Statistics in Medicine

WILEY

Statistics in Medicine

RESEARCH ARTICLE OPEN ACCESS

Bayesian Network Meta-Analysis With One or Two Continuous Outcomes Measured at Multiple Time Points Using Gaussian Random Walks With Drift

Pai-Shan Cheng[1] | Bruno R. da Costa[2] | George Tomlinson[1,3]

[1]Biostatistics Division, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada | [2]Clinical Trial Service Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK | [3]Department of Medicine, University Health Network/Mt Sinai Hospital, Toronto, Ontario, Canada

Correspondence: Pai-Shan Cheng (paishan.cheng@alumni.utoronto.ca)

Received: 27 March 2023 | **Revised:** 23 November 2025 | **Accepted:** 9 December 2025

Keywords: cointegration | longitudinal data | network meta-analysis | osteoarthritis | random walk with drift

ABSTRACT

Network meta-analysis of randomized controlled trials is traditionally conducted on a single outcome measured at one time point. However, many trials also feature a secondary outcome and both outcomes may have been reported at multiple time points. Existing network meta-analysis methods for synthesizing continuous outcome data from such trials focus on either the longitudinal data aspect or the multiple outcomes aspect, but not on both simultaneously. In this paper, we present two Bayesian network meta-analysis models that account for the correlation of outcome measurements over time using Gaussian random walks with drift. The first model is suitable for a single continuous outcome measured at multiple time points, while the second model extends the first model to allow incorporation of a second outcome through cointegration of random walks. A simulation study to evaluate several statistical properties of these models is conducted. The results indicate that both proposed models produce unbiased estimates of relative treatment effect and drift parameters, as well as reasonable coverage. Furthermore, in some scenarios, using the cointegration model yields small gains in precision over using the single outcome model. Based on various performance measures, both proposed models also outperform an existing random walk network meta-analysis model previously used by investigators to synthesize osteoarthritis trials data. The proposed models are illustrated with an application to trials evaluating treatments for knee and hip osteoarthritis. Both models are useful additions to existing tools available to investigators undertaking a network meta-analysis of continuous outcome data at multiple time points.

1 | Introduction

Randomized controlled trials (RCTs) are seen as the gold standard in generating evidence for comparative effectiveness research, although a single RCT seldom provides sufficient evidence to answer research questions concerning multiple treatments of interest in a target population, such as in the context of clinical guideline development. In many cases, a collection of RCTs involving more than two treatments, identified ideally through a systematic review, does not compare the same treatments across all trials. Instead, each RCT in the collection compares only a smaller subset of all treatments of interest, and thus provides only a piece of the evidence. The totality of evidence from the collection of RCTs can be represented as a network of treatment nodes, with edges between nodes denoting that one or more trials involve direct comparisons of the two treatments connected by each edge. To synthesize quantitatively the many pieces of evidence in this network, the investigator undertaking the

systematic review can make use of network meta-analysis (NMA) methods which employ statistical models to combine all available data in a coherent and internally consistent manner.

