Received <day> <Month>, <year>; | Revised <day> <Month>, <year>; | Accepted <day> <Month>, <year>

DOI: xxx/xxxx

ARTICLE TYPE

A comparative review of network meta-analysis models in longitudinal randomized controlled trial

Marta Tallarita*[1] | Maria De Iorio[1,2] | Gianluca Baio[1]

[1]Department of Statistical Science, University College London, London, UK
[2]Yale-NUS College, Singapore

Correspondence
*Email: m.tallarita@ucl.ac.uk

Present Address
Gower Street, London WC1E 6BT UK

Abstract

Network meta-analysis technique extends the standard meta-analysis methods, allowing pairwise comparison of all treatments in a network in the absence of head-to-head comparisons. Traditional network meta-analysis models consider a single endpoint for each trial. However, in many cases trials in the network have different durations and/or report data at multiple time points. Moreover these time points are often not the same for all trials. In this work we review the most relevant methods that incorporate multiple time points and allow indirect comparisons of treatment effects across different longitudinal studies. In particular we focus on the *Mixed Treatment Comparison* (MTC) developed by Dakin *et al* (1), the Bayesian evidence synthesis techniques — integrated two-component prediction (BEST-ITP) developed by Ding *et al* (2) and on the more recent method based on fractional polynomials by Jansen *et al* (3). We highlight the main features of each model and illustrate them in simulations and in a real data application. Our study shows that methods based on fractional polynomials offer a flexible modeling strategy in most applications.

KEYWORDS:
network meta-analysis; mixed treatment comparison; Bayesian evidence synthesis techniques; fractional polynomials; longitudinal studies

1 | INTRODUCTION

Network meta-analysis (NMA) is a well established technique used to synthetise a complex mass of evidence about a potentially large number of interventions (4), (5), (6). NMA is applied in several research areas, but arguably it plays a pivotal role in health technology assessment (HTA) (7). This is the process by which bodies such as the National Institute for Health and Care Excellence (NICE) in the UK, the Pharmaceutical Benefits Advisory Committee in Australia or CADTH in Canada assess whether a new intervention provides value for money, or in other words, whether it is "cost-effective". While often clinical studies are primarily designed to assess the clinical effectiveness of a new intervention, possibly against placebo or standard of care, a full economic evaluation requires a comprehensive assessment against all possible relevant comparators. Thus, HTA is often based on a combination of individual- and aggregated-level data informing the model on a possibly large number of comparisons.

One feature that is often relevant in HTA is that individuals are observed at multiple time points throughout the follow up. From the statistical point of view, this implies that the repeated measurements will tend to be correlated and failure to account for this will lead to biased estimate of the underlying treatment effects. Modelling for NMA of repeated measures has been widely considered in the statistical and health economic literature, in recent times. These methods have been applied in both

frequentist and Bayesian frameworks. Bucher *et al* (8) have presented an indirect method in meta-analysis that preserves the randomization of the originally assigned patient groups. However, this method allows to compare two treatments only if a common comparator exists. Lumley *et al* (9) have proposed a more flexible technique but this model can not handle multiple arm studies. An alternative method to avoid this problem is Olkin's linear regression method (10). Wandel *et al* (11) introduced a Bayesian repeated measures network meta-analysis model that allowed the synthesis of multiple time points. Finally, Lu *et al* (12) have proposed Bayesian hierarchical models for mixed treatment comparison using data sets with more than one follow-up time, and where different trials may report at different times. In particular, they focus on the number of individuals that recover over a specified time interval. By dividing the overall period of observation in intervals, they model the number of patients (still) at risk who are recovered over a specified time interval. In this way, they reconstruct the available data on each trial arm into a series of independent observations (over different time periods): each observation then consists of a separate reporting period, and the observations on each period follow a binomial distribution. The main parameter of interest becomes then the probability of recovering over a pre-specified time interval. Moreover, once a patient is recovered he cannot re-experience the event.

Here we consider the three most recent methods suggested in the literature for NMA in the case of repeated measurements, whose main inferential focus treatment effect over time, allowing for flexible temporal patterns. These are the *Mixed Treatment Comparison* (MTC), developed by Dakin *et al* (1); the Bayesian evidence synthesis techniques — integrated two-component prediction (BEST-ITP) developed by Ding *et al* (2); and the more recent method based on fractional polynomials by Jansen *et al* (3). Our objective is to provide a comparative review of these methods using a simulation study as well as real-world data, with a view of identifying specific characteristics that would make either more appropriate.

We note here that all the models are originally specified under a Bayesian approach; this is mainly due to two reasons. Firstly, NMA is a process that inherently combines different (and possibly not completely homogeneous) sources of evidence. While it is possible to use frequentist methods for inference, a Bayesian approach is recognised as particularly effective as combining "modules", e.g. through the specification of joint prior distributions for some parameters of interest. Secondly, HTA is ultimately focussed on decision-making, rather than just statistical inference. Again, a Bayesian approach is particularly helpful in this case, for example in terms of allowing a straightforward quantification of the impact of model and parameter uncertainty on the optimal decision, given current data (a process often termed "Probabilistic Sensitivity Analysis", which is usually mandatory in many jurisdictions (13)). For greater detail, see Berger 2013 (14) and Baio and Dawid 2012 (**?**).

Our paper is structured as follows. In §2 we briefly review the general set up and distributional assumptions for the three models; then in §3 we present a simulation study aimed at testing the performance of the methods under a range of scenarios in terms of the underlying treatment effect. In section §4 and §5 we illustrate the models using two real-world datasets, one collecting information on patients affected by with chronic obstructive pulmonary disease (COPD) and the other investigating treatments for osteoarthritis(OA) of the knee . Finally, in §6 we summarise our conclusions and recommendations.

2 | MODELS

In this section we briefly review the three main models available in the literature to perform network meta-analysis of longitudinal studies. In this context the main objective is the evaluation over time of some suitably defined response, e.g. continuous outcomes, binary or count data. Let y_{sjt} denote the response in study s, at time t in the jth treatment arm. We assume:

$$y_{sjt} \sim f\left(y_{ij}|\theta_{sjt}, \sigma^2_{sjt}\right) \tag{1}$$

where f is a probability density (usually parametric), $s = 1, \dots, S$, $j = 1, \dots, J$ and $t = 0, \dots, T$. The main parameter of clinical interest is θ_{sjt}, which measures the treatment effect in study s of intervention j at time point t, on a suitable scale (*e.g.* logit in the case of binary variables or log when f is a Poisson distribution).

In this paper we focus on continuous and binary responses. When the measurements y_{sjt} are continuous Equation (1) becomes:

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

while in case of a binary outcome the likelihood function becomes:

$$\begin{aligned} y_{sjt} &\sim \text{Binomial}\left(n_{sjt}, p_{sjt}\right), \\ \text{logit}(p_{sjt}) &= \theta_{sjt}. \end{aligned}$$

The fundamental difference among the three NMA methods is in the way in which the predicted (mean) outcome (θ_{sjt}) and the variance (σ^2_{sjt}), at time point t in arm j of trial s, are specified.

2.1 | Mixed treatment comparison model

In the MTC model the mean outcome is specified in terms of a study- and time- specific baseline and an additive term describing the relative effect of each treatment. For simplicity, we set as the reference arm the treatment $j = 1$. The original specification (1) presents 6 alternative versions of this general structure, where both the baseline and the relative effect may vary over time. Here, we consider a general formulation

$$\theta_{sjt} = \mu_{st} + \delta_{sj},$$

where, μ_{st} is the study- and time-specific effect pooled across treatment arms, while δ_{sj} is the study-specific arm deviation from the reference arm, with $\delta_{s1} := 0$. In this framework the parameters μ_{st} are assumed to be independent across time points in each study and the relative treatment effects δ_{sj} to be constant over time.

In particular, μ_{st} is assumed to be normally distributed for each time-point and each study. Conversely, the parameters δ_{sj} are modelled as structured "random effects", as follows

$$\delta_{sj} \sim \text{Normal}\left(d_j - d_1, \sigma^2_\delta\right)$$

Following Dakin *et al* (1), the parameters d_j are independently assigned vague Normal prior distribution centered on 0 and with large variance, whereas the between-studies standard deviation σ^2_δ is given a uniform prior on a wide interval.

Finally, the variances for the observed data are modelled as

$$\sigma^2_{sjt} = \left(\frac{\text{sd}_{sjt}}{\sqrt{n_{sjt}}}\right)^2 \qquad \text{for all } s, j, t,$$

where sd_{sjt} is an estimate of the standard deviation (possibly available in the literature) and n_{sjt} is the observed sample size for each treatment arm and time point in each study.

2.2 | Bayesian evidence synthesis techniques - integrated two-component prediction model

Ding *et al* (2) present two different models: one is based on fixed effects, while the other includes random effects. The general model with fixed effects specifies the mean and the variance of the observed outcome as

$$\begin{aligned} \theta_{sjt} &= (\phi_s + \delta_j)\left(\frac{1 - e^{p_j t}}{1 - e^{p_j T}}\right) \\ \sigma^2_{sjt} &= \left[\frac{\text{sd}_{sjt}}{\sqrt{n_{sjt}}}\left(\frac{1 - e^{p_j t}}{1 - e^{p_j T}}\right)\right]^2, \end{aligned} \tag{2}$$

where the parameter δ_j indicates the j-th treatment mean effect at the end follow up period (time T). Here, the time period T is the same for all studies. It is straightforward to allow different time periods for different studies, by changing T to T_s. Moreover, the parameter p_j determines the shape of the j-th treatment effect over time. This parameter is assigned a uniform prior distribution defined on a negative interval. This implies that:

$$0 \leq \left(\frac{1 - e^{p_j t}}{1 - e^{p_j T}}\right) \leq 1.$$

More specifically, if p_j is large the rate of the change attributable to treatment j grows fast at the beginning and reaches a plateau quickly, whereas when p_j is close to 0, the trend is almost linear.

