLMM-trees to detect the possibly relevant variables, their crude cut-points and their underlying subgroups. If GLMM-tree results in subgroups, we can evaluate the corresponding variable using a more sophisticated approach. For example, if we identify age and gender as subgrouping variables using the GLMM-tree, we can use a regression technique to model this relationship.

The GLMM tree algorithm (1) starts by fitting a regression model with interaction terms (treatment x subgroup variables) included (2) statistically tests the interaction terms using a user-defined α, i.e. whether treatment is modified with respect to the subgroup variables, then (3) if test indicates modification, splits the dataset in order to create 2 subgroups with the largest difference in treatment effect. This procedure is (4) repeated in each of the resulting subgroups, until a user-defined convergence is achieved.

## 3.3 Regression based approaches

In regression-based approaches it is important to model the appropriate functional form of the association between the effect modifier and the outcome. If the appropriate form is already known, modelling will be straightforward. However, if not, splines and/or fractional polynomials may be used to evaluate the functional form of the association. Another aspect to take into consideration while modelling the functional form, is whether there may be cut-points (knots) where the shape of the association changes. If these exist, this should be also taken into consideration. Again, if the cut-points are known, modelling will be more straightforward than when they are unknown. These two dimensions define how to proceed with the regression model.





Finally, within-trial clustering should be accounted for[19]. This may be achieved, as described in the introduction, either by conducting a two-stage meta-analysis, or by using a mixed effects models or an equivalent method in a one-stage approach. We will describe both, starting with approaches based on global functional forms as they make more assumptions, and adding gradually more flexibility with the segmented functional form approaches.

### 3.3.1. Two-stage approaches

In two-stage approaches a statistical model of choice is directly fitted per trial. The statistical model per trial j is as follows:

Subsequently, we can either pool the coefficients or the fitted functions using typical meta-analytical tools.

#### 3.3.1.1 First stage: Per-trial modelling

The functions are providing the functional shape of the outcome-effect modifier association per trial j for the treated and the control respectively. Depending on the a priori knowledge of the association’s functional form and the cut-points where the effect is altered we may fit.

*Known functional form and known knots (global functions)*

* Global-polynomials:

- Global polynomials typically are limited to either linear, Quadratic or Cubic, but can go to higher degrees.

*Known functional form and known knots and known position of the knots (segmented functions)*

* Piecewise-polynomials with knots
* Piecewise-polynomials mostly used are piecewise constant, linear, quadratic and cubic.

*Unknown functional form and known knots (global function)*

Fractional polynomials [20] are an extension of polynomials, that also include negative powers. FPs provide a global functional form. FPs model the effect of a covariate X as , where m is the degree of the fractional polynomial and the power is derived from a fixed set of powers {-2,-1,-0.5, 0=(log),0.5,1,2}. The Fractional selection procedure FSP algorithm (FSP) has been proposed [21] to explore the best fitting fractional polynomial. The fractional polynomials of a common degree **m** are tested using the deviance difference criterion, whilst fractional polynomials of different degree are compared using a test. When multiple data-sets are present Sauerbrei and Royston [7], have proposed 3 methods to evaluate the general functional form.

* Overall FP, where the FSP is applied in the pooled data, in order to find the best FP (stratified by data-set).
* Study-wise FP2, the best FP2 is selected for each study
* Study wise selected FP, where the best fitting FP is extracted per study

*Unknown functional form and known knots*

Splines are a generalisation of piecewise polynomials and can offer great flexibility to explore the shape of the outcome-effect modifier association. Regression splines of **p** degree should be continuous, have **p-1** continuous derivatives and the derivative should constant across the knots. They are quite similar to piecewise polynomials, with the difference that they are continuous across the knots. A natural spline has an extra assumption that the second derivative of the function over the edges of the association is 0. This is a something that should be considered when the goal of our research is to forecast future outcomes, as for instance in longitudinal or time series studies. When using splines, the real underlying shape is not known or we don’t want to assume it is known, therefore, we explore it. However, information over the position of the knots and their number may be also unknown. Thus, we can introduce even more flexibility into our model by fitting smoothing splines. Smoothing splines by-pass the problem of knot selection by shrinking the coefficients to their basis expansion, which is a piecewise polynomial. In order to do so, they minimise the penalised least squares criterion or equivalently the maximum likelihood criterion with an extra parameter representing the wingliness of the line, equivalently or in algebraic form .