In the last two decades, numerous sophisticated methods have been developed that greatly expanded the toolkit of the investigator interested in conducting a NMA. One area of NMA that has seen significant development is Bayesian methods for synthesizing data measured at multiple time points from RCTs. Traditionally, meta-analyses of longitudinal data from RCTs either focused on a single convenient time point or involved a separate meta-analysis at each time point of interest. However, neither of these approaches account for the correlation of treatment effects between time points. Thus, NMA methods that allow data from multiple time points to be synthesized simultaneously, while accounting for the correlation across time points, have garnered the attention of methodologists. Lu et al. proposed a series of Bayesian hierarchical models for conducting NMA of trials with an outcome measured at more than one follow-up time and with different trials reporting the outcome at possibly different times [1]. These Bayesian models are specified using a piecewise constant hazard model in the context of pooling log hazard ratios comparing healing rates of treatments for a disease and include a random walk model on the log hazard ratios, but they can only be applied in cases where the event of interest can only occur once during the follow-up. Dakin et al. presented several fixed and random effects Bayesian models for conducting NMA of repeated measurements of a continuous outcome with assumptions of constant or time interval-specific relative effects and constant or unconstrained baseline effects [2]. However, many of these models do not provide insights into how the outcome changes over the course of treatment nor how these temporal trends compare across different treatments. Jansen et al. proposed NMA models using fractional polynomials to model the dependence of treatment effects on time [3]. Although fractional polynomials are flexible parametric functions to describe the development of outcomes over time, this flexibility also carries with it certain implications in practice. For instance, for only a second-degree fractional polynomial with two powers p_1 and p_2, and restricting potential values of p_1 and p_2 to the set $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$, there are at least 36 different fractional polynomial NMA models (not accounting for other factors such as fixed versus random effects models) that would need to be fitted to identify the most parsimonious model through comparisons of the deviance information criterion [4]. More recently, da Costa et al. introduced a Bayesian random walk NMA model in the context of synthesizing continuous outcome data from osteoarthritis (OA) RCTs [5]. This NMA model makes use of Gaussian random walks to handle correlations between mean standardized effects across time points, though the random walks are implicitly assumed to have zero drift in this model. Still more recently, Pedder et al. proposed a Bayesian model-based NMA framework for modelling a continuous summary outcome at multiple follow-up times using nonlinear multi-parameter time-course functions [6]. A limitation of this method is that, for models with multiple parameters, a sufficiently rich dataset is required to make all non-nuisance parameters identifiable. Otherwise, additional biologically plausible simplifying assumptions need to be made to constrain these parameters.

Another area of NMA that has seen significant development, though not as extensively as NMA methods for longitudinal data, is Bayesian methods for synthesizing data on multiple outcomes. Traditionally, meta-analyses of RCTs with multiple outcomes involved a separate meta-analysis for each outcome. However, in many RCTs, different outcomes are often correlated with each other, and the traditional approach of a separate meta-analysis for each outcome would ignore such correlation between outcomes. Furthermore, under the traditional approach, several RCTs may need to be excluded from the evidence base if an outcome of interest is not reported for these trials. This can lead to biased estimates of relative treatment effects. To remedy these issues, methods have been developed that enable the borrowing of strength from other correlated outcomes. Efthimiou et al. proposed two bivariate Bayesian models for conducting NMA of two correlated binary, continuous, or time-to-event outcomes [7]. One model involves simplifying both the within- and between-study covariance matrices to reduce the number of parameters and ease computational burden, while the other model involves extending a bivariate meta-analysis model by Riley et al. that uses a single correlation coefficient, an amalgam of within- and between-study correlations, to account for the overall correlation [8]. In a more general manner, Achana et al. developed a multivariate hierarchical NMA model for multiple outcomes, involving both within- and between-study models to account for correlations between treatment effects on different outcomes [9]. Achana et al. also proposed an extension to this multivariate model by making an assumption of constant relative potency of treatments across outcomes [9], thus allowing borrowing of strength across treatments and outcomes. However, with this extension, there are limitations on the number of data points on outcomes allowed to be missing. A more recent multivariate NMA model proposed by Waddingham et al. exploits relationships between outcomes by using structural links in the form of proportional mappings between mean treatment effect parameters [10]. This approach, however, is not suitable for scenarios in which the relative effect may be zero for some outcomes but nonzero for other outcomes.

In this paper, we present two Bayesian NMA models for continuous outcome data measured at multiple time points. The two models were developed with the aim of synthesizing data from OA RCTs. OA is a common cause of disability and is characterized by progressive deterioration of articular cartilage and reactive bone changes [11]. Clinically, OA is associated with regional pain, dysfunction, and stiffness of the affected joint [11]. In OA trials, pain intensity and physical function are two outcomes that are often measured at multiple time points using various different algofunctional scales across different trials. To account for the use of different algofunctional scales, we model standardized effects on pain and physical disability reduction as the two outcomes of interest in this paper. The NMA models we present, however, are not limited to modelling only standardized effects. The first of our two models uses a Gaussian random walk to model mean standardized effects over multiple time points for each treatment in a trial. We additionally introduce drift terms to the Gaussian random walks to account for potential time trends in mean standardized effects. The second model is an extension of the first and allows the inclusion of a second outcome in the

model. Though not strictly a bivariate NMA model in the traditional sense, the second model we present still enables borrowing of strength across outcomes while maintaining the ability to account for the correlation of outcome measurements across time points. To our knowledge, this is the first Bayesian NMA model for continuous outcome data to account for both the longitudinal data and the multiple outcomes aspects in a single model.