In the fixed effect specification, δ_j is assigned a vague Normal prior again centered on 0 and with large variance. The random effects version can be easily obtained by assuming a study-specific treatment effect δ_{sj} in Equation 3, *e.g.*

$$\delta_{sj} \sim \text{Normal}\left(d_j - d_1, \sigma^2_\delta\right),$$

where the parameters d_j are assigned vague Normal priors and the between-studies standard deviation is given a wide uniform prior. Notice that in our simulations in §3 and examples in §4 we use the random effect model for a fairer comparison.

2.3 | Fractional polynomials

Jansen *et al* (3) propose an approach to network meta-analysis based on fractional polynomials (FPs), a family of flexible basis functions used to describe the relationships among variables, *e.g.* in a regression setting. This model assumes nonlinear dynamics of treatment effects over time.

A FP structure of order M for the mean outcome of the j-th treatment in study s is defined as

$$\theta_{sjt} = \begin{cases} \beta_{0sj} + \sum_{m=1}^{M} \beta_{msj} t^{p_m} & \text{if } p_1 \neq \cdots \neq p_M \\ \beta_{0sj} + \beta_{1sj} t^{p} + \sum_{m=2}^{M} \beta_{msj} t^{p} [\ln(t)]^{m-1}, & \text{if } M > 1, p_1 = \cdots = p_M = p \end{cases} \tag{3}$$

with $t^0 := \ln(t)$. Jansen *et al* (3) suggest selecting the power p_m from the set $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$, for $m = 1, \ldots, M$.

The vector $\boldsymbol{\beta_{sj}} = (\beta_{0sj}, \ldots, \beta_{Msj})^\intercal$ is modelled as

$$\begin{pmatrix} \beta_{0sj} \\ \vdots \\ \beta_{Msj} \end{pmatrix} = \begin{pmatrix} \mu_{0s} \\ \vdots \\ \mu_{Ms} \end{pmatrix} + \begin{pmatrix} \delta_{0sj} \\ \vdots \\ \delta_{Msj} \end{pmatrix},$$

where the vector $\boldsymbol{\mu}_s = (\mu_{0s}, \ldots, \mu_{Ms})^\intercal$ denotes the study specific mean, while the vector $\boldsymbol{\delta}_j = (\delta_{0sj}, \ldots, \delta_{Msj})^\intercal$ represents the study- and time-specific effects of treatment j relative to the reference treatment, arbitrarily coded as $j = 1$. The parameters δ_{msj} are modelled assuming a multivariate Normal distribution with the pooled estimates expressed in terms of the overall reference treatment $j = 1$:

$$\begin{pmatrix} \delta_{0sj} \\ \vdots \\ \delta_{Msj} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} d_{0j} - d_{01} \\ \vdots \\ d_{Mj} - d_{M1} \end{pmatrix}, \Sigma \right), \tag{4}$$

where the vector $\boldsymbol{d}_j = (d_{0j}, \ldots, d_{Mj})^\intercal$ are assigned vague Normal priors, for $j > 1$, and $\boldsymbol{d}_1 = \mathbf{0}$ and with

$$\begin{pmatrix} \delta_{0s1} \\ \vdots \\ \delta_{Ms1} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \qquad \text{and} \qquad \Sigma = \begin{pmatrix} \sigma_0^2 & \cdots & \sigma_0 \sigma_M \lambda_{0M} \\ \vdots & \ddots & \vdots \\ \sigma_0 \sigma_M \lambda_{0M} & \cdots & \sigma_M^2 \end{pmatrix}.$$

The covariance matrix Σ captures between-study heterogeneity of the treatment effect parameters δ_{msj}, where $\sigma_m^2 = \text{Var}\left[\delta_{msj}\right]$ and λ_{lk} quantifies the correlation across the treatment effect parameters, with $m, l, k \in \{0, \ldots, M\}$.

Under a fixed-effects model, the formulation in (4) is replaced by

$$\begin{pmatrix} \delta_{0sj} \\ \vdots \\ \delta_{Msj} \end{pmatrix} = \begin{pmatrix} d_{0j} - d_{01} \\ \vdots \\ d_{Mj} - d_{M1} \end{pmatrix}$$

and as a result it is not necessary to estimate the between-study covariance matrix.

Finally, the variance of the main outcome

$$\sigma_{sjt}^2 = \left[\left(\frac{\text{sd}_{sjt}}{\sqrt{n_{sjt}}} \right) \Big/ \left(1 - \rho^2\right) \right]^2$$

is modelled as a function of the corresponding standard error adjusted by a factor $(1 - \rho^2)$, where ρ indicates the correlation coefficient between subsequent time points. Often, ρ is assumed to be known, but in a fully Bayesian setting it can be treated as unknown and object of inference, by specifying an appropriate prior distribution.

3 | SIMULATION STUDY

In order to evaluate the performance of the three models described above, we conduct several simulations studies. We investigate different scenarios to highlight the main differences among the modelling strategies. Our goal is to reproduce different time-pattern for the main effect and investigate the ability of each model to recover such structure. We consider treatment effects that

over time: *a)* are linear; *b)* decrease logarithmically; *c)* are constant and *d)* non-monotonic treatment effect, corresponding to a situation in which a treatment could be first beneficial and then detrimental. Moreover, to provide fair comparisons we also generate data based on the models by Dakin *et al* (1) and Ding *et al* (2). Finally, we extend a continuous outcome for longitudinal studies to a binary outcomes.

For each scenario, we randomly generate 50 datasets according to the following specifications. We then fit the NMA models on each datasets and average the results over the 50 simulated datasets. When generating the data we follow a similar simulation strategy as in Ding *et al* (2).

We simulate data from 3 hypothetical studies with 2 treatment arms. Figure 1 shows graphically the network of studies used for Scenario 1–7; circles indicate treatments, while edges connecting them indicate the availability of direct evidence as provided by a study. In particular, in Figure 1 study 1 compares treatment A and B; study 2 has treatments A and C and study 3 matches B and C. For this simple example, the network is "closed", meaning that there is at least one study providing direct evidence for all possible pairwise comparisons. In Supplementary Material (Section 2) we also perform a simulation study using non-closed network, which shows that the conclusions of the comparison are not affected by the nature of the network.

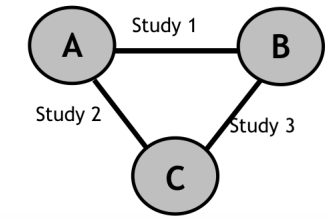

FIGURE 1 Network of studies used for Scenario 1–7 of the simulation study. The different treatments are represented by circles. Two treatments are linked with a black lines if a direct comparison between them is available.

Table 1 describes the details of the simulations in terms of the time points at which follow up occurs as well as the number of individuals in each of the studies arms.

TABLE 1 Simulation parameters. For each study different treatments, times of observation and number of subjects per arm are specified.

Study	Treatment	Time of Observation	Number of Subjects per arm (n_s)
Study 1	A and B	4, 8, 12, and 24 weeks	100
Study 2	A and C	4, 12, and 24 weeks	120
Study 3	B and C	4, 8, and 12 weeks	130

For the first five scenarios, to generate data from a meta-analysis of study, we first simulate individual level observations as $Z_{isjt} \sim \text{Normal}(\theta_{sjt}, \tau_s^2)$, for $i = 1, \dots, n_s$, where n_s is the number of observation in study s. In addition, we set $\theta_{sjt} = \alpha_s + \gamma_{jt}$ with $\alpha_s \sim \text{Normal}(0, 10)$, $\tau_1^2 = 1$, $\tau_2^2 = 2$ and $\tau_3^2 = 4$ and we show γ_{jt}, the true relationships between time and the change from baseline, in Figure 2 . In the simulations, we assume that the variance of the outcome is constant over time and across treatments as our main interest is to investigate the performance of each model in capturing the time structure of the main outcome. Obviously this assumption may not be tenable in many real data applications as shown in section § 4 and § 5 and in Supplementary Material (Section 3).

The main outcome of each study is reported as sample means $Y_{sjt} = \frac{\sum_{i=1}^{n_s} Z_{ijst}}{n_s}$, or $Y_{sjt} \sim \text{Normal}\left(\theta_{sjt}, \sigma_s^2\right)$, with $\sigma_s^2 = \tau_s^2/n_s$. In other words, this setting amounts to including a set of independent study-effects α_s, while the component γ_{jt} in θ_{jst} represents the effect of the treatment over time.

Through the different specifications of the curve γ_{jt} we create the various scenarios described at the beginning of this section, according to the following scheme.