#### 3.3.1.2 Second-stage combination of the first stage results

As a second-stage in the two-stage IPD-MA, we may either pool the estimates or the fitted functions , extracted from the first stage. The simplest approach is to pool the extracted per trial into one , using a multi-variate meta-analysis, whilst assuming common effect or random effects. This approach only works when common powers are present across studies. Therefore, it is applicable in piecewise, global polynomials, and fractional polynomials fitted using the overall FP procedure. We applied a random-effects meta-analysis, using the EB method for the estimation [23] and the HKSJ adjustment [24].

Another suggested pooling method is to pool the fitted functions. For each x in the data (pointwise) we calculate per study the fitted function and its standard error from . This is the per trial predicted line or equivalently the mean expected outcome for given X E(g(x)|X) and the is the confidence intervals of the predicted line. Afterwards, for each x in the data (pointwise), we calculate the pooled estimate and its SE(x), using either common or random-effects meta-analysis [26].

### 3.3.2. One-stage approaches

#### 3.3.2.1 Centred One-stage IPD-MA

We follow recent recommendations [27] and centre per trial the effect modifier. This way we can separate the within and across trial information of the effect modification. As in the two stage methods we can fit piecewise and global polynomials, but using a mixed-effect model to account for within trials clustering. Therefore, assuming that the statistical model will be:

The and can be either piecewise polynomials, global polynomials or splines as described in section *3.3.1.1*. This

For the and we can assume, either fixed (common) effect, fixed effect*s* (stratified betas), or random effects. If we choose the fixed effect approach a common beta is assumed, in the stratified approach j betas will be generated which correspond to the per trial beta, while in the random effects we assume that the per trial coefficients are driven from a common Normal distribution N().

#### 3.3.2.2 Multilevel Fractional polynomials

Fractional polynomials may be used using one stage approach [28]. In this case, we use the same set of powers as in the FSP method. Furthermore, we fit a mixed effect model of our choice, with either stratified, fixed or random effects. For model selection we can use the lowest deviance or the Akaike Information Criterion (AIC) [30], or Bayesian Information Criterion (BIC) [31].

Therefore, the statistical model applied is:

