

RESEARCH ARTICLE

Open Access

Bayesian splines versus fractional polynomials in network meta-analysis



Andreas Heinecke¹, Marta Tallarita² and Maria De Iorio^{1,2*} 

Abstract

Background: Network meta-analysis (NMA) provides a powerful tool for the simultaneous evaluation of multiple treatments by combining evidence from different studies, allowing for direct and indirect comparisons between treatments. In recent years, NMA is becoming increasingly popular in the medical literature and underlying statistical methodologies are evolving both in the frequentist and Bayesian framework. Traditional NMA models are often based on the comparison of two treatment arms per study. These individual studies may measure outcomes at multiple time points that are not necessarily homogeneous across studies.

Methods: In this article we present a Bayesian model based on B-splines for the simultaneous analysis of outcomes across time points, that allows for indirect comparison of treatments across different longitudinal studies.

Results: We illustrate the proposed approach in simulations as well as on real data examples available in the literature and compare it with a model based on P-splines and one based on fractional polynomials, showing that our approach is flexible and overcomes the limitations of the latter.

Conclusions: The proposed approach is computationally efficient and able to accommodate a large class of temporal treatment effect patterns, allowing for direct and indirect comparisons of widely varying shapes of longitudinal profiles.

Keywords: Bayesian evidence synthesis techniques, P-splines, Clinical trials, Evidence-synthesis, Longitudinal studies, Markov chain Monte Carlo methods, Mixed treatment comparison

Background

Scientific and technological advances are steadily adding to the number of different healthcare interventions. To fully exploit their potential requires clinicians and healthcare professionals to make informed and objective choices, based on clinical studies, between a possibly large number of treatment options in terms of relative medical efficacy and cost effectiveness [11, 22]. It is generally accepted that randomized controlled trials provide the most rigorous and conclusive evidence on the relative effects of different interventions. For example, the gold standard for directly comparing two treatments *A* and *B* is a randomized controlled trial. In practice, however,

evidence from direct comparison trials may be limited and it is often impossible to have head-to-head comparisons for all relevant comparators of an intervention, making it necessary to resort to indirect comparisons [11]. For instance, direct comparison from two different studies on treatment *A* versus *C*, and *B* versus *C*, might be available and indirect methods exploit the common comparator *C* to provide an indirect comparison of treatment *A* versus *B*. A large number of individual studies can be mapped out as a network in which treatments are represented by nodes that are connected by edges where data from direct head-to-head studies comparing them are available. Network meta-analysis (NMA) refers to methods attempting to systematically integrate all the information provided by such networks of studies via the entirety of paths between different treatments. The goal is to provide a comparison of each treatment against a

*Correspondence: m.deiorio@ucl.ac.uk

¹Yale-NUS College, 16 College Avenue West, 138527 Singapore, Singapore

²Department of Statistical Science, University College London, Gower Street, WC1E 6BT London, UK



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

common comparator chosen from the network, such as a placebo or standard treatment, which consequently allows for a relative comparison among all available treatments [33, 40–43]. Several statistical methods are available in the literature to effectively integrate findings from individual studies and implement NMA in the context of systematic reviews. Current approaches include the adjusted indirect comparison method with aggregate data, meta-regression, mixed effect models, hierarchical models and Bayesian methods [43–45], and have the potential to yield useful information for clinical decision-making. The main methodological idea beyond most statistical methods is to extend standard pairwise meta-analysis techniques to the entire set of trials and to include direct and indirect comparisons by exploiting the different paths defined on the network. In the recent literature the Bayesian approach has proved successful in addressing the major statistical challenges associated to the design of a systematic review. A summary of the most recurring challenges encountered in NMA is presented in [41]: the total number of trials in a network, the number of trials with more than two comparison arms (introducing an additional layer of complexity) [1], heterogeneity (i.e. clinical, methodological, and statistical variability within direct and indirect comparisons, and across studies), inconsistency (i.e. discrepancy between direct and indirect comparisons) [2], and potential bias that may influence effect estimates [3].

In this work, we focus on NMA techniques for longitudinal data, i.e. on the NMA of studies in which individuals are assessed at multiple time points throughout the follow-up period. Among the additional difficulties this setup brings about are that follow-up times may vary across studies, and, moreover, that the repeated measurements for each individual during the follow-up tend to be correlated. The latter needs to be taken into account to avoid biased estimates of treatment effects. For instance, in [4] the authors compare alternative approaches to handling correlation in linear mixed effect models for longitudinal NMA. In a Bayesian framework [5] proposes a model based on piecewise exponential hazards, while in [6] the treatment effect in multi-arm trials is modelled with a piecewise linear function. Our starting point is the review paper [7], which compares three recently popular NMA methods for longitudinal data, whose main inferential focus is treatment effect over time and which allow for a large class of temporal patterns. The three methods are (i) the mixed treatment comparison (MTC) developed in [8], which assumes random relative effects and allows for the relative treatment effects to vary over time, without temporal pattern restriction, (ii) the Bayesian evidence synthesis techniques – integrated two-component prediction (BEST-ITP) proposed in [9], a parametric model to describe a non-linear relationship between outcome and time for each treatment, assuming a diminishing return

time course of treatment responses, and (iii) the more recent method based on fractional polynomial (FP) temporal patterns of [10], which we will review in more detail below. Each of the three models is formulated under a Bayesian hierarchical modelling approach, which provides a natural tool to combine different (and possibly not completely homogeneous) sources of evidence by decomposing a complex problem in subcomponents which are modelled individually and then linked together via hyperparameters in a probabilistically sound way. Moreover, using a Bayesian approach, it is in principle straightforward to incorporate decision making, for example allowing quantification of the impact of model- and parameter-uncertainty on the optimal decision, given current data. This process is required by many regulatory agencies (see [12]) and would simply translate in adding extra layers to the model hierarchy (see [13, 14]).

The authors of [7] conclude, based on a simulation study and real data applications, that the MTC method appears to be most conservative in the estimation of the underlying effect-size, that the BEST-ITP model fails to capture constant treatment effect over time, and that the FP model seems to offer the most flexible strategy, accommodating different time patterns. In general, it is challenging to capture non-monotonic temporal patterns, with the FP model leading to slightly more accurate estimates (as shown in their simulation study). In terms of computational cost, there are no substantial differences in running time between the three different models, although fractional polynomials require extensive sensitivity analysis to select the optimal order and power terms, as we will discuss below.

Although the FP model turns out as the best of the three reviewed approaches, it still presents limitations both from a methodological and a computational point of view (see Sections on the FP method and “[Real data application](#)” section below). In this work, we propose an NMA method based on B-splines to model temporal behaviour, which allows for the simultaneous analysis of outcomes at different time points, automatically accounting for correlation across time. We illustrate the model in simulations and on real data examples and compare its performance to the FP model. We also consider a variation of our basic model based on P-splines, which provides substantially identical inferential results. The proposed approach has many advantages in terms of model flexibility, computational burden and ease of specification. B-splines (and their extensions such as P-splines) are a natural competitor of FPs and as such they have been previously compared in the literature in different setups (see, among others, [15]).

The article is organized as follows. In the “[Method](#)” section, we describe the model likelihood for the general NMA problem. We review the FP approach, introduce the proposed B-spline model and highlight its advantages.

Further, we specify the prior distribution for the parameters of the model and introduce the P-spline alternative. We then briefly describe the MCMC strategy. In the “Results” and “Discussion” sections we compare the B-spline, P-spline and FP models in simulations and real data applications. In Additional file 1, we describe in detail additional simulation scenarios and their results. We provide JAGS code to fit the proposed approach in Additional file 3.

Method

Network meta-analysis

The main objective of NMA of longitudinal studies is the evaluation of a suitably defined response over time. The response y_{sjt} in study $s \in \{1, \dots, S\}$ and treatment arm $j \in \{1, \dots, J\}$ at time $t \in [0, T]$ is distributed as

$$y_{sjt} \sim f\left(y_{sjt} \mid \theta_{sjt}, \sigma_{sjt}^2\right), \tag{1}$$

where f is a (usually parametric) probability density. The main parameter of clinical interest is θ_{sjt} , measuring the treatment effect of intervention j in study s at time t . As in the generalized linear model framework, a link function $g(\theta_{sjt})$ specifies the relationship between the main clinical effects θ_{sjt} and the response. For example, in the case of continuous outcomes we consider

$$y_{sjt} \sim \text{Normal}\left(\theta_{sjt}, \sigma_{sjt}^2\right), \tag{2}$$

with $g(\theta_{sjt}) = \theta_{sjt}$, while in the case of binary outcomes

$$y_{sjt} \sim \text{Binomial}\left(n_{sjt}, p_{sjt}\right), \tag{3}$$

$$g(p_{sjt}) = \text{logit}(p_{sjt}) = \theta_{sjt}.$$

Other alternatives include $g(\theta_{sjt}) = \ln(\theta_{sjt})$ if f is modelled as a Poisson distribution. In this work, we focus on continuous and binary responses as these are the most common outcomes in clinical trials.

The fundamental difference between the NMA methods compared in the review paper [7] is the way the treatment effect θ_{sjt} is modelled. The review shows that among the considered models FP methods are the most flexible and can accommodate a variety of treatment effect patterns. On the other hand, they require intensive computations for the choice of parameters and present modelling drawbacks, as discussed in the next Section.

To capture the longitudinal data, in this work we focus on basis function models which do not require a priori assumptions on the temporal pattern of the treatment effect. We model the longitudinal curve as a linear combination of basis functions

$$\theta_{sjt} = \sum_{k=0}^M \beta_{ksj} h_k(t),$$

where h_k denotes a basis function, β_{ksj} is the corresponding coefficient for study s and treatment j , and $M + 1$

is the total number of basis functions. In particular, we consider spline and fractional polynomial basis functions, carefully discussing the choice of the type and number of basis functions in the following sections. They are key features of the model as they influence the level of complexity and class of temporal profiles that can be accommodated.