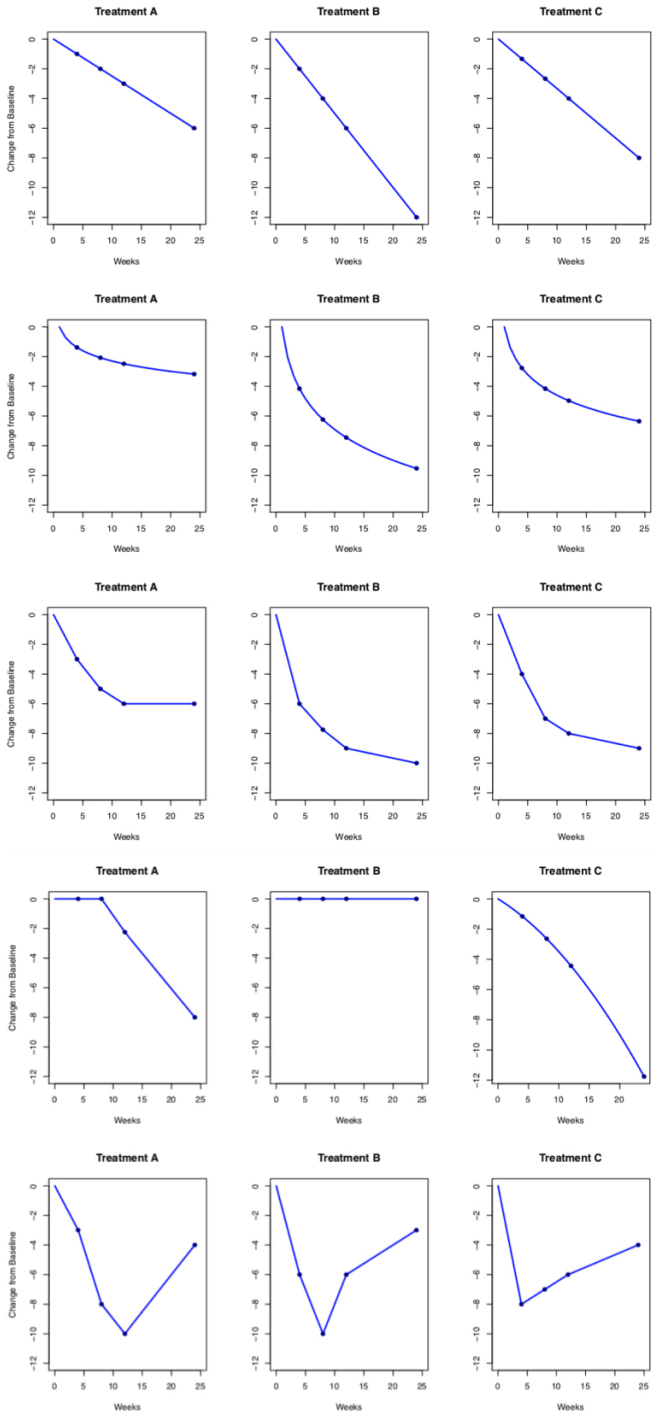

FIGURE 2 Treatment effect, γ_{jt}, over time. The columns refer to different treatments (A, B, C), while the rows represents the different simulation scenarios.

Scenario 1. The treatment effects are linear over the time with different slopes:

- Treatment A: $\gamma_{At} = -\frac{t}{4}$;
- Treatment B: $\gamma_{Bt} = -\frac{t}{2}$;

- Treatment C: $\gamma_{Ct} = -\frac{t}{3}$.

Scenario 2. We consider treatment effects with logarithmic decay over time:

- Treatment A: $\gamma_{At} = -\log(t+1)$;
- Treatment B: $\gamma_{Bt} = -3\log(t+1)$;
- Treatment C: $\gamma_{Ct} = -2\log(t+1)$;

Scenario 3. We assume a monotonic piecewise linear curves:

- Treatment A: $\gamma_{At} = \begin{cases} -\frac{3t}{4} & \text{if} \quad 0 \le t \le 4 \\ -\frac{t}{2} - 1 & \text{if} \quad 4 < t \le 8 \\ -\frac{t}{4} - 3 & \text{if} \quad 8 < t \le 12 \\ -6 & \text{if} \quad 12 < t \le 24 \end{cases}$;
- Treatment B: $\gamma_{Bt} = \begin{cases} -\frac{3t}{2} & \text{if} \quad 0 \le t \le 4 \\ -\frac{10t}{23} - \frac{17}{4} & \text{if} \quad 4 < t \le 8 \\ -\frac{5t}{16} - \frac{21}{4} & \text{if} \quad 8 < t \le 12 \\ -\frac{t}{12} - 8 & \text{if} \quad 12 < t \le 24 \end{cases}$;
- Treatment C: $\gamma_{Ct} = \begin{cases} -t & \text{if} \quad 0 \le t \le 4 \\ -\frac{3t}{4} - 1 & \text{if} \quad 4 < t \le 8 \\ -\frac{t}{4} - 5 & \text{if} \quad 8 < t \le 12 \\ -\frac{t}{12} - 7 & \text{if} \quad 12 < t \le 24 \end{cases}$.

Scenario 4. We introduce also a treatment effect that is constant over the time (treatment B):

- Treatment A: $\gamma_{At} = \begin{cases} 0 & \text{if} \quad 0 \le t \le 8 \\ -\frac{t}{2} + 4 & \text{if} \quad 8 < t \le 24 \end{cases}$;
- Treatment B: $\gamma_{Bt} = 0$;
- Treatment C: $\gamma_{Ct} = \frac{(-t^2 - 25t)}{100}$.

Scenario 5. We assume non-monotonic piecewise linear curves:

- Treatment A: $\gamma_{At} = \begin{cases} -\frac{3t}{4} & \text{if} \quad 0 \le t \le 4 \\ -\frac{5t}{2} + 2 & \text{if} \quad 4 < t \le 8 \\ -\frac{t}{2} - 4 & \text{if} \quad 8 < t \le 12 \\ \frac{t}{2} - 16 & \text{if} \quad 12 < t \le 24 \end{cases}$;
- Treatment B: $\gamma_{Bt} = \begin{cases} -\frac{3t}{2} & \text{if} \quad 0 \le t \le 4 \\ -t - 2 & \text{if} \quad 4 < t \le 8 \\ t - 18 & \text{if} \quad 8 < t \le 12 \\ \frac{t}{4} - 9 & \text{if} \quad 12 < t \le 24 \end{cases}$;
- Treatment C: $\gamma_{Ct} = \begin{cases} -\frac{8t}{4} & \text{if} \quad 0 \le t \le 4 \\ \frac{t}{4} - 9 & \text{if} \quad 4 < t \le 12 \\ \frac{t}{6} - 8 & \text{if} \quad 12 < t \le 24 \end{cases}$.

For **Scenario 6**, we simulate data from the MTC model with the following parameters:

$$\begin{aligned} \mu_{st} &\sim \text{Normal}(-s, 1), \\ \delta_{sj} &\sim \text{Normal}(0.5, 1), \\ sd &= 1.2. \end{aligned}$$

Finally, in the last scenario **Scenario 7**, we simulate data from the BEST-ITP model with the following parameters:

$$\begin{aligned} \phi_s &\sim \text{Normal}\,(-3, 1)\,, \\ \delta &= (0, -0.5, -1), \\ sd &= 1.2, \\ p_j &= (-0.1, -0.15, -0.15). \end{aligned}$$

Posterior inference for these examples, as well as for the real data applications in § 4 and § 5 and in Supplementary Material (Section3), can be performed through a standard Gibbs sampler algorithm, which we implement in JAGS (15). The first 5,000 iterations are discarded as 'burn-in' and the final sample size on which inference is based is 10,000 samples. We check convergence of the chain through the Gelman-Rubin potential scale reduction factor (16). The Gelman-Rubin diagnostic is evaluated by running multiple chains from different initial values and then comparing the estimated between-chains and within-chain variances for each model parameter. Large differences between these variances indicate that convergence has not been reached yet.

3.1 | Models specification

For all the methods, we generally specify vague prior distributions for the parameters of interest. In the case of the MTC we follow the original proposal presented in Dakin *et al* (1) and assume

$$\mu_{st} \sim \text{Normal}\left(0, 10^4\right), \qquad d_{s_j} \sim \text{Normal}\left(0, 10^4\right) \qquad \text{and} \qquad \sigma_\delta \sim \text{Uniform}(0, 20).$$

Following Ding *et al* (2), for the BEST-ITP we assume

$$\phi_s \sim \text{Normal}\left(0, 10^4\right) \qquad \text{and} \qquad p_j \sim \text{Uniform}(-10, 0).$$

Moreover, we follow the suggestion of the previous method (MTC model) to specify the random effect distribution as

$$d_j \sim \text{Normal}\left(0, 10^4\right) \qquad \text{and} \qquad \sigma_\delta \sim \text{Uniform}(0, 20).$$

Finally, regarding the FP approach, in the simulation study we investigate the performance of a first- and a second-order fractional polynomial model ($M = 1, 2$, respectively). In fact, in medical applications, FP1 and FP2 transformations are often, with higher order transformations being used rarely. FP1 and FP2 functions allow representation of a wide range of non-linear relationships. A first-order fractional polynomial is obtained by describing the effect as a function of transformed time t in a linear model: FP of degree 1 functions are always monotonic, while FP degree 2 functions may be monotonic or unimodal. For greater detail, see Royston and Sauerbrei 2008 (17).

When $M = 1$, Equation 3 reduces to

$$\theta_{sjt} = \beta_{0sj} + \beta_{1sj} t^p, \tag{5}$$

while for $M = 2$ it becomes

$$\theta_{sjt} = \begin{cases} \beta_{0sj} + \beta_{1sj} t^{p_1} + \beta_{2sj} t^{p_2} & \text{if } p_1 \neq p_2 \\ \beta_{0sj} + \beta_{1sj} t^p + \beta_{2sj} t^p \left[\ln(t)\right], & \text{if } p_1 = p_2 = p. \end{cases} \tag{6}$$

For consistency with the other models, we consider only specifications that include a random effect. In particular, when $M = 1$, there is only one between-study heterogeneity parameter, σ_0^2, related to the relative treatment effects for β_{0sj}. In our simulations, we consider another separate analysis where the between-study heterogeneity is assumed to affect treatment effects only in terms of β_{1sj}. Moreover, if $M = 2$, we consider also the relative treatment effects for β_{2sj}.

All the models are completed by specifying suitable prior distributions for the vectors $\boldsymbol{\mu}_s$ and $\boldsymbol{d}_j = (d_{0j}, \ldots, d_{Mj})^\intercal$. In particular, we define $\mathbf{0}$ as the zero vector of length $(M + 1)$ and model

$$\boldsymbol{\mu}_s \sim \text{Normal}\left(\mathbf{0}, \boldsymbol{T}_\mu\right) \qquad \text{and} \qquad \boldsymbol{d}_j \begin{cases} = \mathbf{0} & \text{for } j = 1 \\ \sim \text{Normal}\left(\mathbf{0}, \boldsymbol{T}_d\right) & \text{for } j > 1 \end{cases}.$$

The diagonal elements of the prior covariance matrices are set equal to 10^4, while the other elements are equal to zero. Thus, $\boldsymbol{T}_\mu = \boldsymbol{T}_d = 10^4 \mathbb{I}$, with $\mathbb{I}$ denoting the identity matrix of appropriate dimension. Finally, we assume that the correlation ρ follows a Uniform distribution in the interval $(0, 0.95)$.