##### 

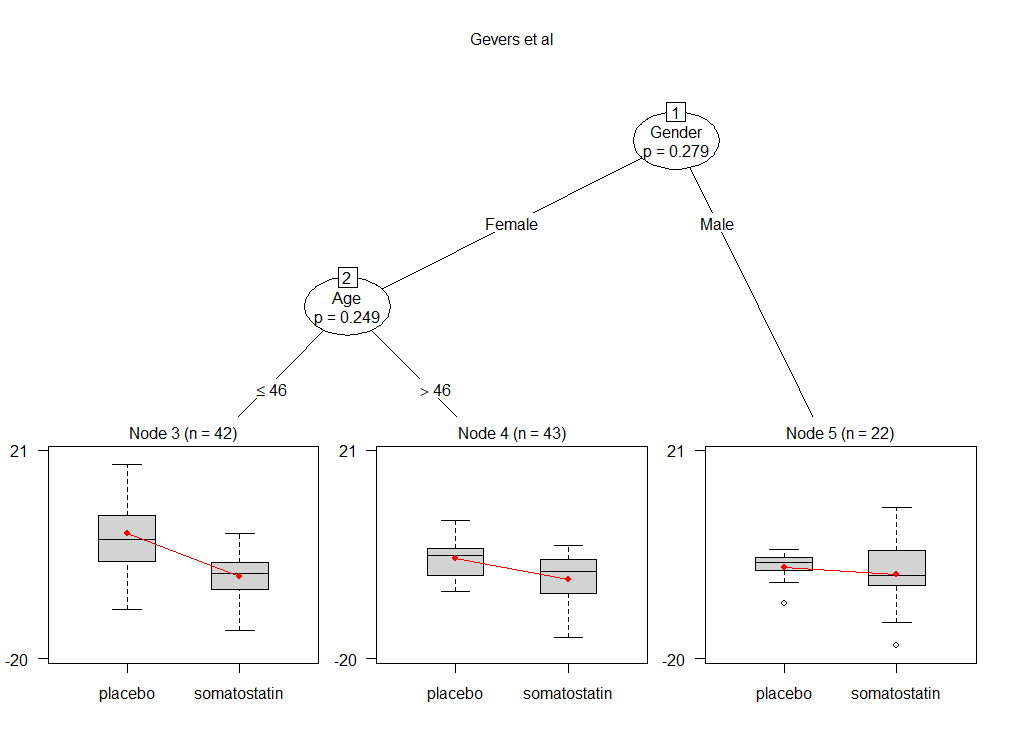
# Results

We will illustrate the results of the aforementioned approaches beginning with the more exploratory to the more confirmatory methods. Therefore, we will start with the tree-based approach and smoothing splines, as we believe that these should be the first exploratory steps. Subsequently, we will show the results of FPs and we end with generalised linear models.

## GLMM-Trees

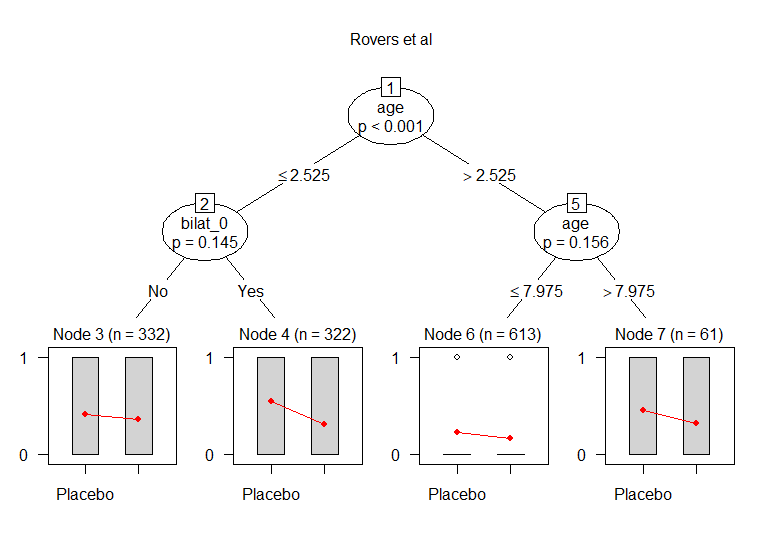
### FIGURE.1 MPLA MPLA

In the Gevers et al . data-set we increased the level of significance into =0.3 , due to small sample size per trial.



We show that there is a difference between the treatment effects of males and females. Furthermore, age seems to be an effect modifier in the subgroup of females (cut-point at 46 years).

### FIGURE.2 MPLA MPLA



On the Rovers et al., dataset we show that age is a potential effect modifier and we may also have a three-way interaction with bilateral otitis.

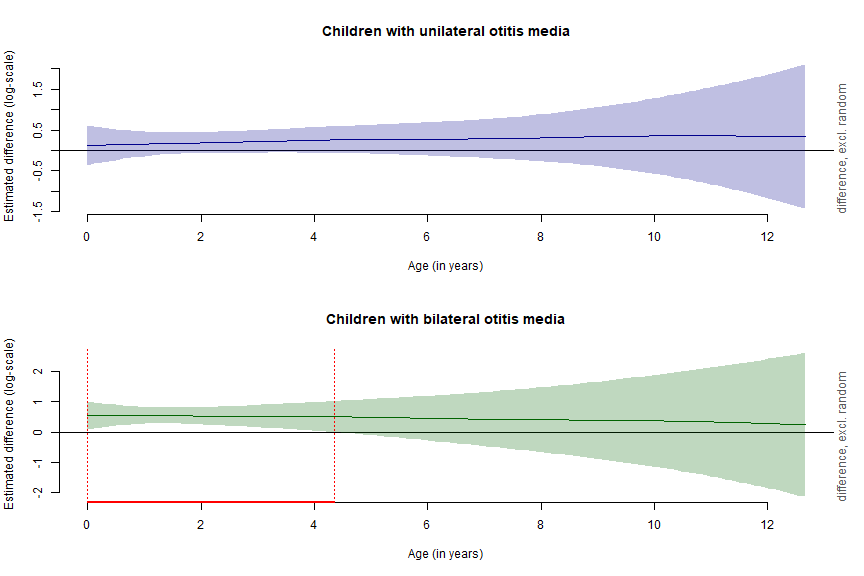
## Regression-based approaches

We will show the results for each approach, first when performing a two-stage and then when using a one-stage method. The somatostatin data set provided trials with few participants, see table 1. For instance, Caroli et.al had only 1 female participant in the somatostatin group. On the other hand, Keimpema et al. and Hogan et al. had limited participation of men. Therefore, the two-stage approaches were limited to linear assumptions, and suffered from trial dropouts.