For the remainder of this paper, we first briefly describe an example dataset containing OA trials data. Then, we present details on the two novel Bayesian NMA models using Gaussian random walks with drift. The presentation of our models is followed by a simulation study which aims to evaluate various statistical properties of these models and demonstrate that these models yield inferences with good frequentist properties. The details of the simulation study, including the data-generating mechanism and performance measures used, are then described. These are followed by the results of the simulation study, which include comparisons of the proposed models with the random walk NMA model of da Costa et al. [5]. The results for the motivating example are then presented, followed by a discussion of key considerations concerning the use of our proposed NMA models.

2 | Motivating Example

The example datasets come from a previous network meta-analysis of RCTs evaluating non-steroidal anti-inflammatory drugs (NSAIDs), paracetamol, and placebo as treatments for knee and hip OA [5]. A systematic search of the Cochrane Central Register of Controlled Trials and other databases for eligible trials was conducted by the authors yielding 76 large randomized trials with at least 100 patients per arm. The prespecified primary outcome for the NMA was pain and the secondary outcome was physical function. Data on the pain outcome were extracted at the following time points whenever available: 1 week (±2 days), 2 weeks (±2 days), 4 weeks (±1 week), 6 weeks (±1 week), 3 months (±1 month), 6 months (±1 month), and 12 months (±1 month). Data on the function outcome were extracted at only a subset of these time points. If data on more than one pain scale were reported for a trial, the authors extracted data for the pain scale that was highest on the following list: (1) global pain score; (2) pain on walking; (3) Western Ontario and McMaster Universities (WOMAC) osteoarthritis index pain subscore; (4) composite pain scores other than WOMAC; (5) pain on activities other than walking; (6) WOMAC global score; (7) Lequesne osteoarthritis index global score; (8) other algofunctional composite scores; (9) patient's global assessment; (10) physician's global assessment. Similarly, if data on more than one physical function scale were reported for a trial, the authors extracted data for the physical function scale that was highest on the following list: (1) global function score; (2) walking disability; (3) WOMAC osteoarthritis index function subscore; (4) composite physical function scores other than WOMAC; (5) physical function on activities other than walking; (6) WOMAC global score; (7) Lequesne osteoarthritis index global score; (8) other algofunctional composite scores; (9) patient's global assessment; (10) physician's global assessment. Due to the variety of algofunctional scales used for each outcome in different trials, the authors computed standardized effects by dividing effect measures in each trial by the median pooled standard deviation across all time points in the trial. The resulting datasets are included in a Supporting Information data file with this paper. We applied the two novel models developed in this paper to the function outcome and used the pain outcome only in an auxiliary manner to aid estimation of parameters for the function outcome. We also applied the random walk NMA model of da Costa et al. using time point one as the reference time point (whereas da Costa et al. used time point four as reference in the original publication) and compared the results with those obtained from fitting our proposed models. Since some of the RCTs included in the original NMA did not report function outcome measurements at any of the relevant time points, the actual number of trials included in the example NMA was reduced to 70 RCTs. These RCTs form the network of 22 treatments shown in Figure 1. Placebo was chosen as treatment 1, the reference treatment, as it had been trialled against the greatest number of other treatments.

3 | Methods

In this section, we first present a Bayesian Gaussian random walk with drift NMA model for a single continuous outcome inspired by the random walk NMA model used by da Costa et al. [5] to conduct a NMA of OA trials comparing NSAIDs, paracetamol, and placebo. The idea of imposing a random walk structure on some measure of treatment effect in the context of conducting a NMA, however, was first presented by Lu et al. [1]. The random walk NMA models discussed in these papers essentially allow information on treatment effects from other time points to be propagated through time to inform estimation of pooled effects associated with a selected reference time point. Although these models are useful because they enable the borrowing of information across time points, they are not designed to allow estimation of pooled effects at any time point other than the reference time point. This motivates the development of model 1 in this paper, which uses Gaussian random walks with drift to capture information about average time trends in treatment effects, thus allowing pooled effects to be estimated at multiple time points. Since physical function is another outcome that is often featured in OA RCTs in addition to pain, we follow up the presentation of model 1 with model 2, an extension of model 1 that allows two outcomes to be incorporated into a single model by considering these outcomes as two random walks exhibiting cointegration. Then, we describe the simulation study in detail, including the data-generating mechanism for simulating the RCT data and the settings of various parameters in different simulation scenarios. We conclude the methods section by defining the performance measures used in the simulation study.