3.2 | Selection of M and p_m in the FP model

Implementing the FP model requires selecting a value for M and fixing the power terms p or (p_1, p_2) in Equations (5) or (6), respectively. Following Jansen *et al* (3), we use the Deviance Information Criterion (DIC), originally proposed by Spiegelhalter *et al* (18) as an estimate of expected predictive error to guide the selection of the FP order and power terms, upon a set of possible choices. The DIC is defined as the sum of $\bar{D}$, the posterior mean of the model deviance and p_D, a posterior estimate of the effective number of parameters, acting as a penalty for more complex models in order to limit the effect of overfitting. Models with smaller DIC should be preferred. For each of the 50 replicates of each scenarios, we compare the values of DIC of the different first-order fractional polynomial models (that is 7 different models, one for each value of p in our pre-specified range) and of the different second-order fractional polynomial models (28 models in total, given by all possible combinations of p_1 and p_2). For each scenario, we select M and corresponding power terms based on the DIC value. Table 2 reports the power that achieves minimum DIC more often corresponding to the selected polynomial order M and the mean DIC evaluated over those models for the Scenario 3 and Scenario 5.

TABLE 2 Mean of DIC, mode of the power terms and posterior estimate of the effective number of parameters (p_D) obtained over 50 simulations for each scenario.

	Scenario 3		**Scenario 5**		
Random effects on	β_{0sj}	β_{1sj}	β_{0sj}	β_{1sj}	β_{2sj}
Power that achieves minimum DIC more often	$p = -1$	$p = -1$	$p_1 = -2$, $p_2 = 0$	$p_1 = -2$, $p_2 = 0$	$p_1 = -2$, $p_2 = 0$
Mean of DIC	49.946	39.710	47.676	44.668	49.976
p_D	21.8	21.9	27.0	28.8	26.0

For scenario 3 the "best" models are associated with $M = 1$ and thus we only report the estimated values for the power term p; for scenario 5, the model fitting the data best among those we have tested is associated with $M = 2$ and thus we report the estimates for both p_1 and p_2. In each scenario, we also report the results upon varying the parameter on which we place the random effect structures. From the inspection of Table 2 , it is possible to see that the DIC is consistently smaller for the models including a random effects on β_{1sj}.

3.3 | Results

To compare the results obtained with the three different NMA methods we display estimated profiles of each treatment from each study obtained for the third and the fifth scenario (the other figures are available in Supplementary Material, Section 1). Figures 3 and 4 show the estimated profiles obtained for the MTC, BEST-ITP and the appropriate FP models for Scenarios 3 and 5, respectively. Red lines indicates the true values used to simulate the data, whereas magenta, green and blue lines indicate the estimates obtained using the MTC model, BEST-ITP model and fractional polynomials model, respectively.

Upon visual inspection, for Scenarios 3 and 5 it appears as though more accurate estimates are obtained with the BEST-ITP and FP models. Moreover, it is evident that the credible intervals obtained from the MTC model are wider than those obtained from the FP and BEST-ITP counterparts. As displayed in Web Figure 1–5 of Supplementary Material, this result is confirmed also by the other Scenarios, with the only exception of Scenario 6. This is to be expected since for scenario 6 we have generated data using the model by Dakin *et al*. Finally, the the width of the credible intervals corresponding to BEST-ITP model increases over time, as it is also evident in Web Figure 1–4 of Supplementary Material. On the other hand, in Web Figure 5 the boundaries of the credible intervals obtained from the BEST-ITP model remain close to the true values over time. This is not surprising since for scenario 7 we have generated data from the model by Ding *et al*.

To better quantify our comparison, we use the mean squared error (MSE) as a measure of goodness of fit. If $\hat{Y}_{sjt}^{(l)}$ denotes the predicted value of the response for study s, treatment j time t and at MCMC iteration l and Y_{sjt} is the "true", simulated values,

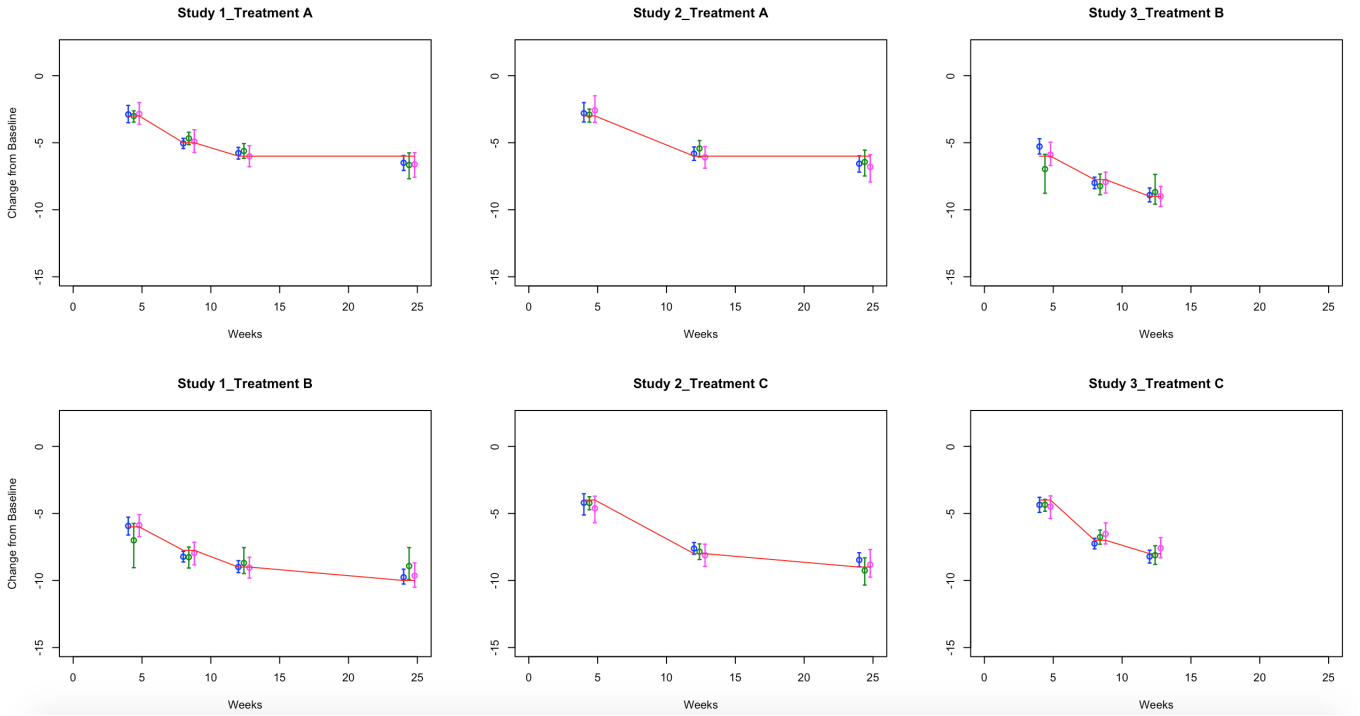

FIGURE 3 *Scenario 3*: posterior estimated profiles obtained. Red lines indicates the values used to simulate the data, whereas magenta lines, green lines and blue lines correspond to estimates obtained using the MTC model, BEST-ITP model and first-order fractional polynomials model respectively.

the MSE is defined as

$$\begin{aligned} \text{MSE} &= \frac{1}{R}\sum_{r=1}^{R}\left[\sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{s=1}^{S}\left(Y_{sjt}-\hat{Y}_{sjt}^{(r)}\right)^2\right] \\ &= \frac{1}{R}\sum_{r=1}^{R}\text{MSE}_r, \end{aligned}$$

where $R = 10000$ is the number of saved MCMC iterations. Lower MSE is associated with better performance.

TABLE 3 Comparison of the 3 NMA models by MSE: Mean (Standard Deviation) of the MSE. The number reported are an average over the 50 simulated datasets for each scenario.

	MTC model	BEST-ITP model	FP model
Scenario 1	8.6004 (0.5524)	1.3959 (0.4903)	0.3505 (0.0388)
Scenario 2	4.8586 (0.5886)	1.1089 (0.3873)	0.3505 (0.0118)
Scenario 3	3.1465 (0.3951)	1.9744 (0.3398)	1.3069 (0.1823)
Scenario 4	27.6757 (3.9365)	81.6728 (8.5083)	1.5747 (0.1338)
Scenario 5	43.3330 (1.4238)	104.2221 (2.6848)	41.7533 (1.8354)
Scenario 6	0.0162 (0.3923)	6.1175 (0.6988)	0.9198 (0.2864)
Scenario 7	2.70591 (0.3365)	0.0125 (0.2658)	0.1701 (0.2898)

The MSE results for each model and for all the simulation scenarios are summarized in Table 3 . We report the mean and the standard deviation of MSE across all 50 simulations. The mean squared error is lower for the FP models for all scenarios.