Following [10], we set the variances of the main outcomes in (2) equal to

$$\sigma_{sjt}^2 = \frac{1}{(1 - \rho^2)} \left(\frac{\text{sd}_{sjt}}{\sqrt{n_{sjt}}} \right)^2, \tag{4}$$

i.e. to the standard error (with sd_{sjt} an estimate of the standard deviation and n_{sjt} the sample size) for the corresponding study, treatment arm and time point, adjusted by a factor $\rho \in [0, 1)$ taking into account the within-study correlation between subsequent time points. While estimates for the correlation coefficient, here assumed constant over time, may be available from expert knowledge, in a fully Bayesian setting it is an object of inference and assigned an appropriate prior distribution. In this study we assume $\rho \sim \text{Uniform}(0, 0.95)$.

Fractional polynomial model

In the FP regression framework, powers of the covariates of interest (in our case time), usually chosen from the set $S = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$, are entered into the linear predictor (see [16] for a detailed account). The authors of [10] propose an FP approach to NMA in which, given an order M_F and powers $p_1, \dots, p_{M_F} \in S$, the mean outcome of the j -th treatment in study s at time $t > 0$ is modelled as

$$\theta_{sjt} = \begin{cases} \beta_{0sj} + \sum_{m=1}^{M_F} \beta_{msj} t^{p_m} & \text{if } p_1 \neq \dots \neq p_{M_F}, \\ \beta_{0sj} + \beta_{1sj} t^p + \sum_{m=2}^{M_F} \beta_{msj} t^p [\ln(t)]^{m-1} & \text{if } p_1 = \dots = p_{M_F} = p, \\ & \text{and } M_F > 1, \end{cases} \tag{5}$$

where $t^0 := \ln(t)$. Implementing the FP model requires selecting an order M_F and powers p_1, \dots, p_{M_F} . In both the original work [10] and the simulation study [7] the Deviance Information Criterion [DIC, [17]] is used as an expected predictive error estimate to guide the selection of the FP order and powers. When using DIC, the aim is to select order and powers that minimize the sum of a goodness of fit term, given by the posterior mean of the model deviance, and a regularization term, given by the posterior estimate of the effective number of parameters. The latter is intended to penalize model complexity in order for the DIC to be a tool that prevents overfitting.

First and second order FPs allow representations of a considerable range of non-linear relationships, and higher orders are rarely used in medical applications (see, for instance, [18, 19]). Even considering only the cases $M_F = 1$ and 2, using the DIC often presents a non-negligible

computational task as it requires the comparison of the DIC values of eight different first-order fractional polynomial models (one for each possible power from S) and of 64 different second-order fractional polynomial models (for all the possible combinations of powers from S). Besides the computational cost necessary to select the order and powers, the FP modelling approach lends itself to more fundamental criticism: (i) any choice of order and combination of powers imposes strong structural properties on the curve and therefore on the set of representable temporal patterns, which may not be supported by the data. Moreover, the selection of powers and order depends also on the prior specification of the other parameters in the model; (ii) FPs generally have singularities at zero or grow polynomially for large arguments. Consequently, more complex FPs may only provide poor fit for values of t that are large or close to zero; (iii) the individual terms in an FP are functions with support over the entire time interval and, as discussed in [16], may be less responsive locally to perturbations of the response at a given point, while the fit at distant points may be affected considerably. To avoid such effects, basis functions with short support may be more appropriate in many situations [see the discussion in [19]]; (iv) first-order FPs describe effects as a function of transformed time t in a linear model $\beta_{0sj} + \beta_{1sj}t^p$, in particular they are always monotonic. Higher order FPs become increasingly complex, with second-order FPs being either monotonic or unimodal. It may however be preferable to have a model in which the linear model is always subsumed in more complex models; (v) although a second order FP as in (5) has three coefficients only (two in the case of modelling change from baseline, where the intercept is equal to zero and can be discarded), the choice of powers increases model complexity, since many different FP models need to be fitted to find the one that best describes the temporal pattern of the treatment effect. Additionally, the use of DIC as model choice criterion effectively hinges the FP model to any potential shortcomings of the DIC which have been pointed out in the literature [e.g. [20, 21]]. We have indeed observed sub-optimal DIC-based selections in the real data applications, and therefore present a more detailed discussion in the “Real data application” section. To address the above issues we next propose to use B-spline basis functions to capture the treatment effect over time.

B-spline model

B-splines are a family of basis functions with many desirable theoretical and computational properties, making them widely used in function approximation and countless applications in statistics and engineering. For details we refer to [23]. Univariate (cardinal) B-splines can be defined inductively, starting from the B-spline of order 1 given by

$$B_1(t) := \begin{cases} 1 & \text{if } t \in [0, 1], \\ 0 & \text{otherwise.} \end{cases}$$

B-splines of higher order $n \in \mathbb{N}$ are defined via consecutive convolutions by

$$B_n(t) := \int_{\mathbb{R}} B_{n-1}(t-s)B_1(t) ds,$$

i.e. with higher order they become increasingly smoother. For $n \geq 2$, the B-spline B_n is

- (i) symmetric and positive valued, with finite support $[0, n]$,
- (ii) $(n - 2)$ -times continuously differentiable,
- (iii) polynomial of degree $n - 1$ when restricted to any interval $[k, k + 1]$, $k \in \mathbb{Z}$, and
- (iv) satisfies $\sum_{k \in \mathbb{Z}} B_n(t - k) = 1$ for all $t \in \mathbb{R}$.

The integer translates $\{B_n(\cdot - k)\}_{k \in \mathbb{Z}}$ are thus a set of highly regular and well structured basis functions. They span the space of functions that are polynomial on each interval $[k, k + 1]$, $k \in \mathbb{Z}$, and $(n - 2)$ -times continuously differentiable in every $k \in \mathbb{Z}$. Such piecewise polynomial functions with global smoothness restrictions are called splines. In the cardinal case the integers are called the knots and the above can be easily generalized to any uniform knot sequence $(hk)_{k \in \mathbb{Z}}$, where $h \in \mathbb{R}$, by considering the basis functions $\{B_n(t/h - k)\}_{k \in \mathbb{Z}}$ of translates of the appropriately dilated cardinal B-splines. In applications in which a finite time interval is considered, only those finitely many basis functions whose support intersects the interval are required.

A main advantage of B-spline basis functions comes from their locality due to their finite support and their structure as translates of only one generating symmetric function. Choosing an order and adjacent knot distance does not impose any overall temporal behavior beyond smoothness. In fact, for sufficiently large $m \in \mathbb{N}$, the functions $\{B_n(2^m t - k)\}_{k \in \mathbb{Z}}$ can approximate any integrable function to arbitrary precision. This has to be contrasted with the FP model, in which the fractional monomials t^{p_m} may be considered as basis functions, and where choosing any combination of fractional monomials imposes rather strong geometrical restrictions on the representable functions. Furthermore, due to the translation structure of the B-spline bases, all parts of the observation time-interval are treated equally, with no performance deterioration for small or large time arguments. The local support of the basis functions makes the B-spline expansions a local-influence model (i.e. perturbations of the response only affect the fit of the model locally), whereas their overlapping support, the size of which goes hand in hand with the order of smoothness, acts as a regularizing mechanism that can facilitate stability [see [16]]. B-spline bases are also stable in the sense that small changes in coefficients

do not perturb the represented function significantly [see [24]]. Small changes in a represented function will therefore result in small changes in its coefficient sequence and vice versa. Together with the locality, and the partition of unity property (iv), these characteristics contribute to a better interpretability of the coefficients in a B-spline expansion as compared to the FP coefficients. Indeed, it is easy to understand changes in the function due to changes in the coefficients, as every coefficient corresponds to a particular interval of the domain of the function. On the other hand, FPs are sums of different global functions and it is therefore harder to visualize the effect of changes in coefficients. Finally, B-spline bases have the property that, with increasing order and additional knots, the spaces of representable functions become strictly larger, containing all previously representable function. This nestedness is desirable since it ensures that simpler models are always embedded in more complex, a property not shared by the FP model.

Practical applications of B-splines are dominated by $n = 4$ as the order of choice, resulting in piecewise cubic polynomials with continuous curvature. In fact, given a number of data points, it is well known [e.g. [25]] that among all smooth regression functions, the unique solution to minimizing a weighted sum of squared approximation errors and the integral squared curvature as regularization is given by a cubic spline with knots at the data time-points. As such, in the remainder of this paper, we will restrict our attention to cubic B-splines.

The smoothness property of the spline functions at their knots makes them relatively insensitive to the precise locations of the knots. While observation times differ across studies, the B-spline basis functions, and in particular their knot positions, have to coincide across studies in order to guarantee the consistency assumption of the NMA in this modelling framework (see next Section). Since studies in NMA of longitudinal data typically have very few observation time points in the interval $[0, T]$, we choose $h = 3/T$, i.e. we consider the four uniformly spaced knots $0, T/3, 2T/3, T$, over the observation interval $[0, T]$. We conducted a sensitivity analysis for the number of knots ranging from two to ten and generally four knots provided the best choice (results not shown).

For cubic B-splines with four knots over the observation time interval, six of the generator translates have support intersecting $[0, T]$. We denote those by $\{B_k\}_{k=0}^{M_B}$ with $M_B = 5$ and model the mean outcome of treatment j in study s at time $t > 0$ as cubic spline using the basis expansion

$$\theta_{sjt} = \sum_{k=0}^{M_B} \beta_{ksj} B_k(t). \tag{6}$$

Despite having a greater number of coefficients, the B-spline model provides a conceptually simpler and computationally more attractive strategy than the FP model (5) since all order and power parameter choices (for instance evaluated via the DIC) have to be additionally considered for (5), whereas we keep the model (6), i.e. the spline order and number of knots, fixed in all applications. The localized support of B-splines allows us to propose a default model choice, with good performance in most applications. This strategy is not possible for FPs, as they lack this property and impose global assumptions on the behaviour of representable functions.