## Two-stage smoothing splines

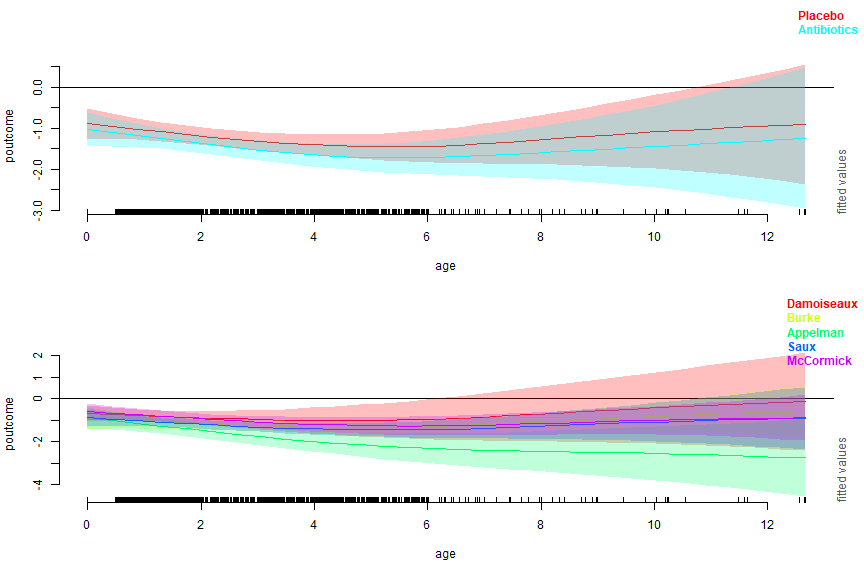
In contrast with linear regression models, in smoothing spline models we cannot interpret the shape of the regression line from the summary. Therefore, visualization is an important tool to investigate effect modification. We show that women younger than 57 have a treatment benefit, probably due to menopause.

In the acute otitis media data-set we show that children 0-5 years have a treatment benefit.

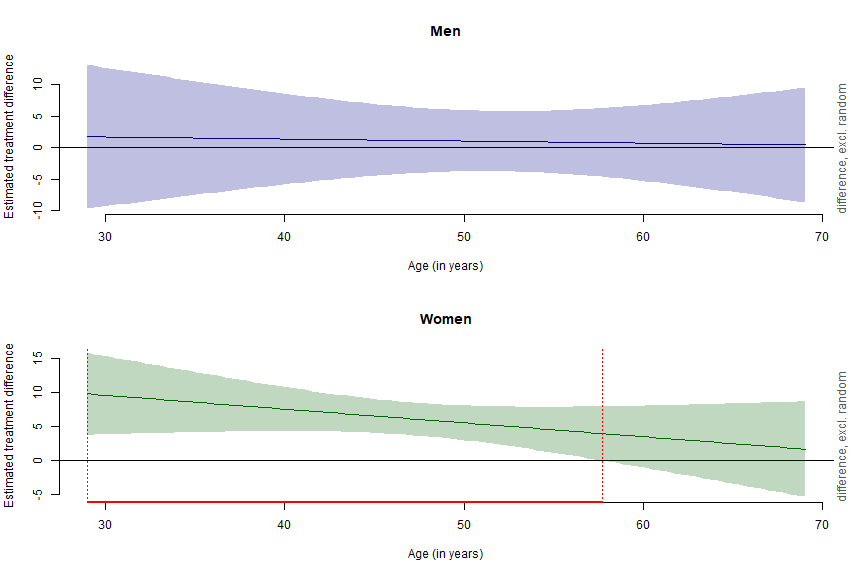


Summary:  
 \* bilat\_0 : factor; set to the value(s): No.   
 \* treat : factor; set to the value(s): Antibiotics, Placebo.   
 \* age : numeric predictor; with 30 values ranging from 0.000000 to 12.655556.   
 \* study : factor; set to the value(s): Saux.

Summary:  
 \* bilat\_0 : factor; set to the value(s): No.   
 \* treat : factor; set to the value(s): Placebo.   
 \* age : numeric predictor; with 30 values ranging from 0.000000 to 12.655556.   
 \* study : factor; set to the value(s): Appelman, Burke, Damoiseaux, McCormick, Saux.