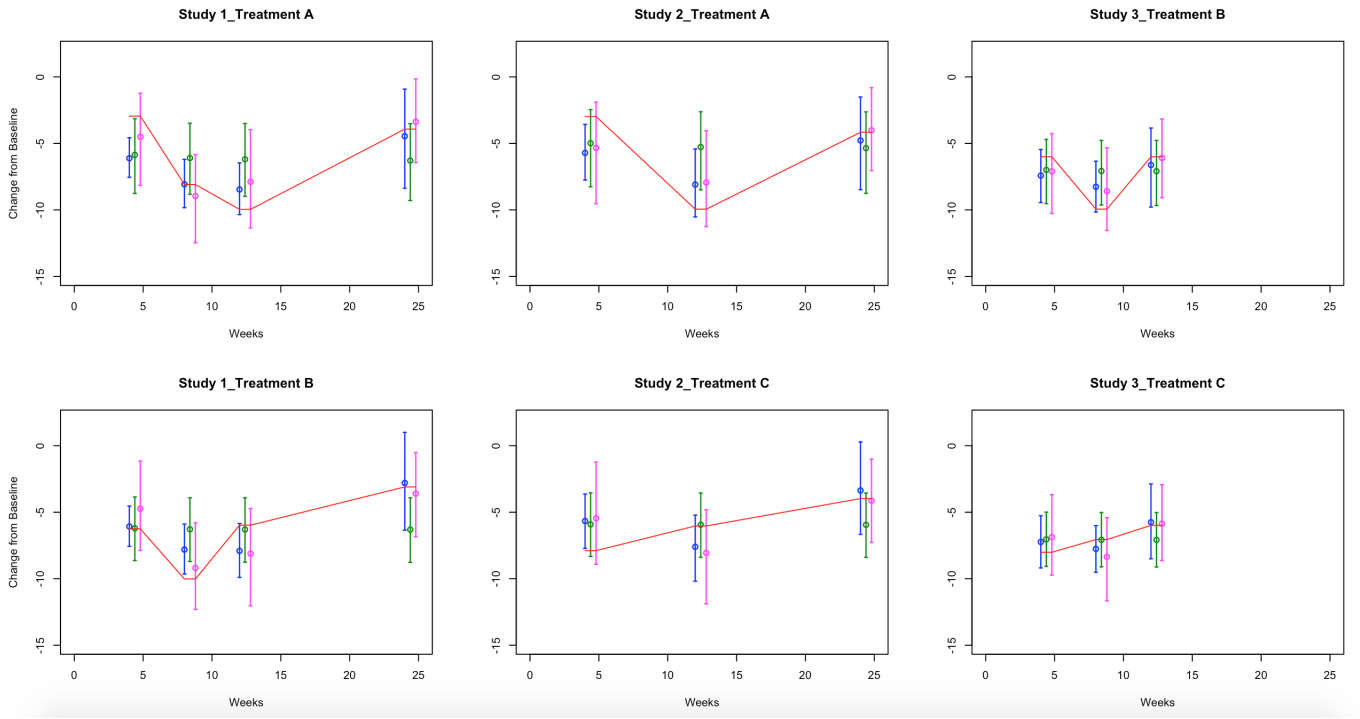

FIGURE 4 *Scenario 5*: posterior estimated profiles obtained. Red lines indicates the values used to simulate the data, whereas magenta lines, green lines and blue lines correspond to estimates obtained using the MTC model, BEST-ITP model and first-order fractional polynomials model respectively.

Moreover, for the first three scenarios, the BEST-ITP model is preferable to the MTC model, while the BEST-ITP model is not able to capture the time constant effect in the fourth scenario, leading to a worse MSE. In general the values of MSE for the Scenario 5 are large, implying that all the model are less able to capture non-monotonic time effects. Finally, in scenario 6 the smallest MSE is reached by the MTC model, while in the last scenario the best model is the BEST-ITP model. This is not unexpected since for scenario 6 we have generated the data from the model by Dakin *et al* and for scenario 7 we have generated data from the model by Ding *et al*.

3.4 | Binary outcomes

Often in longitudinal studies the outcome of interest is discrete. As such it is important to investigate the performance of the above methods when a binary outcomes is available.
Binary data, corresponding to scenario 3, are simulated by simply taking a logit transformation of the Scenario 3 of θ_{sjt}.
We fit the three network meta-analysis model, MTC model, BEST-ITP model and 1^{st} and 2^{nd} order fractional polynomial model (fixed and random effects models) model. The best fit (according to the DIC criterion) for the fractional polynomial model is obtained choosing a first order polynomial and treating β_0 as a fixed effect and β_1 as a random effect, with $p_1 = -2$. In Figure 5 we summarise the posterior distribution of p_{sjt} obtained from MTC model, BEST-ITP model and fractional polynomials. Red lines indicates the true values used to simulate the data, whereas magenta, green and blue lines indicate the estimates obtained using the MTC model, BEST-ITP model and fractional polynomials model, respectively.

From the Figure 5 it is clear that the more accurate estimates are obtained with the MTC and FP models. Moreover, it is evident that the precision of the estimates given by BEST-ITP model increases over time. The values of the MSE confirm the results obtained in the continuous case, as the largest MSE is obtained with the BEST-ITP model (0.0897), while the smallest with the model of Jansen *et al* (3) (0.0060).

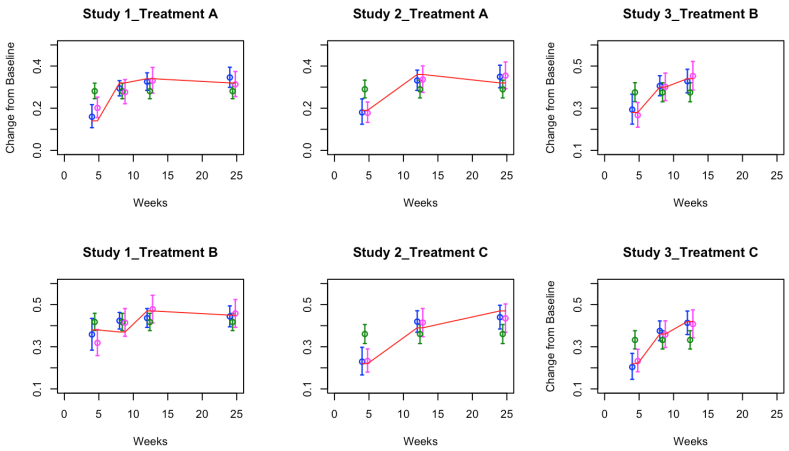

FIGURE 5 *Scenario 3: binary outcome*. Posterior distribution of p_{sjt}. Red lines indicates the values used to simulate the data, whereas magenta lines, green lines and blue lines correspond to estimates obtained using the MTC model, BEST-ITP model and first-order fractional polynomials model respectively.

4 | COPD DATA

We consider here the dataset described in Karabis *et al* (19). The data are related to patients affected by chronic obstructive pulmonary disease (COPD). Subjects can receive a combination of three possible treatments: aclidinium 400 μg BID (AB400), tiotropium 18 μg QD (TIO18) or Placebo. The data are collected from a total of $S = 13$ studies: 3 studies compare aclidinium 400 μg with placebo, 8 compare tiotropium 18 μg with placebo and one compares aclidinium 400 μg and tiotropium 18 μg with placebo. Figure 6 presents the network diagram, showing a total of 12 connections between the comparators. There is one closed loop, providing direct evidence. As all studies were placebo-controlled, placebo is used as the reference treatment. A brief overview of the study design and patient characteristics is presented in Table 4 . For the studies identified that met the selection criteria, the change from baseline (CFB) and the standard error (SE) are reported. We have no information about the loss to follow-up, so we have assumed that the sample sizes n_{sjt} remains the same as at the start of the study. To account properly for loss to follow-up we would need to introduce a new layer in the hierarchy which is beyond the scope of this review. If a study did not report the number of patients (indicated in the right-most column of Table 4 as "NR"), then the sample size has been generated from an uniform distribution ranging from the minimum and the maximum number of patients across all studies. Moreover, where information about the standard errors is not reported, Dakin *et al* (1) suggest to impute the missing standard deviations specifying the following prior distribution

$$\text{sd}_{sjt} \sim \text{Gamma}(\alpha_{se}, \beta_{se}), \qquad \alpha_{sd} \sim \text{Uniform}(0, 10) \qquad \text{and} \qquad \beta_{sd} \sim \text{Uniform}(0, 10).$$

We fit the MTC, BEST-ITP model and second order FP models (using both fixed and random effects models) to the COPD dataset and compare the results in the following.

4.1 | Results

Firstly, we need to select the powers p_1 and p_2 for the FP model. Given that the outcome of interest is reported as change from baseline, the constant term β_{0sj} is always zero and can be ignored. The best fit according to the DIC criterion is obtained treating β_1 as fixed effect and β_2 as a random effect, with resulting estimated posterior means of $p_1 = -0.5$ and $p_2 = 0$.

We present estimated profiles obtained with the three methods in Figure 7 . The solid lines indicate the results as mean differences in treatment effects over time for AB400 vs placebo, TIO18 vs placebo and AB400 vs TIO18 (red, blue and grey lines, respectively).

It appears clear from the inspection of the graphs in Figure 7 that aclidinium has similar effect to tiotropium but it is better than placebo the first 24 weeks. In fact, the grey lines, that represent the results as mean differences in treatment effects over time AB400 vs TIO18, are close to the value 0, whereas, the red lines and the blue lines, that indicate the results as mean differences in treatment effects over time for AB400 vs placebo and for TIO18 vs placebo have positive values. This results agree with

TABLE 4 COPD data: mean Change from baseline (CFB) in liters and corresponding SE at different time points.

Study	Time point (weeks)	CFB aclidinium (SE)	CFB tiotropium (SE)	CFB placebo (SE)	Number of patients
LAS 39	0.14	0.07 (0.016)	0.02 (0.017)	-0.07 (0.022)	NR
	6.00	0.03 (0.018)	-0.01 (0.018)	-0.11 (0.024)	
ATTAIN	1.00	0.09 (0.012)	-	-0.03 (0.012)	269; 273
	4.00	0.08 (0.013)	-	-0.03 (0.013)	
	8.00	0.09 (0.014)	-	-0.03 (0.014)	
	12.00	0.06 (0.015)	-	-0.05 (0.015)	
	18.00	0.08 (0.015)	-	-0.06 (0.015)	
	24.00	0.06 (0.016)	-	-0.07 (0.016)	
ACCORD I	1.00	0.10 (0.012)	-	-0.01 (0.012)	190; 187
	4.00	0.11 (0.014)	-	-0.01 (0.014)	
	8.00	0.10 (0.013)	-	-0.01 (0.014)	
	12.00	0.10 (0.015)	-	-0.02 (0.015)	
ACCORD II	1.00	0.07 (0.013)	-	-0.03 (0.012)	178; 182
	4.00	0.06 (0.014)	-	-0.01 (0.014)	
	8.00	0.08 (0.016)	-	0.00 (0.016)	
	12.00	0.06 (0.016)	-	-0.01 (0.015)	
Tashkin 2008*	4.33	-	0.10 (0.013)	0.01 (0.013)	2987; 3006
	26.00	-	0.10 (0.013)	0.01 (0.010)	
	52.00	-	0.09 (0.013)	-0.01 (0.013)	
Verkindre 2006*	6.00	-	0.15 (0.060)	0.01 (0.055)	46; 54
	12.00	-	0.10 (0.060)	-0.01 (0.055)	
Brusasco 2003*	2.14	-	0.10 (NR)	-0.03 (NR)	402; 400
	24.14	-	0.09 (NR)	-0.05 (NR)	
Casaburi 2000	1.14	-	0.11 (0.010)	-0.01 (0.010)	279; 191
	7.14	-	0.11 (0.010)	-0.01 (0.010)	
	13.14	-	0.11 (0.010)	-0.04 (0.010)	
Donohue 2002*	2.14	-	0.12 (NR)	-0.02 (NR)	209; 201
	24.14	-	0.11 (NR)	-0.04 (NR)	
Moita 2008*	6.00	-	0.05 (NR)	-0.02 (NR)	147; 164
	12.00	-	0.07 (NR)	-0.01 (NR)	
Covelli 2005	8.00	-	0.17 (0.023)	0.003 (0.025)	100: 96
	12.00	-	0.19 (0.025)	0.001 (0.027)	
Donohue 2010*	2.00	-	0.14 (0.011)	0.01 (0.009)	420; 425
	12.00	-	0.15 (0.011)	-0.01 (0.010)	
	26.00	-	0.13 (0.011)	-0.03 (0.011)	
Casaburi 2002*	1.14	-	0.12 (0.010)	-0.01 (0.017)	550;371
	13.14	-	0.12 (0.010)	-0.02 (0.017)	
	25.14	-	0.11 (0.010)	-0.04 (0.017)	
	49.14	-	0.11 (0.010)	-0.03 (0.017)	