Prior specification

We follow the prior specification proposed for the FP approach in [10], which extends earlier approaches of [8] and [9]. The regression coefficient vector $\beta_{sj} = (\beta_{0sj}, \dots, \beta_{M_B sj})^T$ for both the FP model (5), with $M = M_F$, and the B-spline model (6), with $M = M_B$, is expressed as a sum of a study-specific random effect and a study-specific arm deviation from the reference treatment:

$$\beta_{sj} = \mu_s + \delta_{sj}$$

where the study specific means $\mu_s = (\mu_{0s}, \dots, \mu_{M_B s})^T$ in our implementation are assigned a vague prior distribution

$$\mu_s \sim \text{Normal}(\mathbf{0}, 10^4 \mathbf{I}),$$

with $\mathbf{0}$ denoting the zero vector and \mathbf{I} the identity matrix of appropriate dimensions. The study specific treatment effects $\delta_{sj} = (\delta_{0sj}, \dots, \delta_{M_B sj})^T$ in study s of treatment j relative to a reference treatment, indexed as $j = 1$, are modelled as “structured random effects”

$$\delta_{sj} \sim \text{Normal}(\mathbf{d}_j - \mathbf{d}_1, \Sigma), \tag{7}$$

$$\mathbf{d}_j \sim \text{Normal}(\mathbf{0}, 10^4 \mathbf{I}), \tag{8}$$

for $j > 1$, with $\mathbf{d}_1 = (d_{01}, \dots, d_{M_B 1})^T := \mathbf{0}$ and $\delta_{s1} = (\delta_{0s1}, \dots, \delta_{M_B s1})^T := \mathbf{0}$. Clinical effects in (5) and (6) are therefore expressed as change from baseline (CFB) with respect to the reference treatment. The covariance matrix $\Sigma = (\sigma_m \sigma_n \lambda_{mn})_{m,n=0,\dots,M}$ captures heterogeneity between studies, where σ_m^2 (often referred to as the heterogeneity parameter) represents the variance in $(\delta_{msj})_{sj}$ and captures how much variation exists between the results of different studies. We assume σ_m to be constant for all treatment comparisons. This assumption also implies that the between-study variance over effect estimates remain constant over time [see [26]]. The covariances λ_{mn} quantify the correlation between these treatment effect parameters.

When assuming a (partially) fixed-effects model, (7) (or certain components of it) is replaced by $\delta_{sj} = \mathbf{d}_j - \mathbf{d}_1$ (i.e.

all or some σ_m are zero) and it is not necessary to estimate (the respective) between-study covariances. However, in the B-spline case of this work we only consider models including random effects μ_s per study, allowing for heterogeneity in the regression coefficients between studies, but fixed effects for δ_{sj} . For the FP model, as suggested in [10], we impose a random effect only on one of the components δ_{sj} . In particular, when $M_F = 1$, there is only one between-study heterogeneity parameter, related to the relative treatment effects for β_{1sj} . In this case we have only one heterogeneity parameter σ_m to which we assign a uniform distribution on the interval (0, 10). For all real data applications we also consider a uniform distribution on (0, 5) and (0, 50) for σ_m . Posterior inference results are essentially identical to the ones obtained for the interval (0, 10), showing that inference is robust to prior specification for the between study heterogeneity. Moreover, if $M_F = 2$, we consider also the case in which the between-study heterogeneity concerns treatment effects in terms of β_{2sj} .

Note that for the Bayesian NMA model described here, i.e. for (7), to facilitate consistent comparisons of treatments that are not directly compared by the same study with those that are, the implicit assumption regarding the relative treatment effects is that $d_{j_1} - d_{j_2} = (d_{j_1} - d_1) - (d_{j_2} - d_1)$ [10]. Inconsistency can occur if there are systematic differences in relative treatment effect modifiers between different direct comparisons. For the consistency assumption to transfer through (6) and (5), the same B-splines (i.e. order and knots) and the same FPs (i.e. order and powers) have to be used across all studies and treatment arms. Finally, note that the described model does not account for correlations stemming from trials with more than two treatment arms but can be easily extended to consider them [see [26]].

Bayesian p-splines

A possible variation to the B-spline model that arises from a different prior specification for (8) is related to the penalized spline (P-spline) least squares curve fitting approach [27–29], in which B-splines with a relatively large number of equally spaced knots are used to fit a function of desired degree of smoothness by penalizing the curvature. The frequently used regularization introduced in [27] penalizes the sum of second-order differences between consecutive coefficients in the B-spline expansion, and can be adapted to incorporate other potentially available information about the shape and degree of smoothness of the target function. Most importantly, in the P-spline approach the number and location of the knots are not crucial: A relatively large number of uniformly spaced knots (large enough to ensure sufficient flexibility thus preventing oversmoothing) can be fixed a priori since the penalty term prevents overfitting. For a more

thorough discussion on the P-spline penalized least squares approach see [30].

We implement a Bayesian version of the P-spline approach that has been introduced in [31]. The specification of appropriate prior distributions in (9) and (10) introduces locally adaptive smoothing parameters. As before, we assume model (6) for the treatment effects, however, the prior on \mathbf{d}_j in (8) is now replaced by a second order random walk. This provides a Bayesian counterpart to the second order penalty in [27]. Precisely, the prior specification on the \mathbf{d}_j are as follows: we set $\mathbf{d}_1 = \mathbf{0}$, and for $j > 1$ and $k \geq 0$ we assume, following [31],

$$d_{kj} \sim \text{Normal}(2d_{k-1,j} - d_{k-2,j}, \eta_j) \quad (9)$$

$$\eta_j \sim \text{InverseGamma}(1, 0.0005) \quad (10)$$

In case $d_{k-1,j}$ or $d_{k-2,j}$ are not available, we let $d_{kj} \sim \text{Normal}(0, 1000)$.

MCMC algorithm

Posterior inference for all simulated examples and real data applications is performed via standard Gibbs sampling, implemented in JAGS [see [32]]. The first 5000 iterations are discarded as ‘burn-in’ and the final sample size on which inference is based is 10000 samples. Convergence of the chain is checked through the Gelman-Rubin potential scale reduction factor [see [34]]. The Gelman-Rubin diagnostic is evaluated by running multiple chains from different initial values and comparing the estimated between-chains and within-chain variances for each model parameter. Large differences between these variances would indicate that convergence has not been reached yet. MCMC convergence is not a major issue in this model since most prior distributions are conjugate and therefore the chain mixes well. Moreover, we are working with aggregate data that in general exhibits less variability than individual level data. In Additional file 2, we show the traceplots of μ_s and \mathbf{d}_s for one of the replicas of the non-monotonic scenario for the B-spline model. Indeed, good mixing of the chain is also evident from Table 1 in which we report summary statistics of some convergence diagnostics for the non-monotonic scenario with continuous and binary responses. We note that all diagnostics are satisfactory, with the Gelman-Rubin diagnostic close to one, the Effective Sample Size close to 10000 and small MCMC standard error. JAGS code to fit the models is provided in Additional file 3.

Results

Simulation study

We have replicated the simulation study of [7] to compare the B-spline and P-spline model with the FP approach.

Table 1 Median, minimum and maximum evaluated over the 50 replicates of the average across all the parameters of the convergence diagnostics: Gelman-Rubin diagnostic, effective sample size (ESS) and Monte Carlo standard error (SE). Results are reported for our model applied to the non-monotonic scenario with binary and continuous responses

	Non-monotonic			Non-monotonic (binary)		
	Median	Minimum	Maximum	Median	Minimum	Maximum
Gelman-Rubin	1.005	1.002	1.008	1.008	1.003	1.010
ESS	9996.289	9993.762	9998.571	9114.788	9076.123	9480.754
Monte Carlo SE	0.015	0.009	0.023	0.017	0.012	0.019

In [7] various simulation scenarios are designed for continuous and binary responses that involve widely differing temporal patterns of treatment effects. Moreover, the authors consider data sets simulated from the models of [8] and [9], as well as simulated data from non-closed networks. In this context, a network is called closed if for all possible pairs of treatments there is at least one study providing direct comparisons. A detailed description of the simulation study setup and our results is contained in Additional file 1. Moreover, in Additional file 4 we provide R code to simulate data from all the considered scenarios. Here we only discuss one scenario in detail, which highlights the advantages of our proposed approach.

For each scenario of the simulation study, we fit the B-spline, P-spline and FP models to 50 randomly generated data sets and report results as averages. The data sets are generated as follows. Given a network structure with studies s , treatments j and observation times t , individual level observations for continuous outcomes are modelled by $Z_{isjt} \sim \text{Normal}(\theta_{sjt}, \tau_s^2)$, for $i = 1, \dots, n_s$, where n_s is the number of observations in study s . The mean outcomes $\theta_{sjt} = \gamma_{jt} + \alpha_s$ are modelled as a sum of treatment effects γ_{jt} over time and independent study effects $\alpha_s \sim \text{Normal}(0, 10)$. Since our main interest is to investigate the performance of the models in describing the temporal patterns of the main outcomes, we will assume the variances τ_s^2 of the study outcomes as constant over time and across treatment arms. (This assumption may not hold in practice; see the real data applications below.) The main outcomes of each study, treatment and time point are reported as the sample means of the respective $(Z_{isjt})_{i=1, \dots, n_s}$, i.e. $Y_{sjt} = (\sum_{i=1}^{n_s} Z_{isjt}) / n_s$ or

$$Y_{sjt} \sim \text{Normal}(\gamma_{jt} + \alpha_s, \tau_s^2 / n_s).$$

Binary outcomes are simulated by taking the inverse logit transformation of θ_{sjt} and then generating individual level outcomes from a Bernoulli distribution (see (3)).