3.1 | Model 1: Gaussian Random Walk With Drift Network Meta-Analysis Model

Model 1 is specified for a single outcome assuming that the observed standardized effect of treatment k in trial i at time point t, y_{ikt}, follows a Gaussian distribution centred at the mean standardized effect of treatment k in trial i at time point t, θ_{ikt}, with variance σ^2_{ikt}:

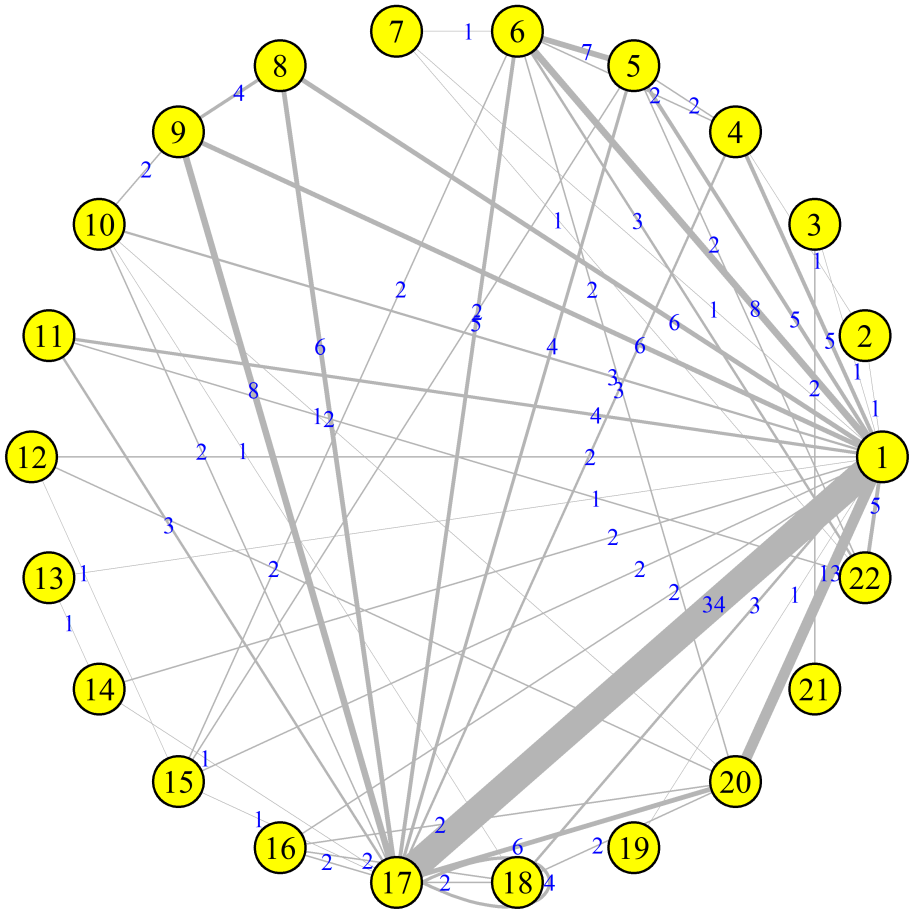

FIGURE 1 | Network diagram showing treatment comparisons in the example network meta-analysis. The number of trials featuring each treatment comparison is shown on each edge; the width of each edge also scales with the number of trials. Treatment labels are: 1 = placebo, 2 = paracetamol < 2000 mg, 3 = paracetamol 3000 mg, 4 = paracetamol 3900–4000 mg, 5 = rofecoxib 12.5 mg, 6 = rofecoxib 25 mg, 7 = rofecoxib 50 mg, 8 = lumiracoxib 100 mg, 9 = lumiracoxib 200 mg, 10 = lumiracoxib 400 mg, 11 = etoricoxib 30 mg, 12 = etoricoxib 60 mg, 13 = diclofenac 70 mg, 14 = diclofenac 100 mg, 15 = diclofenac 150 mg, 16 = celecoxib 100 mg, 17 = celecoxib 200 mg, 18 = celecoxib 400 mg, 19 = naproxen 750 mg, 20 = naproxen 1000 mg, 21 = ibuprofen 1200 mg, 22 = ibuprofen 2400 mg.