the results provided by Karabis *et al* (19). Moreover, the posterior estimates obtained with fractional polynomials show a time pattern, by providing greater difference with increased time, while the results obtained with BEST-ITP model and MTC model are constant over time.

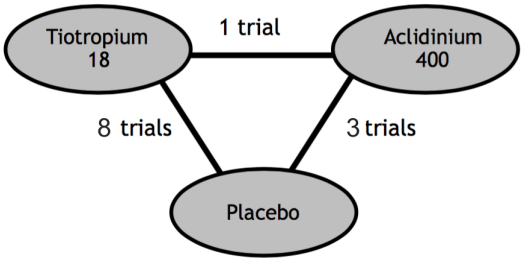

FIGURE 6 Network of studies. The different treatments are represented by circles. Two treatments are linked with a black lines if a direct comparison between them is available.

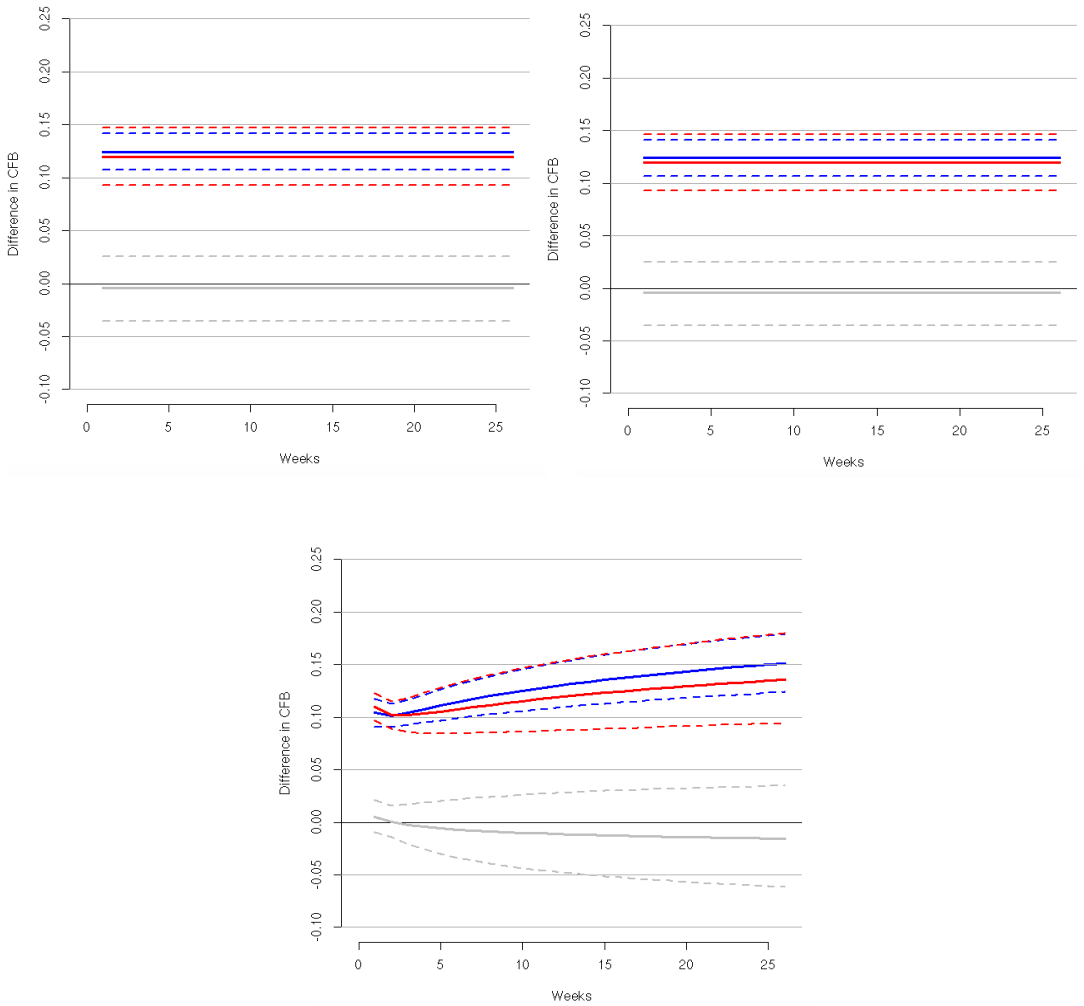

FIGURE 7 Results as mean differences in treatment effects over time for AB400vs placebo (red lines), TIO18 vs placebo (blue lines), AB400 vs TIO18 (grey lines) generated by MTC model (a), BEST-ITP model (b) and Fractional polynomials (c). Dashed lines denote 95% credible intervals (CI), while the solid lines represent the posterior mean.

5 | OSTEOARTHRITIS DATA

We also consider a second dataset described in Jansen *et al* (3). In this dataset there are 16 studies that investigate treatments for osteoarthritis (OA) of the knee based on different hyaluronan (HA)-based viscosupplements. OA is the most common chronic condition of the joints and it is a painful degenerative disease. In OA, the cartilage breaks down, causing pain, swelling and problems moving the joint. Due the effect of HA in synovial fluids, viscosupplementation with HA-based products is used to treat OA. Different treatment are considered: three, four, or five injections of HA with a molecular weight (MV) 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 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). Table 5 and Figure 8 reports an overview of the study design and patient characteristics for the selected studies. Also for this dataset, we use the sample sizes at the initial step across the time points, because the sample sizes at other time points are not available.

Study	Time point (weeks)	CFB (SEM) 5 HY-0.5-0.73	CFB (SEM) 4 HY-0.5-0.73	CFB (SEM) 3 HY-0.5-0.73	CFB (SEM) 3 HYGF20	CFB (SEM) 3 HY-0.62-11.7	CFB (SEM) 3 HY-2.4-3.6	CFB (SEM) placebo	Number of patients
Altman et al. (1998)(20)	1	-15.0 (3.1)	-	-	-	-	-	-15.0 (3.1)	165;165
	2	-19.0 (3.0)	-	-	-	-	-	-20.0 (3.1)	
	3	-23.0 (2.9)	-	-	-	-	-	-21.0 (3.1)	
	4	-27.0 (2.9)	-	-	-	-	-	-27.0 (3.1)	
	5	-29.0 (2.9)	-	-	-	-	-	-30.0 (3.0)	
	9	-30.0 (6.1)	-	-	-	-	-	-32.0 (6.1)	
	12	-31.0 (3.0)	-	-	-	-	-	-31.0 (3.0)	
	16	-33.0 (2.9)	-	-	-	-	-	-33.0 (3.0)	
	21	-37.0 (2.8)	-	-	-	-	-	-34.0 (2.9)	
	26	-36.0 (2.8)	-	-	-	-	-	-31.0 (3.1)	
Carrabba et al. (1995)(21)	2	-8.3 (4.0)	-	-6.0 (4.3)	-	-	-	-7.2 (3.3)	20;20;20
	4	-17.3 (4.2)	-	-14.6 (4.1)	-	-	-	-8.4 (3.3)	
	5	-21.5 (4.2)	-	-18.1 (4.6)	-	-	-	-8.0 (3.6)	
	8	-21.2 (3.9)	-	-16.3 (4.9)	-	-	-	-4.6 (3.8)	
Corrado et al. (1995)(22)	5	-33.3 (7.5)	-	-	-	-	-	-20.7 (7.2)	20;20
	8	-39.0 (6.5)	-	-	-	-	-	-19.1 (6.7)	
Cubukcu et al. (2005)(23)	1	-	-	-	-6.3 (2.0)	-	-	-9.0 (3.9)	30;10
	2	-	-	-	-18.3 (2.4)	-	-	-13.0 (3.7)	
	3	-	-	-	-24.3 (2.5)	-	-	-16.2 (2.8)	
	8	-	-	-	-31.0 (2.4)	-	-	-13.2 (3.1)	
Dougados et al. (1993)(24)	2	-	-20.0 (3.0)	-	-	-	-	-15.0 (2.0)	55;55
	3	-	-31.0 (4.0)	-	-	-	-	-25.0 (3.0)	
	4	-	-23.0 (5.0)	-	-	-	-	-26.0 (3.0)	
	7	-	-35.5 (3.6)	-	-	-	-	-25.8 (2.9)	
	52	-	-38.9 (4.2)	-	-	-	-	-32.7 (3.9)	
Henderson et al. (1994)(25)	1	-12.5 (2.3)	-	-	-	-	-	-16.1 (2.0)	45;46
Huskisson et Donnelly (1999)(26)	5	-38.3 (4.1)	-	-	-	-	-	-21.3 (5.3)	50;50
	8	-33.5 (4.5)	-	-	-	-	-	-19.8 (5.3)	
	16	-32.8 (4.9)	-	-	-	-	-	-13.6 (5.5)	
	26	-26.4 (4.7)	-	-	-	-	-	-8.2 (5.3)	
Jubb et al. (2003)(27)	4	-	-	-12.5 (1.9)	-	-	-	-11.0 (1.9)	204;204
	11	-	-	-10.0 (2.1)	-	-	-	-5.0 (1.9)	
	18	-	-	-9.0 (2.2)	-	-	-	-10.0 (1.9)	
Karlsson et al. (2002)(28)	1	-	-	-	-7.0 (2.0)	-5.0 (2.0)	-	-7.0 (3.2)	70;70;70
	2	-	-	-	-16.0 (2.5)	-12.0 (2.6)	-	-11.0 (3.6)	
	3	-	-	-	-18.0 (2.9)	-20.0 (2.8)	-	-21.0 (4.0)	
	12	-	-	-	-22.0 (3.5)	-22.0 (3.2)	-	-19.0 (4.6)	
	20	-	-	-	-27.0 (3.5)	-21.0 (3.2)	-	-16.0 (3.8)	
	26	-	-	-	-20.0 (3.7)	-16.0 (3.8)	-	-21.0 (4.5)	
Scale et al. (1994)(29)	1	-	-	-	-9.5 (5.8)	-	-	-2.5 (4.2)	40;40
	2	-	-	-	-32.0 (5.8)	-	-	-21.0 (4.2)	
	3	-	-	-	-44.0 (5.0)	-	-	-20.0 (4.2)	
	8	-	-	-	-52.0 (5.8)	-	-	-19.0 (4.2)	
	12	-	-	-	-53.0 (3.6)	-	-	-17.0 (3.6)	
Wobig et al. (1998)(30)	1	-	-	-	-12.0 (4.6)	-	-	-5.0 (4.2)	57;60
	2	-	-	-	-26.0 (4.6)	-	-	-16.0 (4.2)	
	3	-	-	-	-39.0 (4.6)	-	-	-22.0 (4.2)	
	8	-	-	-	-46.0 (4.6)	-	-	-15.0 (4.2)	
	12	-	-	-	-46.0 (4.6)	-	-	-12.0 (4.2)	
Kirchner et al. (2006)(31)	12	-	-	-	-28.7 (2.0)	-	-31.2 (2.0)	-	158;156
Stitik et al. (2007)(32)	1	-5.6 (7.4)	-	-0.9 (5.2)	-	-	-	-	20;20