Different scenarios are created through various specifications of the treatment effect curves γ_{jt} . Here we consider a closed network and simulate data from three hypothetical studies with two treatment arms each. See Fig. 1 (left panel) for a graphical representation. Study 1 directly compares treatments A and B with follow-up at four time

points (weeks 4,8,12 and 24). Study 2 compares treatments A and C at three time points (weeks 4,12,24). Finally, study 3 compares treatments B and C at three time points (weeks 4, 8, 12). The number of subjects per treatment arm in the different studies are chosen as $n_1 = 100, n_2 = 120$ and $n_3 = 130$. Variances are set $\tau_1^2 = 1, \tau_2^2 = 2$ and $\tau_3^2 = 4$.

In Figs. 2 and 3 we present, for continuous and binary outcomes respectively, a comparison of the estimated profiles obtained using the B-spline model, the P-spline model and the FP model, along with the true values used to simulate the data, assuming as treatment effects over time the non-monotonic continuous piecewise linear curves

$$\gamma_{At} = \begin{cases} -\frac{3t}{4}, & t \in [0, 4] \\ -\frac{5t}{2} + 2, & t \in (4, 8] \\ -\frac{t}{2} - 4, & t \in (8, 12] \\ \frac{t}{2} - 16, & t \in (12, 24] \end{cases}, \quad \gamma_{Bt} = \begin{cases} -\frac{3t}{2}, & t \in [0, 4] \\ -t - 2, & t \in (4, 8] \\ t - 18, & t \in (8, 12] \\ \frac{t}{4} - 9, & t \in (12, 24] \end{cases},$$

$$\gamma_{Ct} = \begin{cases} -2t, & t \in [0, 4] \\ \frac{t}{4} - 9, & t \in (4, 12]. \\ \frac{t}{6} - 8, & t \in (12, 24] \end{cases} \tag{11}$$

From the Figures it is evident that the spline models are able to accurately estimate non-monotonic time effects in both the continuous and binary cases, which

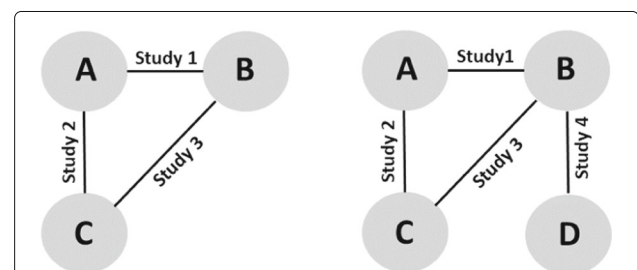


Fig. 1 Network of studies used in the simulation scenarios. Treatments are represented as nodes and connected by an edge if a direct comparison study between them is available. The left network is closed, with one study comparing all possible pairs of treatments, while in the right network it is extended to become non-closed, containing treatment pairs not directly compared by at least one study

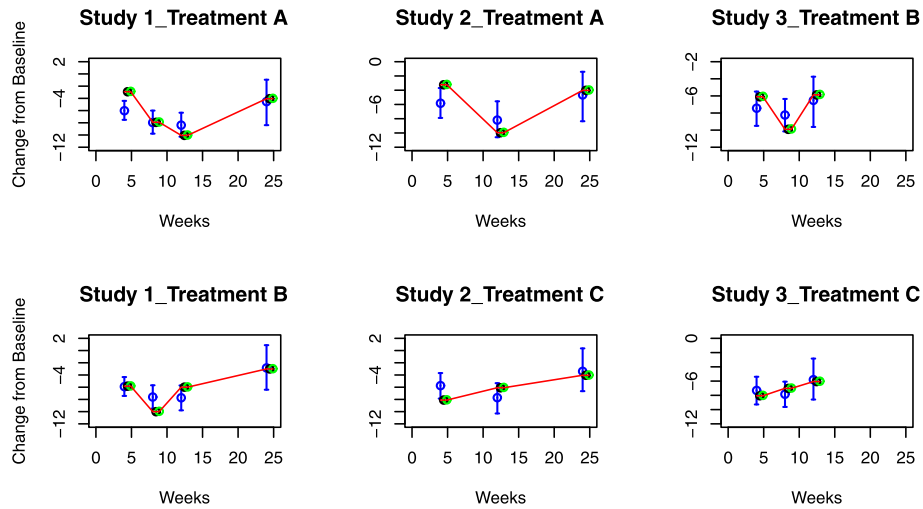


Fig. 2 Posterior estimated profiles of θ_{sjt} obtained under the non-monotonic scenario (11). Red lines indicate the values used to simulate the data. Estimates and 95% credible intervals obtained from the respective models are indicated in blue (FPs), black (B-splines) and green (P-splines). Due to their small size, credible intervals are not visible for B-splines and P-splines

the FP model fails to capture. The 95% credible intervals obtained with the splines are narrow and always cover the true value used to generate the data. In Fig. 4, we show estimated differences in treatment effects for one of the 50 simulation replicates in the continuous case. It is evident that the spline models are able to capture the non-monotonic curve shape for difference in treatment, while the FP model flattens out the difference in treatment with consequent increase of uncertainty. These conclusions are confirmed by the extensive simulation study presented in Additional file 1 for other temporal patterns. We report for an overall quantitative comparison of goodness of fit

the mean squared error (MSE) evaluated for each simulation scenario across all replicates. The MSE is defined as

$$MSE = \frac{1}{R} \sum_{r=1}^R \sum_{s,j,t} (Y_{sjt} - \hat{Y}_{sjt}^{(r)})^2$$

where $\hat{Y}_{sjt}^{(r)}$ denotes the predicted value of the response for study s , treatment j , time t and MCMC iteration r . The number of saved MCMC iterations is $R = 10000$. For all simulation scenarios the MSEs for the B-spline, P-spline and FP model are summarized in Table 2. The

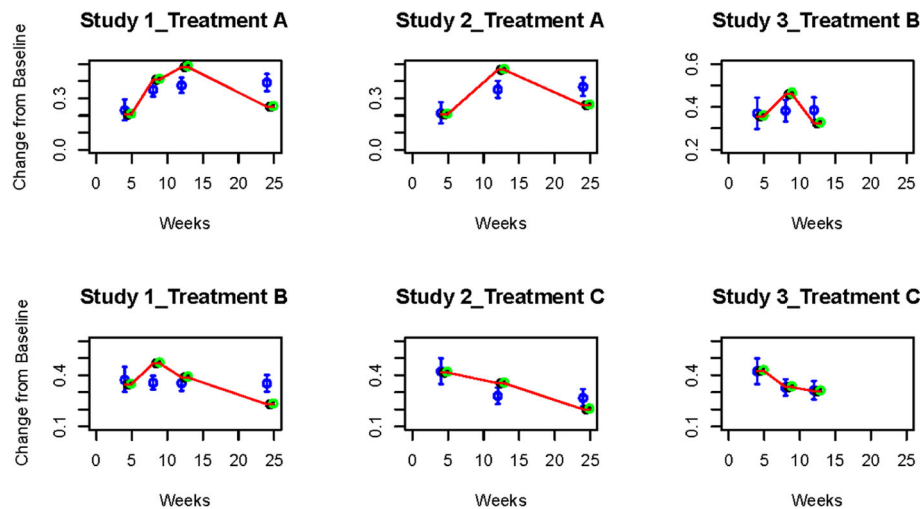


Fig. 3 Posterior estimated profiles of p_{sjt} for binary outcomes under the non-monotonic scenario described in (11). Red lines indicate the values used to simulate the data. Estimates and 95% credible intervals obtained from the respective models are indicated in blue (FPs), black (B-splines) and green (P-splines). Due to their small size, credible intervals are not visible for B-splines and P-splines

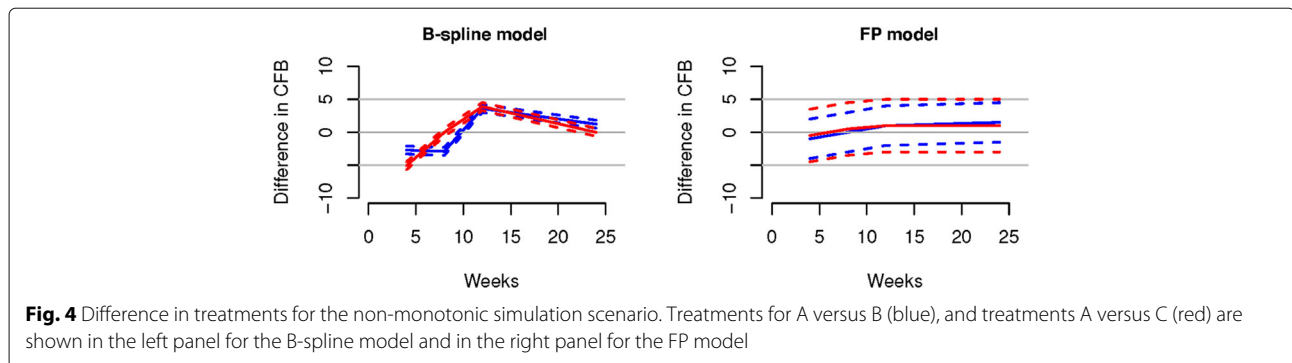


Fig. 4 Difference in treatments for the non-monotonic simulation scenario. Treatments for A versus B (blue), and treatments A versus C (red) are shown in the left panel for the B-spline model and in the right panel for the FP model

table reports the means (and standard deviations) of the MSEs across the 50 simulations, quantitatively showing an increase in estimation accuracy for the spline models as compared to the FP model in all scenarios, and particularly in the non-monotonic scenario simulated in (11). Further, in Table 3 we report summary statistics of the posterior estimates of σ_{ijt} obtained with the B-spline and FP model. The table confirms that the B-spline model performs better in terms of model fit and precision of the estimates. The estimates are in line with the values used to generate the data.