$$y_{ikt} \sim \mathcal{N}(\theta_{ikt}, \sigma^2_{ikt})$$

$$\begin{pmatrix} \theta_{ik1} \\ \vdots \\ \theta_{ikT} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} m_{ik1} \\ \vdots \\ m_{ikT} \end{pmatrix}, \begin{pmatrix} \sigma^2_{\text{time}} & \cdots & \sigma^2_{\text{time}} \\ \vdots & 2\sigma^2_{\text{time}} & \cdots & 2\sigma^2_{\text{time}} \\ & \vdots & \ddots & \vdots \\ \sigma^2_{\text{time}} & 2\sigma^2_{\text{time}} & \cdots & T\sigma^2_{\text{time}} \end{pmatrix} \right)$$

$$m_{ikt} = \begin{cases} \mu_{ib} + \beta_{ib}\, t & \text{if } k = b \\ \mu_{ib} + \delta_{ibk} + \beta_{ik}\, t & \text{if } k \neq b \end{cases}$$

$$\delta_{ibk} \sim \mathcal{N}\left(d_{Ak} - d_{Ab}, \sigma^2_{\text{RE}}\right), \quad d_{AA} = 0$$
$$\beta_{ik} \sim \mathcal{N}\left(B_k, \sigma^2_{\text{drift}}\right)$$

where there are a total of T time points at which outcome measurements are taken, m_{ikt} is the mean of the random walk for treatment k in trial i at time point t, σ^2_{time} is the variance of each random walk step, μ_{ib} is the baseline effect of the control arm in trial i, δ_{ibk} is the trial-specific relative effect of treatment k relative to the treatment in the control arm in trial i, d_{Ak} is the relative treatment effect of treatment k versus the reference treatment A (often taken to be placebo), σ^2_{RE} is the between-trial variance of relative effects, β_{ik} is the trial-specific drift for treatment k in trial i, B_k is the pooled drift associated with treatment k, and σ^2_{drift} is the between-trial variance of drifts. The key assumption of exchangeability [12] in NMA is encapsulated in the random effects distributions of relative treatment effects $\delta_{ibk} \sim \mathcal{N}\left(d_{Ak} - d_{Ab}, \sigma^2_{\text{RE}}\right)$. Note, in particular, that consistency is captured through writing the mean of each random effects distribution as the difference between two basic parameters [13], which consist of contrasts between a given treatment and the reference treatment. We take d_{AA} to be zero, since the relative effect of a treatment versus itself is zero. Also apparent in the random effects distributions is the homogeneous between-trial variance assumption, as σ^2_{RE} has no subscripts indicating any particular treatment contrast and is thus independent of the treatment contrast under consideration. This assumption allows multi-arm trials to be incorporated into the model by specifying a marginal distribution for arm 2 and a conditional distribution for arm $j > 2$ given arms $2, \ldots, j-1$ [14]:

$$\delta_{ibk_2} \sim \mathcal{N}(d_{Ak_2} - d_{Ab}, \sigma^2_{\text{RE}}) \quad \text{for} \quad j = 2$$

$$\delta_{ibk_j} \left| \begin{pmatrix} \delta_{ibk_2} \\ \vdots \\ \delta_{ibk_{j-1}} \end{pmatrix} \right. \sim \mathcal{N}\left((d_{Ak_j} - d_{Ab}) + \frac{1}{j-1} \sum_{\ell=1}^{j-1} (\delta_{ibk_\ell} - (d_{Ak_\ell} - d_{Ab})), \frac{j}{2(j-1)} \sigma^2_{\text{RE}} \right) \quad \text{for} \quad j > 2.$$

