Identifying subgroups based on continuous measurements in individual patient data meta-analysis

Michail Belias

03 April, 2018

*Background*

Individual patient data meta-analysis (IPD-MA) is increasingly used to identify relevant subgrouping effects. Often linearity assumptions are made when examining subgroups based on continuous measurements. However, several more flexible methods exist.

*Objectives*

Our goal is to illustrate, critically review and compare state of the art methods on subgroups effects identification in IPD-MA, based on continuous measurements.

*Methods*

We reviewed META-STEPP, generalised additive mixed effects models, (multi-level) regression models involving fractional polynomials or splines and several tree-based approaches. We applied the methods above on two empirical examples: prescription of antibiotics in children with otitis media and anti-platelet treatment in secondary stroke prevention.

*Results*

We will provide treatment effect plots to visualize subgroup effects within and across studies.

*Conclusions*

We provide advantages and limitations of the aforementioned methods.

1.Introduction

Individual participant data meta-analysis (IPD-MA) is a type of systematic review where the data of multiple studies are combined and analyzed centrally. The possibility to standardize subgroup definitions and outcomes across studies, 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 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 subgroups.

Effect modification may be present over 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 applied. The investigation of subgroup effects is typically conducted with 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 is more challenging. The functional form of the association with the outcome and/or the cut-off point where the effect is altered may not be known a priori. Nevertheless, a commonly applied method is to categorize the continuous covariate. This method is 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 also may lead to deterioration of power, misspecification and spurious results [6]. Finally, although relative measures are of great statistical value, clinical decisions are more consistent when based on absolute measures such as risk difference (RD) and its reciprocal numbers needed to treat (NNT) [7,8]. Attention is needed, because shifting from relative to absolute measures is affecting the results. This phenomenon may even lead on different decisions whether to treat or not, depending on the baseline risk distribution [9,10].

Depending on the knowledge we have over the treatment-effect modifier interaction, recursive partitioning or regression-based approaches may be applied. Recursive partitioning, also known as tree based, techniques are commonly implemented when limited knowledge over the distribution of the effect modification is available. These nonparametric approaches are based on categorizing the data in order to form homogeneous groups. Although a variety of recursive partitioning methods have been developed, most of them ignore the within studies clustering of patients typically present in a IPD-MA. This may lead to spurious subgroups [11,12] since a major drawback of recursive partitioning techniques is their instability. Minor changes in the data and/or assumptions, may cause great differences in either or both the selection of splitting variables and/or their cut-points [13]. Three state of the art approaches, i.e. PALM trees [14], IPD-SIDES [15] and generalised linear mixed-effects model trees (GLMM trees) [16] offer the possibility to account for within study patient clustering.

Regression-based approaches better suited than recursive partitioning techniques for modelling the association between an outcome and one or more covariates. The potential effect modifier is already known, and the goal is to test whether a treatment effect difference is present or not. Typically, this is accomplished with hypothesis testing. Regression-based approaches can be applied in either one or two stages. In two-stage approaches, each trial is first analysed separately, using a statistical model of choice. Subsequently, the estimates are pooled in the second stage using meta-analytical methods. For instance, the first stage may estimate the different effects observed per subgroup, or the treatment-covariate interaction effect. We can account for non-linearity either by fitting a polynomial GLM if we know beforehand the shape of the outcome effect modifier association. Otherwise more flexible approaches such as fractional polynomials (FPs)[17,18], polynomial splines or generalised additive models (GAMs)[19,20] may also be preferred . A state of the art approach meta-STEPP [21] is proposed, when categorical outcomes are present for investigating the treatment effect modification over a continuous covariate. It is a plot-based two-stage approach, where a moving average technique is applied. In one-stage approaches, all IPD from every trial are analysed simultaneously whilst accounting for the clustering of participants within studies. It is suggested to per trial mean-center the covariates, in order to account for potential ecological bias and model treatment-effect modifier interactions [22]. It is often unclear whether one-stage or two-stage approaches should be preferred. Also, it is unclear if the treatment effect function [23] or interaction term analysis [24] is more appropriate for clinical decisions. In our study we focus on the clinical perspective of identifying the best treatment for a patient. Therefore, we will examine both the absolute and relative risk difference between treated and control. Our objective is to review available methods for investigating continuous covariate effect modification in multiple IPD-sets.