Real data application

We consider five real NMA data sets. The networks of studies for all data sets are provided in Figs. 5 and 6. Main features of all data sets are summarized in Table 4.

- (i) The first data set contains information from 13 studies on patients affected by chronic obstructive pulmonary disease (COPD) and is described in [35] and [7]. COPD is a long term lung condition affecting

adults worldwide, for which there is no cure but treatment can help slow down disease progression. Subjects receive one of three possible treatments: Acclidinium 400 μg BID (AB400), Tiotropium 18 μg QD (TIO18), or placebo. Three studies compare AB400 with placebo, nine compare TIO18 with placebo, and one compares AB400 and TIO18 with placebo. There is one closed loop, providing direct evidence. As all studies are placebo-controlled, placebo is used as the reference treatment.

- (ii) The second data set contains information from 16 studies on treatments for osteoarthritis (OA) of the knee and is described in [10] and [7]. OA is a painful chronic degenerative joint condition. The treatments in the systematic review are based on different hyaluronan (HA)-based viscosupplements. The different treatments considered are: three, four, or five injections of HA with a molecular weight (MW) of 0.5-0.73 million Da (Hyalgan) (3 Hy-0.5-0.73; 4 Hy-0.5-0.73; 5 Hy-0.5-0.73); three injections of HA MW of 0.62-11.7 million Da (Supartz) (3

Table 2 Mean Square Error (MSE) comparison for B-spline, P-spline and fractional polynomial (FP) network meta-analysis models. Displayed are the mean and standard deviation (SD) of the respective MSEs averaged over 50 simulated data sets for each scenario. Unless stated otherwise, considered outcomes are continuous. Scenarios that contain non-monotonic temporal behaviors appear to be challenging for the FP method

Scenario	B-spline model		P-spline model		FP model	
	Mean	SD	Mean	SD	Mean	SD
Linear	0.0019	0.0213	0.0021	0.0243	0.3505	0.0388
Logarithmic	0.0014	0.0111	0.0015	0.0135	0.3505	0.0118
Piecewise linear monotonic	0.0011	0.0923	0.0052	0.1437	1.3069	0.1823
Mixed	0.0037	0.0313	0.0040	0.0931	1.5747	0.1338
Non-monotonic	0.0019	0.0389	0.0027	0.1039	41.7533	1.8354
MTC	0.0374	0.0483	0.0402	0.0984	0.9198	0.2864
BEST-ITC	0.0397	0.0503	0.0402	0.1003	0.1701	0.2898
Piecewise linear monotonic (binary)	0.0034	0.0118	0.0053	0.0978	0.0897	0.3698
Non-monotonic (binary)	0.0004	0.0019	0.0004	0.0178	0.9748	0.2976
Piecewise linear (non-closed network)	0.0172	0.0173	0.0235	0.0774	0.0903	0.3854

Table 3 Comparison between estimates of σ_{sijt} (p_{sijt} for binary outcomes) obtained with the B-spline, P-spline and Fractional Polynomial (FP) network meta-analysis models. We report median, minimum and maximum evaluated over the 50 replicates of the average σ_{sijt} (respectively p_{sijt}) for each scenario

Scenario	B-spline model			FP model		
	Median	Min	Max	Median	Min	Max
Linear	0.1007	0.1003	0.1010	0.1423	0.1172	0.1867
Logarithmic	0.1016	0.1014	0.1018	0.1439	0.1258	0.1761
Piecewise linear monotonic	0.1300	0.1190	0.1468	0.5701	0.5552	0.5875
Mixed	0.1321	0.1200	0.1497	0.6794	0.6666	0.6850
Non-monotonic	0.1399	0.1240	0.1760	0.9321	0.9321	0.9322
MTC	0.0829	0.0746	0.0939	0.1831	0.1236	0.3685
BEST-ITP	0.1352	0.1262	0.1426	0.2507	0.1712	0.4110
Piecewise linear monotonic (binary)	0.6128	0.5923	0.6246	0.6989	0.6830	0.7037
Non-monotonic (binary)	0.5940	0.5763	0.6011	0.7732	0.7372	0.8189
Piecewise linear (non-closed network)	0.1483	0.1203	0.1749	1.0382	1.0365	1.0408

- Hy-0.62-11.7); three injections of HA MW of 2.4-3.6 million Da (Euflexxa) (3 Hy-2.4-3.6); and three injections of Hylan GF-20 MW 6 million Da (Synvisc) (3 HyGF20). Placebo is the baseline treatment.
- (iii) The third data set contains information from 4 studies on patients affected by type 2 diabetes (T2D) and is described in [9] and [[7], Supplementary Materials]. Diabetes poses an enormous individual and societal burden, with high risk of major complication and diminished quality and length of

- life. The authors consider four treatments of which one is placebo and the others are oral anti-diabetic agents. Since the authors do not give further details we also label the treatments as 1 through 4, the baseline being the placebo treatment 1. The primary clinical outcome of interest is the hemoglobin A1c (HbA1c, %) reduction from baseline, with larger absolute value indicating better efficacy.
- (iv) The fourth data set is part of a large collection of studies on benefits and risks of drugs for treating

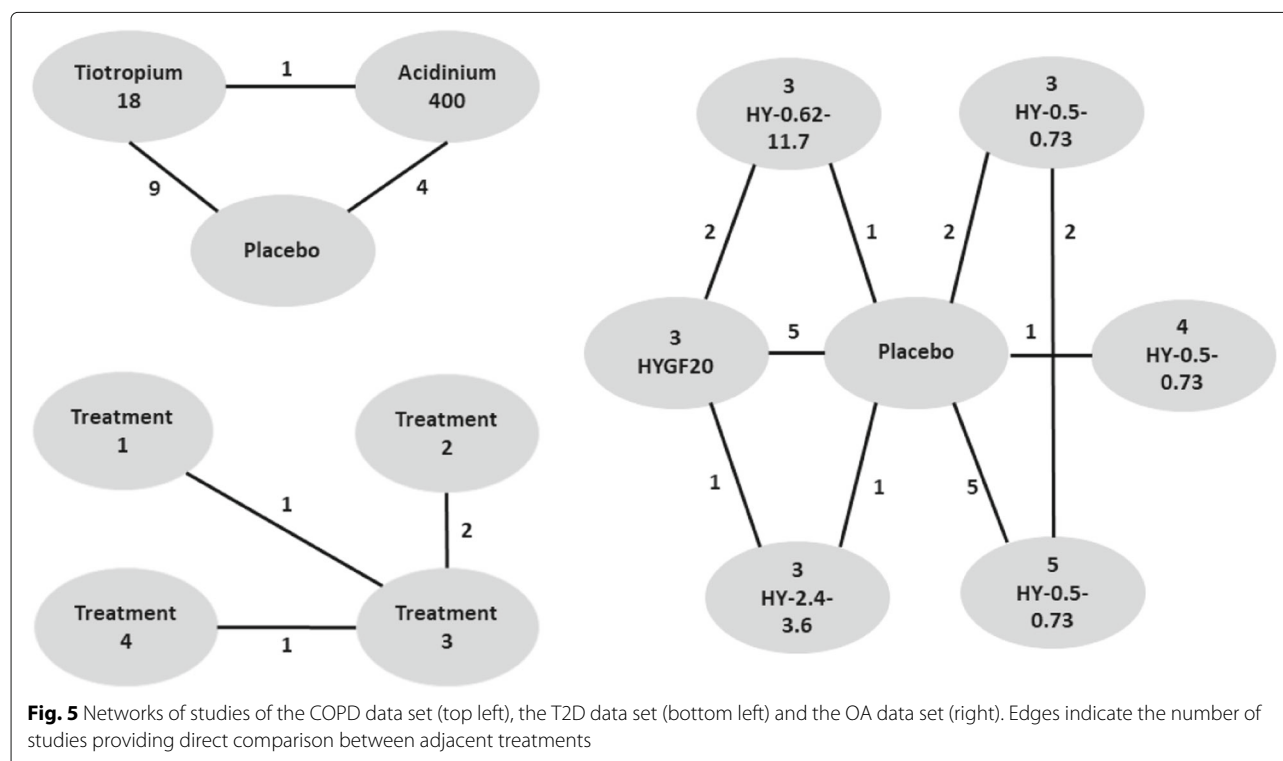
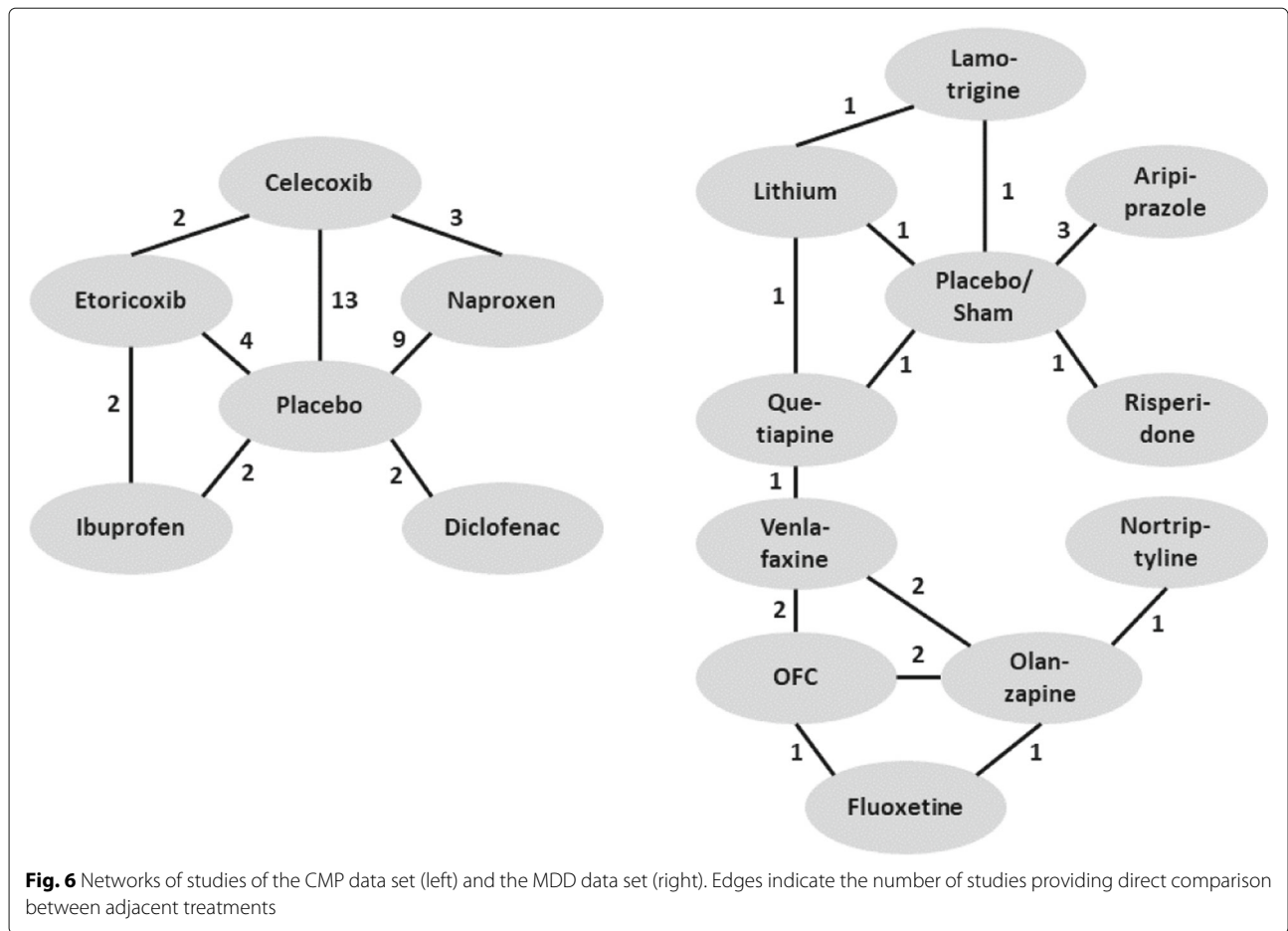