Since $\{\theta_{ikt}\}$ is assumed to follow a Gaussian random walk with drift, we have by definition

$$\theta_{ikt} = \theta_{ik(t-1)} + \beta_{ik} + z_{ikt},$$

where $z_{ikt} \sim \mathcal{N}(0, \sigma^2_{\text{time}})$, $t = 1, 2, \ldots$. Iterating this equation, we obtain

$$\begin{aligned} \theta_{ikt} &= \theta_{ik(t-1)} + \beta_{ik} + z_{ikt} \\ &= \theta_{ik(t-2)} + 2\beta_{ik} + z_{ik(t-1)} + z_{ikt} \\ &= \ldots \\ &= \theta_{ik0} + \beta_{ik}\, t + \sum_{j=1}^{t} z_{ikj} \end{aligned} \tag{1}$$

where θ_{ik0} is assumed to be some known initial value of the random walk. Taking conditional expectations of both sides of Equation (1) given θ_{ik0} and β_{ik}, we obtain the mean function

$$\begin{aligned} E(\theta_{ikt} | \theta_{ik0}, \beta_{ik}) &= \theta_{ik0} + \beta_{ik}\, t + \sum_{j=1}^{t} E(z_{ikj}) \\ &= \theta_{ik0} + \beta_{ik}\, t. \end{aligned} \tag{2}$$

The definition of m_{ikt} in model 1 then follows from this result by setting $\theta_{ik0} = \mu_{ib}$ if $k = b$ and $\theta_{ik0} = \mu_{ib} + \delta_{ibk}$ if $k \neq b$. Similar to a meta-regression model, the drift parameters are introduced as interaction terms, each of which represents the additional treatment effect per unit change in time point for a given treatment. Unlike a typical meta-regression model, however, we also have a drift term acting in the control arm of each trial. This is a necessary specification because given the initial value and drift, the mean function of a random walk with drift must take the form of Equation (2). A type of exchangeability assumption is also made on the β_{ik}, as encapsulated by $\beta_{ik} \sim \mathcal{N}\left(B_k, \sigma^2_{\text{drift}}\right)$. That is, the trial-specific drift for treatment k is independently drawn from a Gaussian distribution centred at B_k, the pooled drift parameter associated with treatment k. The covariance matrix of the multivariate Gaussian distribution in model 1 also follows from our choice to use a Gaussian random walk with drift to model the mean standardized effects over time. This can be seen by computing the conditional covariance between θ_{iks} and θ_{ikt} given θ_{ik0} and β_{ik} for $s \leq t$:

$$\begin{aligned} Cov(\theta_{iks}, \theta_{ikt} \mid \theta_{ik0}, \beta_{ik}) &= Cov\left(\theta_{ik0} + \beta_{ik}\, s + \sum_{j=1}^{s} z_{ikj},\ \theta_{ik0} + \beta_{ik}\, t + \sum_{\ell=1}^{t} z_{ik\ell} \mid \theta_{ik0}, \beta_{ik} \right) \\ &= Cov\left(\sum_{j=1}^{s} z_{ikj}, \sum_{\ell=1}^{t} z_{ik\ell} \right) \\ &= \sum_{j=1}^{s} \sum_{\ell=1}^{t} Cov\left(z_{ikj}, z_{ik\ell}\right) \\ &= \sum_{j=1}^{s} \text{Var}(z_{ikj}) \\ &= s\sigma^2_{\text{time}} \end{aligned} \tag{3}$$

where the second-to-last line follows since $Cov(z_{ikj}, z_{ik\ell}) = 0$ for $j \neq \ell$ due to statistical independence. Hence, in the covariance matrix in model 1, we have the upside-down L-shape bands that not only reflect the property of increasing variance with increasing time characteristic of a Gaussian random walk, but also the property shown in Equation (3) for the covariance more generally.