## One-stage smoothing splines



## Two-stage Fractional Polynomials

*Overall FP, where the FSP is applied in the pooled data, in order to find the best FP (stratified by data-set).*

We extracted the fractional polynomial transformations for age per treatment group and stratified per Study. Then we fit a GLM per trial with this overall FP. The FP for the somatostatin data was linear. In this method we couldn’t investigate threeway interactions, as 2 studies had insufficient **female** participants and limited sample size.

Little et al. had no information over the bilateral otitis media and Appelman et al. didn’t have enough participants to investigate the FP shape. So we droped these studies.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | z | Pr(>|z|) | 95%ci.lb | 95%ci.ub |
| Intercept | -0.6923949 | 0.2437843 | -2.8401954 | 0.0045086 | -1.1702033 | -0.2145865 |
| Treatment | -0.3445359 | 0.1745985 | -1.9733036 | 0.0484610 | -0.6867428 | -0.0023291 |
| FP(Age) | -0.0547115 | 0.5665480 | -0.0965699 | 0.9230679 | -1.1651253 | 1.0557022 |
| Bilateral | 0.1979062 | 0.2989969 | 0.6619006 | 0.5080350 | -0.3881170 | 0.7839295 |
| Treatment x FP(Age) | -0.4483191 | 0.7706942 | -0.5817082 | 0.5607633 | -1.9588520 | 1.0622138 |
| Treatment x Bilateral | -0.3637427 | 0.3259802 | -1.1158431 | 0.2644893 | -1.0026521 | 0.2751666 |
| Treatment x FP(Age) | -3.2650027 | 1.3581054 | -2.4040864 | 0.0162129 | -5.9268402 | -0.6031651 |
| Treatment x FP(Age) x Bilateral | 0.5487662 | 1.9508748 | 0.2812924 | 0.7784862 | -3.2748781 | 4.3724104 |

Chi-squared test:  
X2 = 9.309, df = 3, P(> X2) = 0.02545

We couldn’t show that age is an effect modifier on 0.05 .

## Study-wise FP2, the best FP2 is selected for each study

The best fitting FP2s for the somatostatin per trial suffered from the same problems as the overall FP.

Only 2 studies (Damoiseaux, Burke) could converge using the FP2 method.

*Study wise selected FP, where the best fitting FP is extracted per study*

The best fitting FPs for the somatostatin per trial were linear, so the analysis is identical with the two-stage global polynomial.

*Two-stage Global polynomials (coefficient pooling)*

For the Gevers et al. data-analysis, we assumed a linear functional form, as we had limited data (108 observations) and spending (degree of polynomial) (trial number) was considered inefficient. Furthermore, the initial pooled plot showed no significant non-linearity. In contrast, for the Rovers et al. we observed an overall quadratic shape. Damoiseaux et al. had age rounded and the participants were only 1- and 2-years old children. To avoid non-convergence of the log-binomial models we created a slight artificial deviation for the Damoiseaux et al. using the **jitter()** command.

In two-stage meta-analysis using linear assumptions, effect modification across age was not statistically significant.

For the Gevers et al the two-stage meta-analysis of interaction terms showed no statistically significant results.

Little et al had no bilateral information. The study was dropped.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | z | Pr(>|z|) | 95%ci.lb | 95%ci.ub |
| Intercept | 0.0669383 | 0.2821166 | 0.2372718 | 0.8124460 | -0.4860001 | 0.6198767 |
| Treatment | 0.0978052 | 0.4139674 | 0.2362630 | 0.8132286 | -0.7135561 | 0.9091664 |
| Age | -0.4653124 | 0.1394981 | -3.3356175 | 0.0008511 | -0.7387237 | -0.1919011 |
| Bilateral | 0.5250112 | 0.4494114 | 1.1682195 | 0.2427182 | -0.3558190 | 1.4058414 |
| Age^2 | 0.0429164 | 0.0142156 | 3.0189784 | 0.0025363 | 0.0150545 | 0.0707784 |
| Treatment x Age | -0.2684232 | 0.2187709 | -1.2269601 | 0.2198376 | -0.6972064 | 0.1603600 |
| Treatment x Bilateral | -1.6486816 | 0.7156003 | -2.3039142 | 0.0212275 | -3.0512323 | -0.2461308 |
| Age x Bilateral | 0.0209487 | 0.3030520 | 0.0691258 | 0.9448894 | -0.5730224 | 0.6149198 |
| Treatment x Age^2 | 0.0306737 | 0.0232333 | 1.3202458 | 0.1867530 | -0.0148628 | 0.0762102 |
| Age^2 x Bilateral | -0.0071464 | 0.0382012 | -0.1870718 | 0.8516043 | -0.0820194 | 0.0677267 |
| Treatment x Age x Bilateral | 0.8894708 | 0.5336328 | 1.6668218 | 0.0955499 | -0.1564303 | 1.9353720 |
| Treatment x Age^2 x Bilateral | -0.1071213 | 0.0764659 | -1.4009031 | 0.1612431 | -0.2569916 | 0.0427491 |