Fig. 5 Networks of studies of the COPD data set (top left), the T2D data set (bottom left) and the OA data set (right). Edges indicate the number of studies providing direct comparison between adjacent treatments



chronic musculoskeletal pain (CMP) in patients with osteoarthritis or rheumatoid arthritis. CMP disorders are associated with some of the poorest health related quality of life. Patients with pain experience severe restrictions on their functioning and ordinary daily activities. The data are collected in [36] and the authors conduct a Bayesian NMA. Here we concentrate on pain relief as the outcome of interest, measured by visual analogue scale (VAS), and only include studies with two follow-ups. Treatments are Diclofenac, Naproxen, Ibuprofen, Celecoxib, and Etoricoxib, which are compared to placebo. Outcomes are reported at 6 and 12 weeks (within 2-week range).

- (v) The fifth data set is taken from a comparative review of efficacy and tolerability of interventions for treatment resistant major depressive disorder (MDD) [37]. MDD is a mental disorder affecting about 10 – 15% of the general population. It is associated with depressed mood and/or loss of interest or pleasure in daily activities for more than two weeks. MDD is associated with significant morbidity and mortality, and often patients fail to achieve full

remission. Here we focus on efficacy outcome measured as CFB in the Montgomery-Asberg Depression Rating Scale (MADRS). Interventions are Aripiprazole, Fluoxetine, Lamotrigine, Lithium, Nortriptyline, Olanzapine, Olanzapine/Fluoxetine combination (OFC), Quetiapine, Risperidone, Venlafaxine, which are compared to placebo/sham. Outcomes are reported at 4,6 or 8 weeks.

For the data sets for which no information about the loss to follow-up is available, we assume that the sample sizes remain the same as at the start of the study. For studies within the NMAs that do not report the number of patients, we generate the sample size from a uniform distribution ranging from the minimum to the maximum number of patients across all studies of the respective NMA. Finally, following [8], we impute any missing standard deviation information specifying the prior distributions

$$sd_{ijt} \sim \text{Gamma}(\alpha_1, \alpha_2),$$

$$\alpha_i \sim \text{Uniform}(0, 10).$$

Posterior inference results are shown in Figs. 7, 8, 9, 10, 11.

Table 4 Summary of main features of the chronic obstructive pulmonary disease (COPD), osteoarthritis (OA), type 2 diabetes (T2D), chronic musculoskeletal pain (CMP), and treatment resistant major depressive disorder (MDD) data sets

	COPD	OA	T2D	CMP	MDD
Number of studies	13	16	4	27	13
of treatments	3	7	4	6	10
Number of follow ups	2–6	1–10	2–5	2	3–4
Number of patients per study	46–3006	20–295	88–575	12–6769	24–1147
Network type	closed	non-closed	non-closed	non-closed	non-closed

Discussion

From the simulation study (see Figs. 2, 3 and Table 2) (see Additional file 1 for further scenarios) it appears evident that the results obtained from the B-spline and the P-spline models are virtually identical. This is not surprising due to the use of the same basis functions and the small number of time points and knots in our application, resulting in a very small number of second order differences. Hence the local smoothing, the most attractive feature of P-splines, has very limited opportunity. Moreover, penalization is naturally incorporated in the Bayesian framework through the prior distribution. Indeed, the prior in (8) already shrinks small coefficients to zero while keeping substantial effects large. Furthermore, it is well known that the prior distribution performs a similar role as the penalty term in classical penalized regression, with the advantage that the penalty parameters can be estimated jointly with the other parameters of the model. From a Bayesian perspective, the main difference between the priors in (8) and in (9) is that the former assumes globally constant smoothing while the latter facilitates locally adaptive smoothing. This is relevant in applications with larger time horizon and appealing when the underlying function is oscillating. Given the limited number of time observations and the often regular behaviour of treatment effects over time in our application, the difference between B-splines and P-splines is negligible. For

this reason we only fit the B-spline model in the real data applications.

The simulation study results show that the spline based models offer a flexible strategy that is able to accommodate different time patterns and non-linearities in the underlying curve without the computational burden of the FP model, while still leading to better results. Cubic splines with four uniformly spaced knots appear to be able to provide accurate estimates and seem a natural choice, since it is common in NMA that few time measurements are available per treatment arm and study.

A possible extension of our model is the inclusion of within-study correlation between different time points. It has been shown that ignoring it might lead to increased residual variability, standard error of the pooled estimates and mean squared error. Including a specific parameter for within-study variability influences the borrowing of strength between time points and ultimately across studies. These considerations have been discussed in the literature and different methods to accommodate within-study correlation in meta-analysis have been proposed [see, e.g. [38, 39]]. In our work, the within-study correlation is mostly accounted for by flexibly modelling the mean function, which describes the time evolution of the treatment effect, by using B-splines. This strategy allows to capture structure in the data. Conditionally on the mean function and the remaining parameters in the