	2	-15.2 (8.0)	-	-5.7 (5.9)	-	-	-	-	
	5	-20.5 (8.0)	-	-13.2 (5.9)	-	-	-	-	
	12	-24.3 (8.0)	-	-23.2 (5.9)	-	-	-	-	
	26	-23.3 (8.0)	-	-21.8 (5.9)	-	-	-	-	
	36	-26.3 (8.0)	-	-10.4 (5.9)	-	-	-	-	
	52	-35.4 (8.0)	-	-16.1 (5.9)	-	-	-	-	
Wobig et al. (1999)(33)	1	-	-	-	-8.0 (3.6)	-10.0 (4.5)	-	-	38;35
	2	-	-	-	-18.0 (3.6)	-20.0 (4.5)	-	-	
	3	-	-	-	-29.0 (4.2)	-30.0 (4.5)	-	-	
	8	-	-	-	-37.0 (4.2)	-30.0 (4.5)	-	-	
	12	-	-	-	-37.0 (4.2)	-30.0 (4.5)	-	-	
Altman et al. (2009)(34)	1	-	-	-	-	-	-13.5 (1.0)	-10.0 (1.0)	293;295
	2	-	-	-	-	-	-20.0 (1.0)	-17.5 (1.0)	
	3	-	-	-	-	-	-24.0 (1.0)	-22.5 (1.0)	
	7	-	-	-	-	-	-25.0 (1.0)	-22.0 (1.0)	
	12	-	-	-	-	-	-25.5 (1.0)	-22.0 (1.0)	
	18	-	-	-	-	-	-27.5 (1.0)	-21.0 (1.0)	
	26	-	-	-	-	-	-25.0 (1.0)	-17.5 (1.0)	
Dickson et al. (2001)(35)	12	-	-	-	-39.0 (4.0)	-	-26.0 (4.0)	-	NR

TABLE 5 OA dataset: mean CFB (Change from baseline) in liters and corresponding standard error (SE) at different time points..

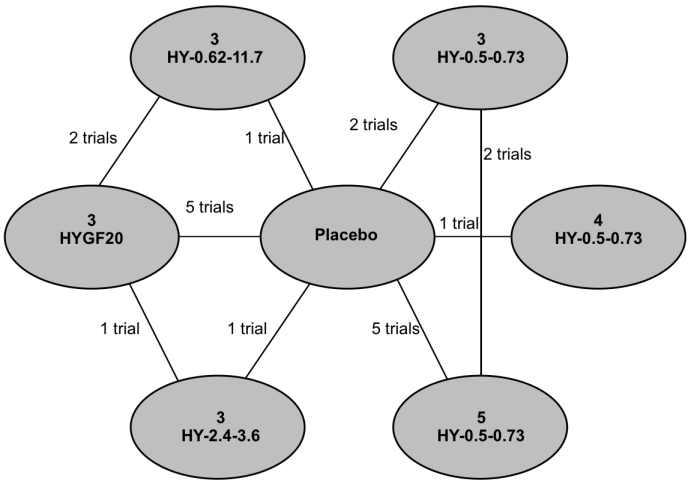

FIGURE 8 Network of studies. The different treatments are represented by circles. Two treatments are linked with a black lines if a direct comparison between them is available.

5.1 | Results

Also in this example, we apply the MTC model, BEST-ITP model and 1^{st} and 2^{nd} order fractional polynomial model (fixed and random effects models)(FP) models to perform a network meta-analysis of OA studies. In our analysis placebo is treated as reference treatment. The best fit for fractional polynomial model according to DIC is obtained using a second order polynomial and treating β_1 as a fixed effect and β_2 as a random effect, with p_1=0.5 and p_2=1.

In Figure 9 posterior estimated profiles obtained from the three models are shown. We report the posterior mean of the treatment effect relative to placebo over time for 5 HY-0.5-0.73 (magenta lines), 4 HY-0.5-0.73 (black lines), 3 HY-0.5-0.73 (red lines), 3 HYGF20 (green lines), 3 HY-0.62-11.7 (blue lines) and 3 HY-2.4-3.6 (light blue lines).

All treatments are better than placebo, consistently over the first 24 weeks. In particular, the best result are obtained with 3 HYGF20. Moreover, the estimated treatment effects obtained from the fractional polynomials model change in time; for example for the first few weeks receiving 5 injections of HY-0.5-0.73 is better than 4 injections of the same drug, while later on we find that the opposite is true. This type of change is not detected by the MTC model and the BEST-ITP model.

6 | CONCLUSIONS

In this paper, we review the MTC, BEST-ITP and fractional polynomials models for Network Meta-analysis, highlighting the main features of the different methods. The model presented by Dakin *et al* (1) assumes that relative treatment effects may vary over time. Ding *et al* (2) suggested a parametric model for the relationship between outcome and time for each treatment. Finally, Jansen *et al* (3) suggested a more flexible approach where the relationship between outcome and time is a fractional polynomial. The three NMA models make different assumptions on the behaviour of the treatment effect over time. The MTC model assumes that the treatment effect is constant over time as

$$\begin{aligned} \theta_{sjt} &= \mu_{st} + \delta_{sj} \\ \delta_{sj} &\sim \text{Normal}\left(d_j - d_1, \sigma_\delta^2\right) \end{aligned}$$

