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

*Michail Belias*

*08 October, 2018*

## Abstract (116 out of 200 words)

## Objective

To 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), such as: covariate-centred IPD-MA, mixed effects fractional polynomials, splines, meta-stepp and glmm-trees.

### Study Design and Setting

We applied the aforementioned approaches into two empirical data-sets. The first is investigating the effect of somatostatin treatment versus placebo in liver reduction percentage, on participants with polycystic liver disease. The second investigates the effect of antibiotics in fever/earpain reduction, on children with acute otitis media(AOM).

### Results

Non-linear association was detected in AOM IPD-MA.

### Conclusion

We conclude that subgroup detection in IPD-MA requires knowing the underlying assumptions and careful modelling. Effect modification may be distorted by a non-linear association if left unadjusted.

# 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 personalized healthcare is the investigation of subgroup effects. For this task, IPD-MA is considered a gold standard as single trials rarely have sufficient power to identify relevant effect modification.

Effect modification is known to be present in both categorical and/or continuous covariates. For example, differences in the treatment effect may be present between smokers and non-smokers. In this case, subgroups are already defined and therefore, only hypothesis testing may be conducted. The investigation of subgroup effects is typically conducted using statistical tools, such as t-tests, contingency tables or fitting an appropriate generalised linear model (GLM) with interaction terms included. On the other hand, effect modification across a continuous covariate may be more challenging. A commonly applied method is to categorize the continuous covariate. This method has been criticised for misspecification, loss of information and power, inflation of the type I error rate, when adjusting for confounding and biased results [1–5]. Another common practice is to assume linearity over the link function, a method that may also lead to deterioration of power, misspecification and even spurious results [6]. Therefore, along with the confirmation of a variable as an effect modifier, we may have to the explore the functional form of the outcome-effect modifier association also. Furthermore, clinical desicions are based in cut-points in which the treatment effect is altered. These cut-points may be altered if the outcome-effect modifier functional form is misspecified.

IPD-MA using regression-based techniques, may be conducted either in one or two stages. In two-stage IPD-MA, each trial is first analysed separately, using an appropriate statistical model. For instance, the first stage may estimate the main treatment effect, or the treatment-covariate interaction effect. Subsequently, these effects are combined into a summary estimate in the second stage of the meta-analysis. For instance, subgroup effects may be investigated by modelling the association of the estimated main treatment effect with a trial-level covariate (meta-regression) or by pooling the interaction terms estimated in the first stage (two-stage meta-analysis of interaction terms). Meta-regression and per-subgroup meta-analysis have been critisized for lack of power and biased results [7], in subgroups effects investigation. Therefore, we will only consider two-stage meta-analysis of interaction terms.

In one-stage IPD-MA, all IPD from every trial are analysed simultaneously whilst accounting for the clustering of participants within studies. Hereby, researchers may model interactions between treatment and patient-level covariates either directly (naive one-stage IPD-MA), or after the covariates are mean-centred per study in order to account for potential ecological bias (centred one-stage IPD-MA). *Clustering can be also accounted using stratified, fixed or random intercept and/slopes [10]* Finally, state of the art plot and tree-based methods are developed for subgroup effects investigation. Generalized linear mixed-effects model trees(glmertree) introduced by Fokkema et al.[11] is accounting for non-linear associations, while also adjusting for within studies clustering. Meta-stepp is a plot based moving average method that approximates non-linear effects from clustered data [12].

It is often unclear when each method should be preferred. Also, it is unclear if the treatment effect function [13] or interaction term analysis [14] is most appropriate and when. We aim to describe and illustrate the aforementioned methods. For that task, we will use both regression based approaches such as meta-stepp, centred one-stage IPD-MA, mixed effects fractional polynomials and splines, and tree based approaches such as generalized linear mixed-effects model trees.

# 2. Methods

## 2.1 Data-sets

We used 2 empirical IPD-sets. The first data-set [15] 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). Rovers et al. 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 also investigated separately for potential effect modification and only bilateral acute otitis media showed a significant result.

