A comparison of Fractional polynomials, meta-STEPP and smoothing splines to investigate effect modification in IPD-MA : a simulation study

Michail Belias

21 September, 2019

Abstract

Introduction

Individual participant data (IPD) meta-analysis (MA) offer great opportunities, such as to simultaneously investigate effect modification while also modelling for potential non-linear association between the outcome and the effect modifier. Several regression based approaches have been proposed such as fractional polynomials, meta-STEPP and splines.

Objective

Our objective is to compare their properties and their precision to investigate effect modification in IPD-MA.

Methods

To that goal we first describe the aforementioned methods. Subsequently, we conduct a simulation study covering 3 distinct scenarios. All scenarios consist of 5 studies with 200 participants each and equal treatment allocation reflecting the design of a randomised clinical trial. In the first scenario we introduced heterogeneity but the domain of the continuous effect modifier was the same across the studies. In the second scenario we introduced heterogeneity and the the domain of the continuous effect modifier was the different across the studies, while in the third scenario we added ecological bias.

Results

Conclusions

Introduction

Individual participant data (IPD) meta-analysis (MA) is well established as the gold standard to synthesize evidence from multiple studies [1,2]. IPD-MA offers great opportunities and showed great increase over the last decades [3]. Two opportunities that can be addressed in IPD-MA is the investigation of effect modification and modelling non-linear associations with the outcome. Estimating the real underlying functional shape between the outcome and a continuous variable is an important task, as failing to do so may lead to biased results and subsequently biased conclusions. Therefore, a plethora of modelling approaches have been proposed such as fractional polynomials, splines and sliding window techniques. Most of these approaches have been initially developed in a single studies and extended in multi-level schemes to account for the within study clustering of the participants. The aforementioned approaches can be classified in two categories. Fractional polynomials is a **global** method, while splines and sliding windows approaches are **fragmented** methods.

One of the first attempts to extent linear models to account non-linearities was polynomial regression approach. Hereby, powers of the continuous variable higher than one (linear) are generated and regressed with the outcome. Different degree polynomials are typically compared using tests such as likelihood ratio and Wald test or criteria such as AIC and BIC. Fractional polynomials [4,5] use the same approach as polynomial regression, but the powers that can be used come from a specific set $S : [-2; -1; -0 : 5; 0; 0 : 5; 1; 2; 3]$. Note that FPs introduce also negative powers, which offers great modeling capability. In our manuscript, we will not adress polynomial regression, as we believe that they are covered by fractional polynomials due to their similarity.

Splines are based on piece-wise polynomials. Although, a wide variety of methods has has been developed splines are generally considered semi-parametric. Nevertheless, some approaches tend to be more parametrica than others. As the term piece-wise implies the independent variables are splitted into sub-domains. The cut points for these splits are often called knots and the domains within each knot intervals. Note that simply fitting a polynomial within two adjacent interval, would produce discontinuous fitted lines. Therefore, splines are constrained to be continuous and in most approaches smooth over the knots. Splines can be classified in two categories **regression splines** and **penalised splines**. The most commonly used **regression splines** are restricted splines [cit] and B-splines [6,7]. On the other hand, the most commonly used **penalised splines** are smoothing splines [8] and p-splines [9].

Sliding window - often called moving average or moving regression- are non-parametric approaches, as they make no distributional assumptions. As the term sliding window implies, hereby, we focus in small subsets of data starting from the lowest value of the continuous variable and gradually reaching the highest. These small subsets are often called windows and the distance between the two consequent windows is called sliding step. The most common of sliding window approaches are Locally Weighted Scatterplot Smoothing (lowess) and Subpopulation Treatment Effect Pattern Plot (STEPP)[10].

All aforementioned modelling approaches have been also extended to meta-analysis. In general IPD-MA may be applied either in one-stage or two-stage. In two-stage IPD-MA, we may apply per study a statistical model of choice and then extract either 1) the estimated coefficients along with their standard errors or 2) their fitted lines along with their 95% confidence bands. Subsequently, we pool them with multivariate meta-analysis [11] or point-wise meta-analysis [12]. Apparently, methods that don't provide coefficients can not use this multi-variate meta-analysis. Furthermore, another limitation for this approach is that it may be applied only whenever the domain of the continuous variable is approximately common between studies. On the other hand, point-wise meta-analysis can be applied in both cases. Nevertheless, in scenarios where the domain of the independent variable is different point-wise meta-analysis may provide jagged fitted lines and confidence intervals. On the other hand, in one-stage meta-analysis a typically a mixed-effects model is applied to account for within study clustering of the participants. This way we assume that the coefficients and/or the fitted lines come from a higher distribution and they are different across studies due to randomness.

Little is known about the properties of the aforementioned techniques. To address that we perform a simulation study with three types of response and in three different scenarios to reflect multi-level structures that are typically encountered in practise. As the performance of each method is highly dependent on the simulated functional shape we do not expect to find a single technique that outperforms all other in every case. Instead, our goal is to provide guidance on choosing the approach that fits best a given occasion. Our manuscript is organised as follows. In section 2 we describe the simulated scenarios and explain the simulation design and the performance measures we used. In section 3, we describe the aforementioned approaches and in section 4 we provide their results. Section 5 contains discussion and our conclusions.

2. Simulation

2.1

Perfomance measurement

We calculated the mean prediction error of our approaches and

3. Methods

4. Results

5. Discussion and conclusions

References

[1] Riley RD. Commentary: Like it and lump it? Meta-analysis using individual participant data. International Journal of Epidemiology 2010;39:1359–61. doi:10.1093/ije/dyq129.

[2] Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: Rationale, conduct, and reporting. BMJ 2010;340:c221–1. doi:10.1136/bmj.c221.

[3] Simmonds M, Stewart G, Stewart L. A decade of individual participant data meta-analyses: A review of current practice. Contemporary Clinical Trials 2015;45:76–83. doi:10.1016/j.cct.2015.06.012.

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

[5] 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.

[6] Boor C de. A practical guide to splines. Springer New York; 1978. doi:10.1007/978-1-4612-6333-3.

[7] Curve and surface fitting with splines. Choice Reviews Online 1993;31:31–2162–31–2162. doi:10.5860/choice.31-2162.

[8] Best DJ, Green PJ, Silverman BW. Nonparametric regression and generalized linear models: A roughness penalty approach. Biometrics 1994;50:1228. doi:10.2307/2533467.

[9] Eilers PHC, Marx BD. Flexible smoothing with b -splines and penalties. Statistical Science 1996;11:89–121. doi:10.1214/ss/1038425655.

[10] Yip W-K, Bonetti M, Cole BF, Barcella W, Wang XV, Lazar A, et al. Subpopulation treatment effect pattern plot (STEPP) analysis for continuous, binary, and count outcomes. Clinical Trials: Journal of the Society for Clinical Trials 2016;13:382–90. doi:10.1177/1740774516643297.

[11] Berkey CS, Anderson JJ, Hoaglin DC. Multiple-outcome meta-analysis of clinical trials. Tatistics in Medicine 1996;15:537–57. doi:10.1002/(SICI)1097-0258(19960315)15:5<537::AID-SIM176>3.0.CO;2-S.

[12] 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.