The Best-ITP model assumes that the treatment effect over time is given by

$$(d_j - d_1)\left(\frac{1 - e^{p_j t}}{1 - e^{p_j T}}\right)$$

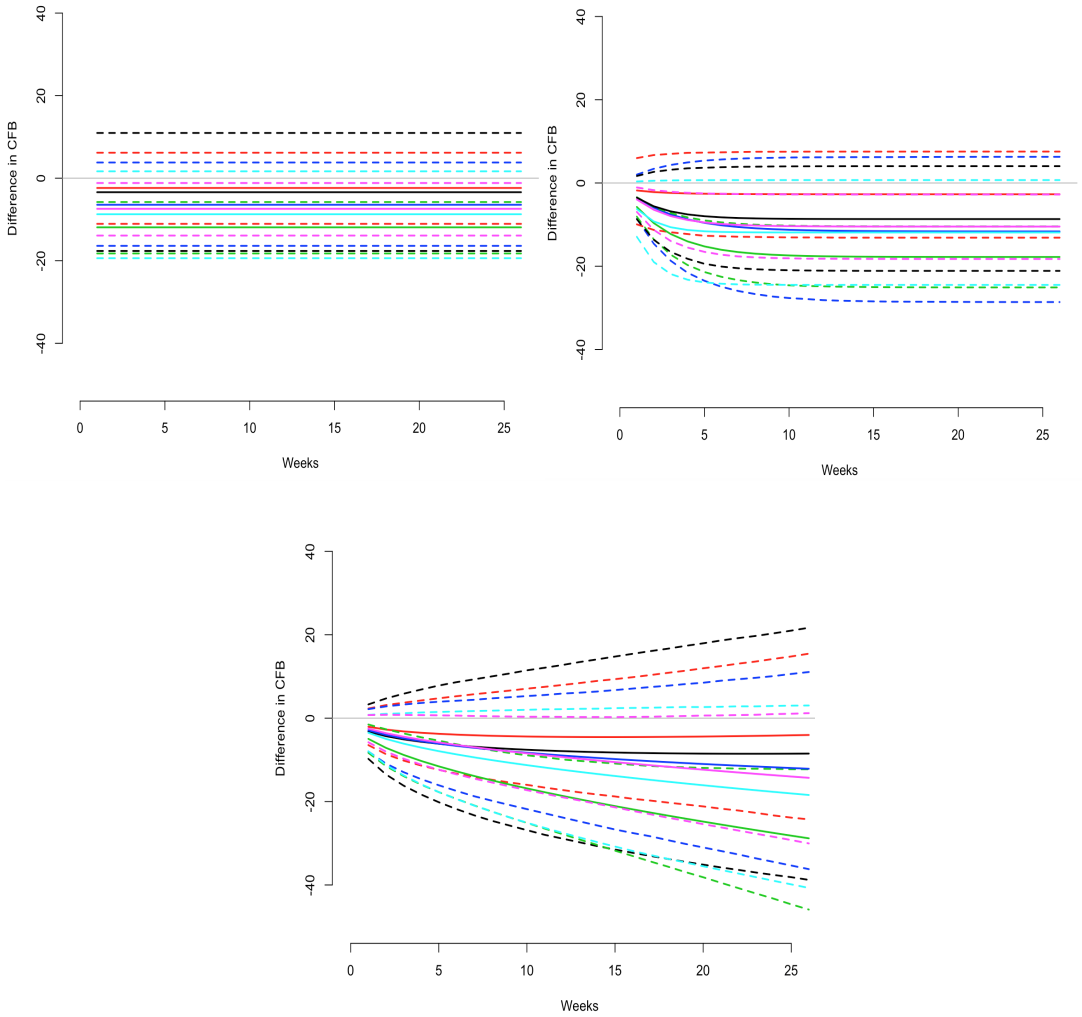

FIGURE 9 Posterior mean of the different treatment effects relative to placebo over time for 5 HY-0.5-0.73 (magenta lines), 4 HY-0.5-0.73 (black lines), 3 HY-0.5-0.73 (red lines), 3 HYGF20 (green lines), 3 HY-0.62-11.7 (blue lines) and 3 HY-2.4-3.6 (light blue lines) generated by MTC model (a), BEST-ITP model (b) and Fractional polynomials (c). Dashed lines denote 95% credible intervals (CI), while solid lines represent the posterior mean.

which admits a constant effect when p_j is close to zero. Finally, when employing fractional polynomials, we are making the assumption that the treatment effect can be described by a parametric function of time. This allows borrowing information across time and for a variety of temporal patterns.

We evaluated the performance of the models in a a simulation study and on real data applications. Based on the results of our simulations, Daikin's model appears to be the most conservative in terms of estimation of the underlying effect-size, while fractional polynomial seems to offer the most flexible strategy, able to accommodate for different time patterns. The BEST-ITP model can not capture the constant course, i.e. an effect size that does not change over time. In general, it is more difficult to capture the non-monotonic temporal patterns, with the Fractional Polynomial model leading to slightly better estimates.

In terms of computational time, there are no substantial differences in running times between the three different models, although fractional polynomials require extensive sensitivity analysis to select the optimal order and the power terms of the polynomials.

Acknowledgements

Dr Gianluca Baio is partially supported as the recipient of an unrestricted research grant sponsored by Mapi Group at University College London. We would like to thank Andreas Karabis for useful suggestions and comments on the work and for providing the COPD data.

References

[1] Dakin HA, Welton NJ, Ades AE, 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. *Statistics in medicine.* 2011;30(20):2511–2535.

[2] Ding Y, Fu H. Bayesian indirect and mixed treatment comparisons across longitudinal time points. *Statistics in medicine.* 2013;32(15):2613–2628.

[3] Jansen JP, Vieira MC, Cope S. Network meta-analysis of longitudinal data using fractional polynomials. *Statistics in medicine.* 2015;34(15):2294–2311.

[4] Borenstein Michael, Hedges Larry V, Higgins Julian PT, Rothstein Hannah R. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research synthesis methods.* 2010;1(2):97–111.

[5] Higgins Julian, Thompson Simon G, Spiegelhalter David J. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society).* 2009;172(1):137–159.

[6] Lipsey Mark W, Wilson David B. *Practical meta-analysis.* Sage Publications, Inc; 2001.

[7] Welton N J, Sutton A J, Cooper N J, Abrams K R, Ades A E. *Evidence Synthesis for Decision Making in Healthcare*. John Wiley & Sons, Ltd; 2012.

[8] Bucher Heiner C., Guyatt Gordon H., Griffith Lauren E., Walter Stephen D.. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of Clinical Epidemiology.* 1997;50(6):683 - 691.

[9] Lumley Thomas. Network meta-analysis for indirect treatment comparisons. *Statistics in medicine.* 2002;21(16):2313–2324.

[10] Cooper Harris, Hedges Larry V, Valentine Jeffrey C. *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation; 2009.

[11] Wandel Simon, Jüni Peter, Tendal Britta, et al. Effects of glucosamine, chondroitin, or placebo in patients with osteoarthritis of hip or knee: network meta-analysis. *Bmj.* 2010;341:c4675.

[12] Lu G, Ades AE, Sutton AJ, Cooper NJ, Briggs AH, Caldwell DM. Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Statistics in Medicine.* 2007;26(20):3681–3699.

[13] Adalsteinsson Erpur, Toumi Mondher. Benefits of probabilistic sensitivity analysis–a review of NICE decisions. *Journal of market access & health policy.* 2013;1(1):21240.

[14] Berger James O. *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media; 2013.

[15] Plummer M. *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.* 2003.

[16] Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statistical science.* 1992;:457–472.

[17] Royston Patrick, Sauerbrei Willi. *Multivariable model-building: a pragmatic approach to regression anaylsis based on fractional polynomials for modelling continuous variables*. John Wiley & Sons; 2008.

[18] Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* 2002;64(4):583–639.

[19] 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. *International Journal of COPD.* 2013;8:405–423.

[20] Altman RD. Intra-articular sodium hyaluronate (Hyalgan®) in. *J Rheumatol.* 1998;25:2203–12.

[21] Carrabba M, Paresce E, Angelini M, Re KA, Torchiana EEM, Perbellini A. The safety and efficacy of different dose schedules of hyaluronic acid in the treatment of painful osteoarthritis of the knee with joint effusion. *European journal of rheumatology and inflammation.* 1995;15(1):25–31.

[22] Corrado EM, Peluso GF, Gigliotti S, et al. The effects of intra-articular administration of hyaluronic acid on osteoarthritis of the knee: a clinical study with immunological and biochemical evaluations. *European journal of rheumatology and inflammation.* 1995;15(1):47–56.

[23] Cubukcu D, Ardıç F, Karabulut N, Topuz O. Hylan GF 20 efficacy on articular cartilage quality in patients with knee osteoarthritis: clinical and MRI assessment. *Clinical rheumatology.* 2005;24(4):336–341.

[24] Dougados M, Nguyen M, Listrat V, Amor B. High molecular weight sodium hyaluronate (hyalectin) in osteoarthritis of the knee: a 1 year placebo-controlled trial. *Osteoarthritis and cartilage.* 1993;1(2):97–103.

[25] Henderson EB, Smith EC, Pegley F, Blake DR. Intra-articular injections of 750 kD hyaluronan in the treatment of osteoarthritis: a randomised single centre double-blind placebo-controlled trial of 91 patients demonstrating lack of efficacy.. *Annals of the rheumatic diseases.* 1994;53(8):529–534.

[26] Huskisson EC, Donnelly S. Hyaluronic acid in the treatment of osteoarthritis of the knee.. *Rheumatology.* 1999;38(7):602–607.

[27] Jubb RW, Piva S, Beinat L, Dacre J, Gishen P. A one-year, randomised, placebo (saline) controlled clinical trial of 500-730 kDa sodium hyaluronate (Hyalgan) on the radiological change in osteoarthritis of the knee.. *International journal of clinical practice.* 2002;57(6):467–474.

[28] Karlsson J, Sjögren LS, Lohmander LS. Comparison of two hyaluronan drugs and placebo in patients with knee osteoarthritis. A controlled, randomized, double-blind, parallel-design multicentre study. *Rheumatology.* 2002;41(11):1240–1248.

[29] Scale D, Wobig M, Wolpert W. Viscosupplementation of osteoarthritic knees with hylan: a treatment schedule study. *Current Therapeutic Research.* 1994;55(3):220–232.

[30] Wobig M, Dickhut R, Vetter G. Viscosupplementation with hylan GF 20: a 26-week controlled trial of efficacy and safety in the osteoarthritic knee. *Clinical therapeutics.* 1998;20(3):410–423.

[31] Kirchner M, Marshall D. A double-blind randomized controlled trial comparing alternate forms of high molecular weight hyaluronan for the treatment of osteoarthritis of the knee. *Osteoarthritis and cartilage.* 2006;14(2):154–162.

[32] Stitik TP, Blacksin MF, Stiskal DM, et al. Efficacy and safety of hyaluronan treatment in combination therapy with home exercise for knee osteoarthritis pain. *Archives of physical medicine and rehabilitation.* 2007;88(2):135–141.

[33] Wobig M, Bach G, Beks P, et al. The role of elastoviscosity in the efficacy of viscosupplementation for osteoarthritis of the knee: a comparison of hylan GF 20 and a lower-molecular-weight hyaluronan. *Clinical therapeutics.* 1999;21(9):1549–1562.

[34] Altman RD, Rosen JE, Bloch DA, Hatoum HT, Korner P. A double-blind, randomized, saline-controlled study of the efficacy and safety of EUFLEXXA® for treatment of painful osteoarthritis of the knee, with an open-label safety extension (The FLEXX Trial). In: :1–9Elsevier; 2009.

[35] Dickson DJ, Hosie G, English JR. A double-blind, placebo-controlled comparison of hylan GF 20 against diclofenac in knee osteoarthritis. *JOURNAL OF DRUG ASSESSMENT.* 2001;4(3):179–190.

How cite this article: Tallarita M., M. De Iorio, A. Karabis, and G. Baio (2017), A comparative review of network meta-analysis models in longitudinal studies, *Statistics in Medicine*, *2017;00:1–6*.