# 2.Methods

Depending on the research question and the prior knowledge a researcher may choose exploratory and/or confirmatory data analysis. Exploratory data analysis strives on investigating patterns and features of the data that represent deviations from the model, while confirmatory data analysis quantifies these deviations and investigates to what level could they occur by chance (hypothesis testing) [25,26]. These two types of data analysis are usually complementary. For instance, a researcher may be interested in confirming a subgroup effect, while simultaneously exploring it’s -association with the outcome- functional form.

In our study we include six regression-based and three recursive partitioning approaches: META-STEPP, GLMMs, Polynomial GLMs, fractional polynomials, splines, GAMMs, IPD-SIDES, GLMM Trees, PALM trees, see *Figure.1*. Our aim is to review the aforementioned approaches in the detection of the optimal treatment for a given individual, based on a continuous biomarker. First, we introduce the regression-based approaches and subsequently we describe the recursive partitioning approaches.

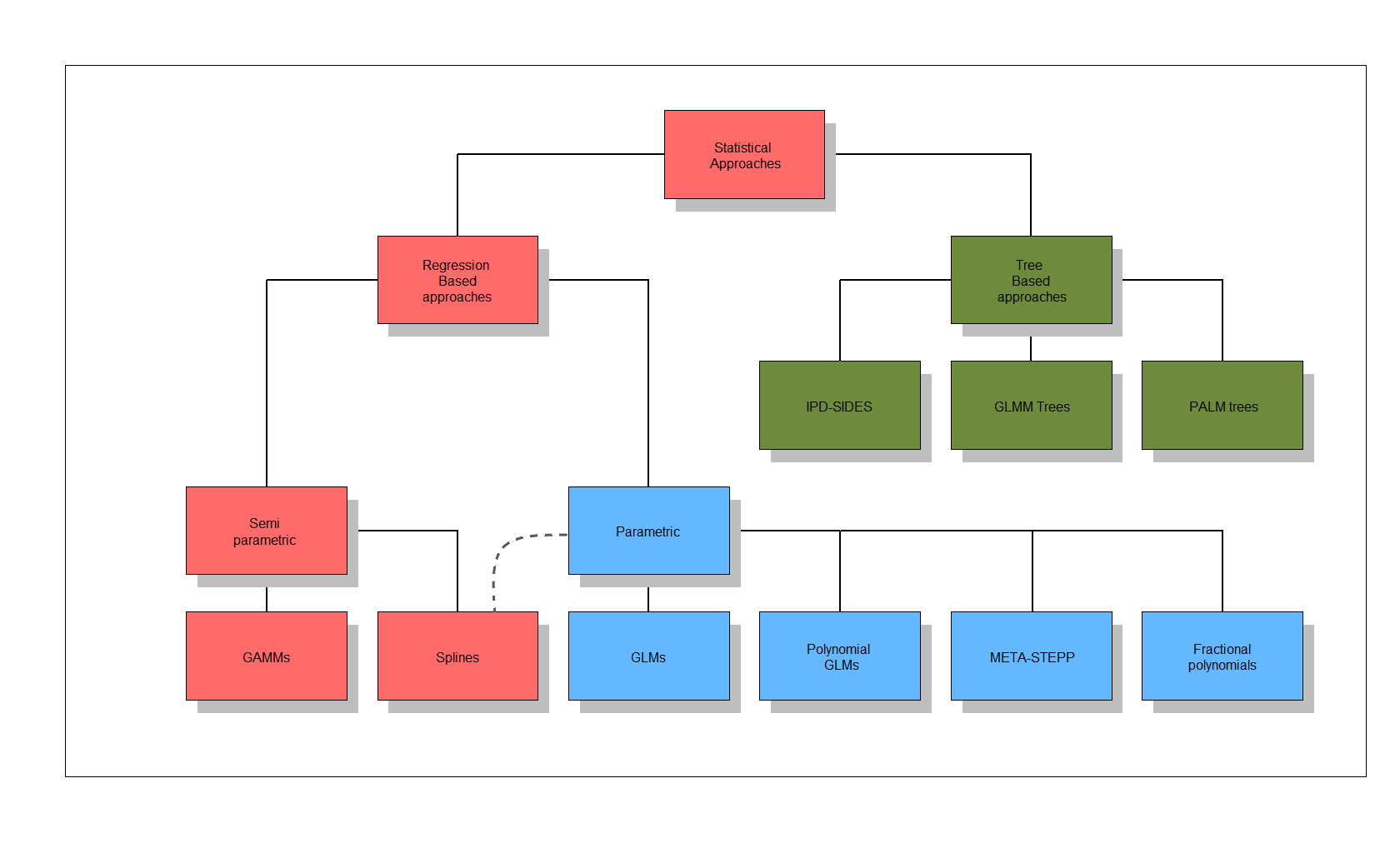


Figure. 1 Diagram of critically reviewed approaches

## 2.1 Regression-based approaches

We identify three variations over the regression-based approaches. We start with the (1) interaction term-based approaches, followed by the (2) treatment effect approaches and finally we review the (3) plot-based approaches. Because we investigate effect modification in an IPD-setting we also have to account for within trial clustering using either one or two stage approaches. Two-stage approaches are intuitively simpler, but come with a potential cost in power, especially when small studies are included [27]. On the other hand, one-stage methods need careful modeling to avoid potential misspecification. Generalised polynomial models are typically applied, but a priori assumptions of the underlying functional form are needed, with linear being the most common made. More flexibility is offered through fractional polynomials, splines and generalised additive models, as the functional shape of the dependent and independent variables association is explored and determined by the data.

### 2.1.1 Methods analysing the treatment covariate interaction terms

#### 2.1.1.1 Two-stage meta-analysis of interaction terms (MA -IT)

In meta-analysis of interaction terms (MA-IT), the interaction between the potential continuous effect modifier and treatment is directly modelled per trial. An appropriate model, logistic, CoxPH or linear, is fitted and subsequently the interaction terms estimates are pooled using meta-analytical methods. The statistical model per trial **j** is as follows:

, where = and and and are per trial intercepts and estimates of treatment. If is equal to one, the model is linear, while second and third-degree polynomials are quadratic and cubic respectively. The estimates are extracted and pooled using meta-analytical methods.

*Meta-analytical methods*