Chi-squared test:  
X2 = 3.762, df = 4, P(> X2) = 0.4391

*One-stage IPD-MA*

##### 

# Discussion

##### 

# References

[1] Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: A bad idea. Statistics in Medicine 2005;25:127–41. doi:[10.1002/sim.2331](https://doi.org/10.1002/sim.2331).

[2] Altman DG. The cost of dichotomising continuous variables. BMJ 2006;332:1080–0. doi:[10.1136/bmj.332.7549.1080](https://doi.org/10.1136/bmj.332.7549.1080).

[3] Austin PC, Brunner LJ. Inflation of the type i error rate when a continuous confounding variable is categorized in logistic regression analyses. Statistics in Medicine 2004;23:1159–78. doi:[10.1002/sim.1687](https://doi.org/10.1002/sim.1687).

[4] Maxwell SE, Delaney HD. Bivariate median splits and spurious statistical significance. Psychological Bulletin 1993;113:181–90. doi:[10.1037/0033-2909.113.1.181](https://doi.org/10.1037/0033-2909.113.1.181).

[5] Weinberg C. How bad is categorization? Epidemiology 1995;6:345–6. doi:[10.1097/00001648-199507000-00002](https://doi.org/10.1097/00001648-199507000-00002).

[6] Jørgensen TSH, Osler M, Ängquist LH, Zimmermann E, Christensen GT, Sørensen TI. The u-shaped association of body mass index with mortality: Influence of the traits height, intelligence, and education. Obesity 2016;24:2240–7. doi:[10.1002/oby.21615](https://doi.org/10.1002/oby.21615).

[7] Sauerbrei W, Royston P. A new strategy for meta-analysis of continuous covariates in observational studies. Statistics in Medicine 2011;30:3341–60. doi:[10.1002/sim.4333](https://doi.org/10.1002/sim.4333).

[8] Royston P, Sauerbrei W. Interaction of treatment with a continuous variable: Simulation study of significance level for several methods of analysis. Statistics in Medicine 2013;32:3788–803. doi:[10.1002/sim.5813](https://doi.org/10.1002/sim.5813).

[9] Legha A, Riley RD, Ensor J, Snell KI, Morris TP, Burke DL. Individual participant data meta-analysis of continuous outcomes: A comparison of approaches for specifying and estimating one-stage models. Statistics in Medicine n.d. doi:[10.1002/sim.7930](https://doi.org/10.1002/sim.7930).

[10] Wang XV, Cole B, Bonetti M, Gelber RD. Meta-STEPP: Subpopulation treatment effect pattern plot for individual patient data meta-analysis. Statistics in Medicine 2016;35:3704–16. doi:[10.1002/sim.6958](https://doi.org/10.1002/sim.6958).

[11] Royston P, Sauerbrei W. Interactions between treatment and continuous covariates: A step toward individualizing therapy. Journal of Clinical Oncology 2008;26:1397–9. doi:[10.1200/jco.2007.14.8981](https://doi.org/10.1200/jco.2007.14.8981).

[12] Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. BMJ 2010;340:c117–7. doi:[10.1136/bmj.c117](https://doi.org/10.1136/bmj.c117).

[13] 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. doi:[10.1016/s0140-6736(06)69606-2](https://doi.org/10.1016/s0140-6736(06)69606-2).