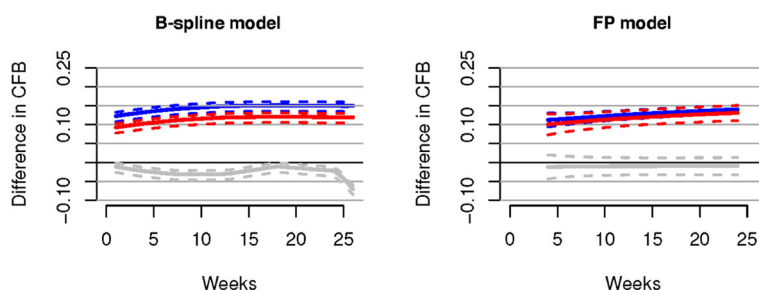
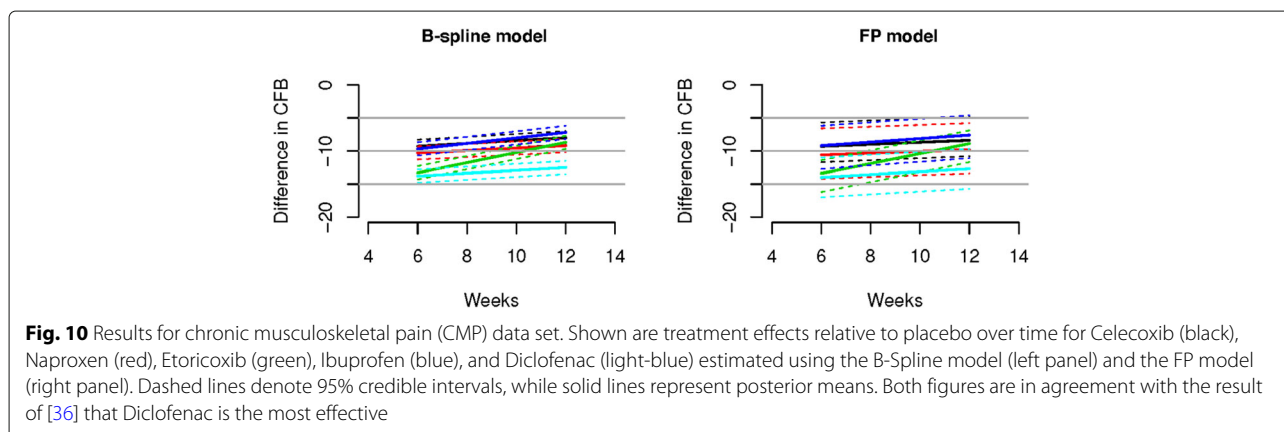
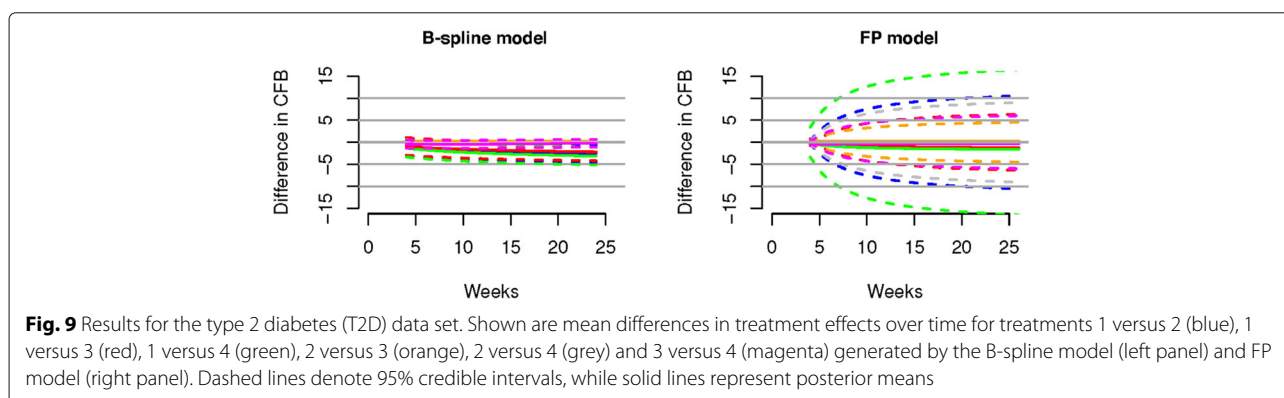
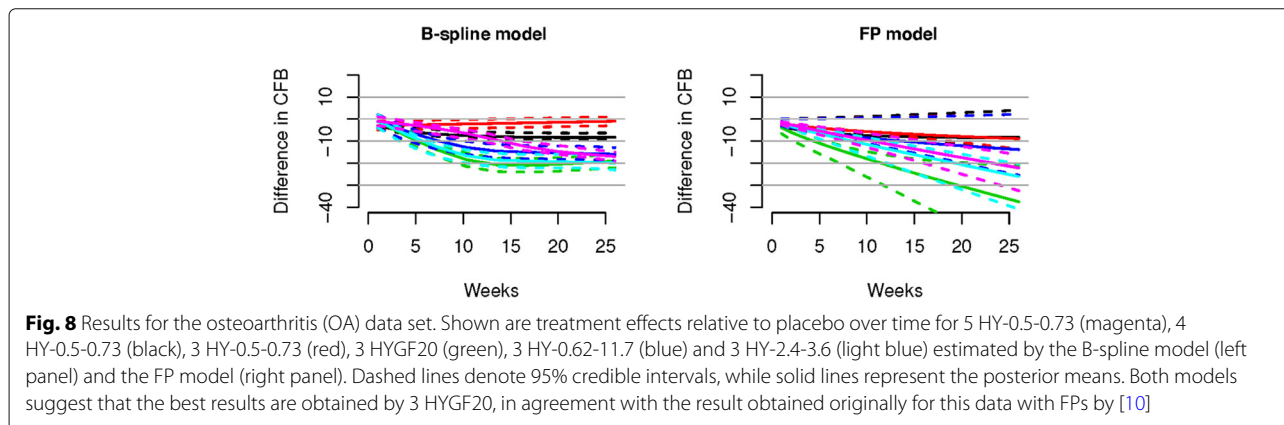


Fig. 7 Results for the chronic obstructive pulmonary disease (COPD) data set. Shown are estimation results as mean differences in treatment effects over time for AB400 versus placebo (red), TIO18 versus placebo (blue) and AB400 versus TIO18 (grey) generated by the B-spline model (left panel) and the FP model (right panel). Dashed lines denote 95% credible intervals, while solid lines represent the posterior mean. These results are in agreement with the conclusions in [35]



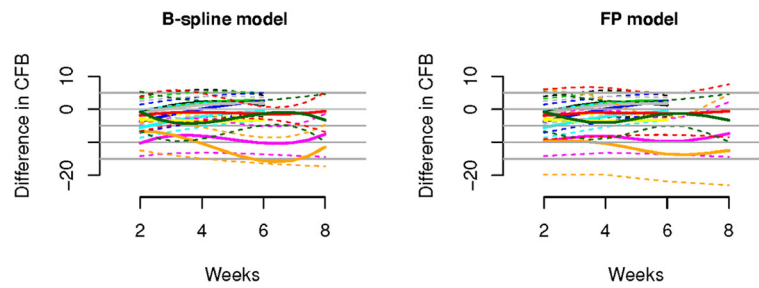


Fig. 11 Results for the major depressive disorder (MDD) data set. Shown are treatment effects relative to placebo over time for Nortriptyline (black), Lithium (red), Fluoxetine (green), Olanzapine (blue), OFC (light-blue), Risperidone (magenta), Aripiprazole (yellow), Venlafaxine (grey), Quetiapine (orange) and Lamotrigine (dark green) estimated using the B-Spline model (left panel) and the FP model. Dashed lines denote 95% credible intervals, while solid lines represent posterior means. The results are in qualitative agreement with [37]

model, the observations are independent. Borrowing of strength for inference across studies is mainly achieved through the joint modelling of the study-specific mean functions. In this framework it is in principle possible to include correlated errors among observations at different time points, and some proposals in this direction can be found in the literature. Here we do not introduce this further source of uncertainty given the limited amount of time-points per study per treatment, the fact that the observations are rarely close in time, the use of aggregated data, and finally because often (as shown in our data examples) the time pattern of the treatment effect is not noticeably oscillating but, on the contrary, is often linear or piecewise linear. Moreover, in the case of continuous responses we rescale in (4) the variances of the main outcomes, which achieves a similar effect as the rescaling of the variance by the autoregressive coefficient in an Autoregressive Model of order 1. For different types of responses other rescaling strategies need to be adopted. For example, for binomial or count data explicit correlation between time points could be introduced by modelling the outcome variables using a multivariate Binomial or a multivariate Poisson distribution, respectively.

The results of the real data applications confirm those of the simulation study. For the FP fit we considered first and second order FPs. The estimated profiles using the B-spline and FP models are presented in Figs. 7-9. B-splines provide narrower credible intervals for the posterior mean estimates, while still being able to detect slight changes over time. As a rule of thumb, if the 95% credible interval of the posterior distribution of the difference between two treatments does not cover zero, then we can conclude there is evidence in support of different effectiveness between them. In the results for the OA data it is evident that the B-spline model is able to detect non-monotonic time patterns, while the FP model forces a monotonic trend given the parameter choice which is dictated by the DIC. Furthermore, for the T2D and the CMP examples, the shape and size of the credible intervals obtained with

the FP model is due to the choice (by DIC) of a first order polynomial with power $p = -2$. Any choice of FP parameters generally imposes structure on the temporal pattern, which might not always be supported by the data. In particular, this is the case for the choice of the function β/t^2 when, as for the CMP data, only two data points per study and treatment arm are available. In Table 5 we report the specific FP selected by DIC for each data set. These examples confirm that the DIC might not be the optimal model choice criterion when confronted with non-linear effects. Indeed, while the DIC has been shown to be an approximation to a penalized loss function based on the deviance with a penalty derived from a cross-validation argument, it has been warned that this approximation is in fact only valid when the effective number of parameters p_{eff} in the model is considerably smaller than the number of independent observations n [see [20]]. Since the latter assumption is usually not satisfied in the case of NMAs the DIC can tend to under-penalize more complex models. The poor empirical performance of DIC compared to cross-validation when the assumption $p_{\text{eff}} \ll n$ is violated has also been highlighted by [21].

Conclusions

In this article we propose a random effect model for NMA of repeated measurements based on splines. The model is able to accommodate a large class of temporal

Table 5 Real data sets: DIC selection of the order M_F and of the power terms p_m in the FP model. RE and FE refer to random and fixed effects

Data set	M_F	p_1	p_2	RE/FE
COPD	2	-0.5	0	RE on β_2
OA	2	0.5	1	RE on β_2
T2D	1	-2		RE on β_1
CMP	1	-1		FE
MDD	2	0.5	0	FE

treatment effect patterns, allowing for direct and indirect comparisons of widely varying shapes of longitudinal profiles. We argue for a fixed choice of the order and automatic selection of uniformly spaced knots of the B-spline. The model is not restricted to continuous or binary outcomes, but can be extended to any type of response variable by specifying an appropriate link function. An important extension of the model is to account for the correlation structure between trial specific treatment effects, δ_{sj_i} and δ_{sj_k} , in multiple treatment arm studies. This can be achieved by decomposing the multivariate Normal distribution into a sequence of conditional univariate distributions as described in [10, 26]. In the context of longitudinal NMA, B-spline and P-spline models provide equivalent posterior inference but we believe that B-splines achieve a satisfactory level of penalization and smoothness for this application. Furthermore, we show that the B-spline model overcomes the methodological and computational limitations of the FP approach, which, beyond requiring extensive sensitivity analysis for each new scenario, might also suffer from potential shortcomings of the DIC. In detail investigation of model choice criteria for FP selection is needed, but it is beyond the scope of this work. One of the main consequences of a sub-optimal choice of the polynomial can be observed in the uncertainty quantification as represented by the credible intervals for the FP model. The B-spline model is useful in understanding treatment effects as well as between-study variability and naturally allows for different numbers of observations per study, different times of observations, different sample sizes across studies, as well as missing data, and there are arguments for better interpretability of its coefficients as compared to the FP model. The concerns of [10], about splines with fixed number of uniformly spaced knots being too restrictive in possible curve shapes, are not supported by our results in simulations and real data applications, which highlight the flexibility and efficiency of the B-spline approach.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12874-020-01113-9>.