Meta-analysis is a statistical procedure, that pools estimates of interest using a weighted average. In cases where the estimates are assumed to be derived from a common effect and vary due to randomness, fixed effects are used. On contrary when extended heterogeneity is observed then random effects are more appropriate. We applied a random effect meta-analysis for pooling the estimates of interaction. We extracted the pooled estimates and their corresponding p-values.

#### 2.1.1.2 Two-stage meta-analysis fractional polynomials interaction terms (FPMA-IT)

If we don’t want to make assumptions over the functional shape, we can use fractional polynomials and the function selection procedure (FSP) proposed by Royston et al. and Sauerbrei et al [18,28]. As Sauerbrei et al. suggested, we may determine the functional form in each study using three methods:

* The overall fractional polynomial, where the best FP is for the pooled data-set is selected. We used = 1% significance level, as this method uses all of the data and sufficient power to detect even weaker non-linear effects.
  + The statistical model per trial **j** is as follows:

or

, when

* The study wise FP2, where the best FP2 function for each study is selected
  + The statistical model per trial **j** is as follows:
* The study-wise selected FP, where the best FP driven out of the FSP procedure for each study is selected.

The statistical model per trial **j** is as follows:

Linear:

FP1:

FP2:

Afterwards, the average function is generated by either pooling the standardised per study functions, using inverse-variance weighting.

, or by averaging the estimates using either fixed or random effects weights.

#### 2.1.1.3 Centred one-stage IPD-MA

An alternative approach is to investigate the treatment-effect modifier interaction using all available data simultaneously, accounting for within trial clustering. Similarly, with the above we can fit a generalised polynomial mixed effects model.

~

~

For even more flexibility we can fit a generalised additive models (GAMs), an extension of generalised linear models, which we describe above. Being additive the interaction terms are generated per subgroup. One smoothing function for the control and one for the treated is generated. Their main difference is that instead of simple terms we fit a generalised linear model over the , where are non-parametric functions of the independent variable :

# Critique of the methods

## Generalised mixed effects model

)

## Generalised Fractional Polynomial Mixed Model

## Generalised additive mixed effects model (GAMMs)

A generalised additive model [citation] is a generalised linear model with a linear predictor involving a sum of smooth functions of covariates. Equivalently, to generalised linear model

## Recursive partitioning approaches

Recursive partitioning approaches are typically non-parametric, thus high-dimensional interactions can be explored with flexibility accounting for non-linearity. Their intuitive algorithm splits the data in order to form homogeneous groups.