[14] Gevers TJ, Inthout J, Caroli A, Ruggenenti P, Hogan MC, Torres VE, et al. Young women with polycystic liver disease respond best to somatostatin analogues: A pooled analysis of individual patient data. Gastroenterology 2013;145:357–365.e2. doi:[10.1053/j.gastro.2013.04.055](https://doi.org/10.1053/j.gastro.2013.04.055).

[15] Breiman L. Classification and regression trees. Routledge; 2017. doi:[10.1201/9781315139470](https://doi.org/10.1201/9781315139470).

[16] Zeileis A, Hothorn T, Hornik K. Model-based recursive partitioning. Journal of Computational and Graphical Statistics 2008;17:492–514. doi:[10.1198/106186008x319331](https://doi.org/10.1198/106186008x319331).

[17] Su X, Tsai C-L, Wang H, Nickerson DM, Li B. Subgroup analysis via recursive partitioning. SSRN Electronic Journal 2009. doi:[10.2139/ssrn.1341380](https://doi.org/10.2139/ssrn.1341380).

[18] Fokkema M, Smits N, Zeileis A, Hothorn T, Kelderman H. Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. Behavior Research Methods 2017. doi:[10.3758/s13428-017-0971-x](https://doi.org/10.3758/s13428-017-0971-x).

[19] Abo-Zaid G, Guo B, Deeks JJ, Debray TP, Steyerberg EW, Moons KG, et al. Individual participant data meta-analyses should not ignore clustering. Journal of Clinical Epidemiology 2013;66:865–873.e4. doi:[10.1016/j.jclinepi.2012.12.017](https://doi.org/10.1016/j.jclinepi.2012.12.017).

[20] Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. Applied Statistics 1994;43:429. doi:[10.2307/2986270](https://doi.org/10.2307/2986270).

[21] Ambler G, Royston P. Fractional polynomial model selection procedures: Investigation of type i error rate. Journal of Statistical Computation and Simulation 2001;69:89–108. doi:[10.1080/00949650108812083](https://doi.org/10.1080/00949650108812083).

[22] Sauerbrei W, Royston P. Building multivariable prognostic and diagnostic models: Transformation of the predictors by using fractional polynomials. Journal of the Royal Statistical Society: Series A (Statistics in Society) 1999;162:71–94. doi:[10.1111/1467-985x.00122](https://doi.org/10.1111/1467-985x.00122).

[23] Stijnen T, Houwelingen JCV. Empirical bayes methods in clinical trials meta-analysis. Biometrical Journal 2007;32:335–46. doi:[10.1002/bimj.4710320316](https://doi.org/10.1002/bimj.4710320316).

[24] Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. Statistics in Medicine 2001;20:3875–89. doi:[10.1002/sim.1009](https://doi.org/10.1002/sim.1009).

[25] IntHout J, Ioannidis JP, Borm GF. The hartung-knapp-sidik-jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-laird method. BMC Medical Research Methodology 2014;14. doi:[10.1186/1471-2288-14-25](https://doi.org/10.1186/1471-2288-14-25).

[26] White IR, Kaptoge S, Royston P, and WS. Meta-analysis of non-linear exposure-outcome relationships using individual participant data: A comparison of two methods. Statistics in Medicine 2018. doi:[10.1002/sim.7974](https://doi.org/10.1002/sim.7974).

[27] Hua H, Burke DL, Crowther MJ, Ensor J, Smith CT, 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. Statistics in Medicine 2016;36:772–89. doi:[10.1002/sim.7171](https://doi.org/10.1002/sim.7171).

[28] Long J, Ryoo J. Using fractional polynomials to model non-linear trends in longitudinal data. British Journal of Mathematical and Statistical Psychology 2010;63:177–203. doi:[10.1348/000711009x431509](https://doi.org/10.1348/000711009x431509).

[29] Johnson W, Balakrishna N, Griffiths PL. Modeling physical growth using mixed effects models. American Journal of Physical Anthropology 2012;150:58–67. doi:[10.1002/ajpa.22128](https://doi.org/10.1002/ajpa.22128).

[30] Akaike H. A new look at the statistical model identification. In:. Springer series in statistics, Springer New York; 1974, pp. 215–22. doi:[10.1007/978-1-4612-1694-0\_16](https://doi.org/10.1007/978-1-4612-1694-0_16).

[31] Schwarz G. Estimating the dimension of a model. The Annals of Statistics 1978;6:461–4. doi:[10.1214/aos/1176344136](https://doi.org/10.1214/aos/1176344136).