Additional file 1: Table S6 contains simulated temporal patterns for additional simulation scenarios (i) linear, (ii) logarithmic, (iii) piecewise linear monotonic, (iv) mix of the previous with one treatment effect being constant. Figures S12, S13, S14 and S15 show respective estimated profiles obtained for the B-spline, P-spline and FP models, along with true values used to simulate the data. Figure S16 shows estimated profiles for case (iii) with binary outcomes. Figure S17 shows estimated profiles for a scenario, as in the "Simulation study" section, in which temporal patterns have been generated according to the mixed treatment comparison (MTC) model. Figure S18 shows estimated profiles for a scenario in which temporal patterns have been generated from the Bayesian evidence synthesis techniques – integrated two-component prediction (BEST-ITP) model. Figure S19 illustrates the influence of a non-closed network.

Additional file 2: Figures S20, S21 and S22 contain MCMC traceplots.

Additional file 3: JAGS code for B-spline model.

Additional file 4: R code to generate simulation scenario data.

Abbreviations

BEST-ITP: Bayesian evidence synthesis techniques – integrated two-component prediction; CFB: change from baseline; CMP: chronic musculoskeletal pain; COPD: chronic obstructive pulmonary disease DIC: deviance information criterion; FP: fractional polynomial; MCMC: Markov Chain Monte Carlo; MDD: major depressive disorder; MSE: mean squared error; MTC: mixed treatment comparison; NMA: network meta-analysis; T2D: type 2 diabetes

Acknowledgements

We would like to thank Andreas Karabis for providing the COPD data.

Authors' contributions

AH developed the spline component of the model, interpreted and drafted the manuscript. MT conducted the simulations and analyzed the application data. MDI proposed the idea, coordinated the research work and contributed to the discussion of the results and to the writing the manuscript. All author(s) read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

Data sets for the real data examples are publicly available in the cited references.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 10 December 2019 Accepted: 2 September 2020

Published online: 20 October 2020

References

- Béliveau A, Boyne DJ, Slater J, Brenner D, Arora P. BUGSnet: an R package to facilitate the conduct and reporting of Bayesian network meta-analyses. *BMC Med Res Methodol*. 2019;19(1):196.
- Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades A. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Med Dec Making*. 2013;33(5):641–56.
- Welton N, Ades A, Carlin J, Altman D, Sterne J. Models for potentially biased evidence in meta-analysis using empirically based priors. *J R Stat Soc Ser A Stat Soc*. 2009;172(1):119–36.
- Ishak KJ, Platt RW, Joseph L, Hanley JA, Caro JJ. Meta-analysis of longitudinal studies. *Clin Trials*. 2007;4(5):525–39.
- Lu G, Ades A, Sutton A, Cooper N, Briggs A, Caldwell D. Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Stat Med*. 2007;26(20):3681–99.
- Pedder H, Dias S, Bennetts M, Boucher M, Welton NJ. Modelling time-course relationships with multiple treatments: Model-based network meta-analysis for continuous summary outcomes. *Res Synth Methods*. 2019;10(2):267–86.
- Tallarita M, De Iorio M, Baio G. A comparative review of network meta-analysis models in longitudinal randomized controlled trial. *Stat Med*. 2019;38(16):3053–72.
- Dakin HA, Welton NJ, Ades A, Collins S, Orme M, Kelly S. Mixed treatment comparison of repeated measurements of a continuous endpoint: an example using topical treatments for primary open-angle glaucoma and ocular hypertension. *Stat Med*. 2011;30(20):2511–35.

9. Ding Y, Fu H. Bayesian indirect and mixed treatment comparisons across longitudinal time points. *Stat Med*. 2013;32(15):2613–28.
10. Jansen J, Vieira M, Cope S. Network meta-analysis of longitudinal data using fractional polynomials. *Stat Med*. 2015;34(15):2294–311.
11. Song F, Loke YK, Walsh T, Glenny A-M, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ*. 2009;338:1147.
12. Adalsteinsson E, Toumi M. Benefits of probabilistic sensitivity analysis—a review of nice decisions. *J Mark Access Health Policy*. 2013;1(1):21240.
13. Baio G, Dawid A. Probabilistic sensitivity analysis in health economics. *Stat Methods Med Res*. 2015;24(6):615–34. <https://doi.org/10.1177/0962280211419832>.
14. Berger JO. *Statistical decision theory and bayesian analysis*. New York: Springer; 2013.
15. Strasak AM, Umlauf N, Pfeiffer RM, Lang S. Comparing penalized splines and fractional polynomials for flexible modelling of the effects of continuous predictor variables. *Comput Stat Data Anal*. 2011;55(4):1540–51.
16. Royston P, Sauerbrei W. *Multivariable Model-building: a Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables*. New York: John Wiley & Sons; 2008.
17. Spiegelhalter D, Best N, Carlin B, Van Der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B Stat Methodol*. 2002;64(4):583–639.
18. Austin PC, Park-Wyllie LY, Juurlink DN. Using fractional polynomials to model the effect of cumulative duration of exposure on outcomes: applications to cohort and nested case-control designs. *Pharmacoepidemiol Drug Saf*. 2014;23(8):819–29.
19. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *J R Stat Soc Ser C Appl Stat*. 1994;43(3):429–53.
20. Plummer M. Penalized loss functions for Bayesian model comparison. *Biostatistics*. 2008;9(3):523–39.
21. Vehtari A, Lampinen J. Bayesian model assessment and comparison using cross-validation predictive densities. *Neural Comput*. 2002;14(10):2439–68.
22. Welton NJ, Sutton AJ, Cooper NJ, Abrams KR, Ades AE. *Evidence synthesis for decision making in healthcare*. New York: John Wiley & Sons, Ltd; 2012.
23. de Boor C. *A practical guide to splines*. New York: Springer; 2001.
24. Christensen O. *An introduction to frames and riesz bases*. New York: Springer; 2016.
25. Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction*. New York: Springer; 2009.
26. Cooper NJ, Sutton AJ, Morris D, Ades A, Welton NJ. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Stat Med*. 2009;28(14):1861–81.
27. Eilers PH, Marx BD. Flexible smoothing with B-splines and penalties. *Stat Sci*. 1996;11(2):89–102.
28. Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M. A review of spline function procedures in R. *BMC Med Res Methodol*. 2019;19(1):46.
29. Ruppert D, Wand MP, Carroll RJ. *Semiparametric Regression*. Cambridge: Cambridge university press; 2003.
30. Aguilera AM, Aguilera-Morillo M. Comparative study of different B-spline approaches for functional data. *Math Comput Model*. 2013;58:1568–79.
31. Lang S, Brezger A. Bayesian P-splines. *J Comput Graph Stat*. 2004;13(1):183–212.
32. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd international workshop on distributed statistical computing*, vol 124, Issue 125.10. 2003. p. 1–10.
33. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods*. 2010;1(2):97–111.
34. Gelman A, Rubin D. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992;7(4):457–72.
35. Karabis A, Lindner L, Mocarski M, Huisman E, Greening A. Comparative efficacy of aclidinium versus glycopyrronium and tiotropium, as maintenance treatment of moderate to severe COPD patients: a systematic review and network meta-analysis. *Int J Chronic Obstructive Pulm Dis*. 2013;8:405–23.
36. van Walssem A, Pandhi S, Nixon RM, Guyot P, Karabis A, Moore RA. Relative benefit-risk comparing diclofenac to other traditional non-steroidal anti-inflammatory drugs and cyclooxygenase-2 inhibitors in patients with osteoarthritis or rheumatoid arthritis: a network meta-analysis. *Arthritis Res Ther*. 2015;17(1):66.
37. Papadimitropoulou K, Vossen C, Karabis A, Donatti C, Kubitz N. Comparative efficacy and tolerability of pharmacological and somatic interventions in adult patients with treatment-resistant depression: a systematic review and network meta-analysis. *Curr Med Res Opin*. 2017;33(4):701–11.
38. Ahn JE, French JL. Longitudinal aggregate data model-based meta-analysis with NONMEM: approaches to handling within treatment arm correlation. *J Pharmacokinet Pharmacodyn*. 2010;37(2):179–201.
39. Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. *J R Stat Soc Ser A Stat Soc*. 2009;172(4):789–811.
40. Higgins J, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc*. 2009;172(1):137–59.
41. Li T, Puhan MA, Vedula SS, Singh S, Dickersin K. Network meta-analysis highly attractive but more methodological research is needed. *BMC Medicine*. 2011;9(1):79.
42. Lipsey MW, Wilson DB. *Practical meta-analysis*. Thousand Oaks, California: Sage Publications, Inc; 2001.
43. Lu G, Ades A. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23(20):3105–24.
44. Glenny A, Altman D, Song F, Sakarovitch C, Deeks J, D'amico R, Bradburn M, Eastwood A. Indirect comparisons of competing interventions. *Health Technol Assess*. 2005;9(26):1–134.
45. Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, Lee K, Boersma C, Annemans L, Cappelleri JC. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR task force on indirect treatment comparisons good research practices: part 1. *Value Health*. 2011;14(4):417–28.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