### IPD-SIDES

IPD-SIDES [15] is recursive partitioning method for subgroup effects investigation. The a

We first describe the algorithm for the SIDES procedure followed by a more detailed description of the splitting criterion and the continuation criterion. The algorithm for growing the tree is as follows:

Start at the root node consisting of the entire dataset Step 1 - Evaluate the splitting criterion for all possible splits of every covariate, excluding any covariates already used to define the parent node, retaining only the best split for each covariate. Order the covariates from smallest adjusted P-value to largest adjusted P-value where the adjusted P-values are computed using the Sidak-based multiplicity adjustment which adjusts for the number of splits searched for a given covariate (see below). Step 2 - Select the best M covariates from the ordered best splits. The value of M is specified by the user where the recommended value is 5. For each of the M splits, form the split creating 2 child nodes and retain the child node with the larger positive treatment effect, provided it satisfies the continuation criterion. The retained nodes now become parent nodes for the next iteration. Step 3 – Repeat steps 1 and 2 for the newly formed parent nodes Step 4 – Repeat steps 1 to 3 until either a pre-specified maximum number of levels (L) is reached or if no more splits can be formed, i.e., the continuation criterion is not satisfied. In both cases, the previously formed parent nodes become terminal nodes.

### Generalised Linear Mixed Model Trees (GLMM trees)

Generalised Linear Mixed Model Trees is a state of the art technique, proposed by Fokkema et al [16] for the detection of treatment-effect modifier interaction.

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.

First, we introduce the recursive partitioning techniques and subsequently we describe the regression based methods.Regression-based approaches rely on parametric or semi-parametric assumptions and can be further categorized into one stage and two stage, according to the analysis. Finally, inferences can be made either by investigating the interaction term or by calculating the treatment effect function.

the simplest and most widely used regression based approach is generalised linear models (GLMs) including treatment-effect modifier interaction terms. A drawback of GLMs is that they assume linearity over the link function, which is often not true. Polynomial functions, such as cubic and quadratic, are applied for better fit. Although prior knowledge of the real underlying functional form is required. More flexibility is offered by fractional polynomials [17,18], an approach that uses a predefined set of powers both negative and positive. Fractional polynomials may be fitted in IPD-MA using a two stage approach. Three second stage pooling methods have been proposed. In contrast with these global functional form approaches, piece wise polynomials and splines are also proposed for fitting smoother and flexible non linear functions. In polynomial splines we predefine the number and/or knots (cut-points). Across the continuous covariate data are divided over the knots and a predefined degree polynomial GLM is fitted [29]. If we have no knowledge of the number or position of knots, we can use smoothing splines, loess or penalized, which do not require a predefined number of knots. Instead a *roughness* penalty is required, in order to account for over-fitting, under-fitting and wingliness of the curve. A combination of GLMs and the aforementioned smoothed predictor function is generalised additive models (GAMs) and generalised additive mixed models (GAMMs).

# Application to an IPD meta‐analysis in epilepsy

we apply the aforementioned methods to two empirical example. The first is an IPD meta-analysis of 1.225 patients from five randomized controlled trials, comparing the effects of two antiepileptic drugs [22,30–32]. The second is an IPD meta-analysis from six randomised trials of the effects of antibiotics in children with acute otitis media [33].

## Description of the data.

# Appendix

### Fractional polynomials

Fractional polynomials are the extension of quadratic and cubic polynomials. Royston proposed a set of powers M = {-2,-1,-0.5, 0 (=log), 0.5, 1, 2, 3}. Furthermore, he advocated that in most cases first and second degree FP are sufficient.

* FP1 models: Y =
* FP2 models:
* Y = , for
* Y = , for

Rarely a third or more power parameters are used (FPm models).

Royston et. al has proposed a stepwise forward technique for detecting the appropriate model. The algorithm for optimal selection is :

The best selected model is then tested against the null model, a straight line, and the best-fitting first-degree FP.

#### Meta-analysis extension

Royston et. al [citation] has proposed the following technique for the multi-level framework:

* Step 1. Find the confounder model If multiple variables are present we may need to adjust for confoundship. So the first step is to determine the confounder model. If confounders are not a priori know we derive them per study, and summarise the corresponding per trial linear predictors .
* Step 2. Determine functional form. Rouston et al. [citation] have been proposed three methods for determining the functional form of a continuous predictor in each study. In all the approaches, estimates are adjusted for the confounder models found in step 1.

1. Overall fractional polynomial (FP). Using the function selection procedure (FSP), find the best FP transformation for the pooled data set (stratified by data set). Select using a small nominal significance level. or when .
2. Study-wise FP2. The best FP2 function for each study is selected .
3. Study-wise selected FP. Use the FSP to select the best FP for each study individually. In this method an FP is determined separately in each study. Depending on the statistical power and the strength of the relationship with Z, the selected function in study i may be linear, FP1 or FP2:
   * Linear: or
   * FP1: or
   * FP2:

* Step 3. Averaging the functions

We propose three methods for determining an average function. In each method, estimates are adjusted using the confounder indexes 1;:::;k determined in step 1. 1. Pooled function 2. Function estimated with fixed-effects weights 3. Function estimated with random-effects weights

# Discussion

potential limitations: Two stage fractional polynomials assume common functional form for the Treated and control.