The second data-set [16] was investigating the effect of somatostatin in the liver volume reduction. Gevers et al. collected IPD from 3 randomized placebo-controlled trials with a total of 107 participants. The outcome was continuous (liver volume reduction) and age, sex, baseline liver volume, and diagnosis of autosomal dominant polycystic liver or kidney disease has been investigated for effect modification. Gevers et al. concluded that therapy using somatostatin was more beneficial for young female patients. One of the included trials [Caroli et al.] had a cross-over design, threfore participants were in both treatment groups (control and treated) in different time periods. We matched the participants per age and gender and picked half on the treated and half on the control group. Therefore some differences between our results and those reported in the original article may occur.

## 2.2 Statistical approaches

## 2.2.1 Recurcive partitioning methods (tree based)

### Generalised Linear Mixed Model Trees (glmm or glmer trees)

Generalised linear mixed model trees is a state of the art technique, proposed by Fokkema et al [11] for the detection of treatment-effect modifier interaction. The algorithm is based on model-based recursive partitioning [17,18] typically applied, while also taking into account the clustered structure of datasets.

The GLMM tree algorithm:

(1) fit the parametric model to the dataset,
(2) statistically test for parameter instability with respect to each of a set of partitioning variables,
(3) if there is some overall parameter instability, split the dataset with respect to the variable associated with the highest instability,
(4) repeat the procedure in each of the resulting subgroups.

## 2.2.2Regression based methods

### Two-stage meta-analysis of interaction terms

latex In meta-analysis of interaction terms (MA-IT), the interaction between the potential effect modifier and treatment is directly modelled per trial. The statistical model per trial j is as follows:

$g(Y_{ij}) = \beta_{0j} + \beta_{Tj} \times Treatment + \beta_{Aj} \times f(Age) + \beta_{xj} \times Treatment \times f(Age))$

Subsequently, an inverse variance weighted average of $\beta_{xj}$ is calculated.

### Centered One-stage IPD-MA

We follow recent recomendations [19] and centre per trial the effect modifier. This way we separate the within and across trial information.

$g(Y_{ij}) = \beta_{0j} + \beta_{Tj} \times Treatment + \beta_{Aj} \times (Age - \bar{Age}) + \beta_{xj} \times Treatment \times (Age - \bar{Age}))$

mixed

$\beta_{0j} \sim N(\beta_0, \tau_0^2)$

$\beta_{Tj} \sim N(\beta_T, \tau_T^2)$

$\beta_{Aj} \sim N(\beta_A, \tau_A^2)$

$\beta_{xj} \sim N(\beta_x, \tau_x^2)$

### Mixed effect fractional polynomials

We follow recent recomendations [19] and centre per trial the effect modifier. This way we separate the within and across trial information.

$g(Y_{ij}) = \beta_{0j} + \beta_{Tj} \times Treatment + \beta_{Aj} \times fp(Age - \bar{Age}) + \beta_{xj} \times Treatment \times fp(Age - \bar{Age}))$

mixed

$\beta_{0j} \sim N(\beta_0, \tau_0^2)$

$\beta_{Tj} \sim N(\beta_T, \tau_T^2)$

$\beta_{Aj} \sim N(\beta_A, \tau_A^2)$

$\beta_{xj} \sim N(\beta_x, \tau_x^2)$

## 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.

[2] Altman DG. The cost of dichotomising continuous variables. BMJ 2006;332:1080–0. doi: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.

[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.

[5] Weinberg C. How bad is categorization? Epidemiology 1995;6:345–6. doi: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.

[7] Simmonds MC, Higgins JPT. Covariate heterogeneity in meta-analysis: Criteria for deciding between meta-regression and individual patient data. Statistics in Medicine 2007;26:2982–99. doi:10.1002/sim.2768.

[8] 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. doi:10.1136/bmj.j573.

[9] Fisher D, Copas A, Tierney J, Parmar M. A critical review of methods for the assessment of patient-level interactions in individual participant data meta-analysis of randomized trials, and guidance for practitioners. Journal of Clinical Epidemiology 2011;64:949–67. doi:10.1016/j.jclinepi.2010.11.016.

[10] 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.;0. doi:10.1002/sim.7930.

[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. Behavior Research Methods 2017. doi:10.3758/s13428-017-0971-x.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[18] 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.

[19] 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.