# Bibliography

[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] Cook RJ, Sackett DL. The number needed to treat: A clinically useful measure of treatment effect. BMJ 1995;310:452–4. doi:[10.1136/bmj.310.6977.452](https://doi.org/10.1136/bmj.310.6977.452).

[8] Ranganathan P, Pramesh C, Aggarwal R. Common pitfalls in statistical analysis: Absolute risk reduction, relative risk reduction, and number needed to treat. Perspectives in Clinical Research 2016;7:51. doi:[10.4103/2229-3485.173773](https://doi.org/10.4103/2229-3485.173773).

[9] Dahabreh IJ, Hayward R, Kent DM. Using group data to treat individuals: Understanding heterogeneous treatment effects in the age of precision medicine and patient-centred evidence. International Journal of Epidemiology 2016:dyw125. doi:[10.1093/ije/dyw125](https://doi.org/10.1093/ije/dyw125).

[10] Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients. JAMA 2007;298:1209. doi:[10.1001/jama.298.10.1209](https://doi.org/10.1001/jama.298.10.1209).

[11] Sela RJ, Simonoff JS. RE-EM trees: A data mining approach for longitudinal and clustered data. Machine Learning 2011;86:169–207. doi:[10.1007/s10994-011-5258-3](https://doi.org/10.1007/s10994-011-5258-3).

[12] Martin D. Efficiently exploring multilevel data with recursive partitioning. PhD thesis. University of Virginia, n.d. doi:[10.18130/v3ps0f](https://doi.org/10.18130/v3ps0f).

[13] Philipp M, Zeileis A, Strobl C. A toolkit for stability assessment of tree-based learners. Machine Learning 2016.

[14] Seibold H, Hothorn T, Zeileis A. Generalised linear model trees with global additive effects 2016.

[15] Mistry D, Stallard N, Underwood M. A recursive partitioning approach for subgroup identification in individual patient data meta-analysis. Statistics in Medicine 2018. doi:[10.1002/sim.7609](https://doi.org/10.1002/sim.7609).

[16] 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).

[17] 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).

[18] 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).

[19] Hastie T, Tibshirani R. Generalized additive models. Defense Technical Information Center; 1984. doi:[10.21236/ada147454](https://doi.org/10.21236/ada147454).

[20] Brown RA, Hastie TJ, Tibshirani RJ. Generalized additive models. Biometrics 1991;47:785. doi:[10.2307/2532174](https://doi.org/10.2307/2532174).

[21] 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).

[22] 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).

[23] 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).

[24] 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).

[25] Gelman A. Exploratory data analysis for complex models. Journal of Computational and Graphical Statistics 2004;13:755–79. doi:[10.1198/106186004x11435](https://doi.org/10.1198/106186004x11435).

[26] Hoaglin DC, Mosteller F, Tukey JW, editors. Fundamentals of exploratory analysis of variance. John Wiley & Sons, Inc. 1991. doi:[10.1002/9780470316832](https://doi.org/10.1002/9780470316832).

[27] 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](https://doi.org/10.1002/sim.2768).

[28] Royston P, Sauerbrei W. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. Statistics in Medicine 2004;23:2509–25. doi:[10.1002/sim.1815](https://doi.org/10.1002/sim.1815).

[29] Lin X, Zhang D. Inference in generalized additive mixed modelsby using smoothing splines. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 1999;61:381–400. doi:[10.1111/1467-9868.00183](https://doi.org/10.1111/1467-9868.00183).

[30] Smith CT, Williamson PR, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. Statistics in Medicine 2005;24:1307–19. doi:[10.1002/sim.2050](https://doi.org/10.1002/sim.2050).

[31] Williamson PR, Clough HE, Hutton JL, Marson AG, Chadwick DW. Statistical issues in the assessment of the evidence for an interaction between factors in epilepsy trials. Statistics in Medicine 2002;21:2613–22. doi:[10.1002/sim.1044](https://doi.org/10.1002/sim.1044).

[32] Marson A, Williamson P, Clough H, Hutton J, Chadwick D, Monothera OBOTE. Carbamazepine versus valproate monotherapy for epilepsy: A meta-analysis. Epilepsia 2002;43:505–13. doi:[10.1046/j.1528-1157.2002.20801.x](https://doi.org/10.1046/j.1528-1157.2002.20801.x).

[33] 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).