In model 1, when estimating relative treatment effects at different time points, it is the relative drifts (i.e., the differences between two B parameters) that are needed, not just an individual drift parameter. To obtain the relative treatment effect at each time point beyond time zero, we take the pooled effect at time point zero and add to it a relative drift multiplied by time term. Suppose D_{pqt} represents the pooled effect of treatment q relative to treatment p at time point t, then we have

$$D_{pqt} = d_{pq} + (B_q - B_p)t. \tag{4}$$

To understand this, imagine each of the trials included in the NMA actually features all treatments in the network, but with many of the reported treatment effects missing at random. Then, it is reasonable to consider as the pooled effect of treatment q versus p at time t the expected difference at time t between the means of the random walks representing the treatment effects in the arm with treatment q and in the arm with treatment p over the population of all trials. That is, to obtain the pooled effect of treatment q versus p at time point t, we compute

$$\begin{aligned} D_{pqt} &= E\left[m_{iqt} - m_{ipt}\right] \\ &= E\left[(\mu_{ib} + \delta_{ibq} + \beta_{iq}t) - (\mu_{ib} + \delta_{ibp} + \beta_{ip}t)\right] \\ &= E\left[(\delta_{ibq} - \delta_{ibp}) + (\beta_{iq} - \beta_{ip})t\right] \\ &= (d_{Aq} - d_{Ab}) - (d_{Ap} - d_{Ab}) + (B_q - B_p)t \\ &= (d_{Aq} - d_{Ap}) + (B_q - B_p)t \\ &= d_{pq} + (B_q - B_p)t. \end{aligned}$$

Model fitting is conducted in a Bayesian framework using any software program that supports Markov chain Monte Carlo (MCMC) simulations, such as JAGS [15] or OpenBUGS [16]. Being a Bayesian model, prior distributions are required to be specified for all unknown parameters in model 1. We take the approach of specifying weakly informative prior distributions, which correspond to a $\mathcal{N}(0, 1)$ prior for the relative treatment effects, d_{Ak}, the pooled drifts, B_k, and the trial-specific baseline effects, μ_{ib}; and a $Unif(0, 2)$ prior for the unknown standard deviation parameters including σ_{time}, σ_{RE}, and σ_{drift}. A value of 2 for the upper bound of the uniform prior distribution is sufficiently vague, as we are dealing with standardized effect sizes.

3.2 | Model 2: Cointegrated Gaussian Random Walks With Drift Network Meta-Analysis Model

Model 2 is an extension of model 1 and accommodates two outcomes measured at multiple time points and assumed to be cointegrated Gaussian random walks. Conceptually, cointegration of two time series occurs when they tend to move together and never stray arbitrarily far from each other. Mathematically, two time series, x_t and y_t, are said to be cointegrated of order (q, p) if x_t and y_t are both integrated of order q and there exists some λ such that the linear combination

$$z_t = \begin{pmatrix} 1 & -\lambda \end{pmatrix} \begin{pmatrix} y_t \\ x_t \end{pmatrix} = y_t - \lambda x_t$$

is integrated of order $q - p$ [17]. A time series is said to be integrated of order q, denoted $I(q)$, if the number of times it is required to be differenced to achieve covariance stationarity is q. The vector $(1 - \lambda)'$ is referred to as a cointegrating vector. In this paper, we treat specifically the case in which two random walks are cointegrated of order $(1, 1)$; that is, each random walk is $I(1)$ and taking a linear combination with the cointegrating vector results in a stationary, $I(0)$, time series. For this reason, the two random walks representing the mean standardized effects over time for the two outcomes are referred to simply as cointegrated random walks, without additional reference to order.

Similar to existing multivariate meta-analysis models, we aim to use the relationship between the two outcomes to obtain parameter estimates with improved precision [18]. Unlike many multivariate meta-analysis models, however, we avoid difficulties often associated with multivariate modelling, such as requiring rarely-reported within-trial correlation estimates or involving sophisticated reparameterization techniques to enable estimation of the between-trial covariance matrix [9, 19]. Although our proposed model makes use of data from two outcomes (i.e., pain and physical function), the roles of the two outcomes are not interchangeable and one of the two outcomes must be chosen as the main outcome, while the other is relegated to the role of auxiliary outcome. The distinction between the main outcome and auxiliary outcome is that model 2 only allows the pooled effects to be estimated for the main outcome, while the auxiliary outcome is only used to improve the precision of the estimates for the main outcome.

Assuming the observed standardized effects for treatment k in trial i at time point t for the main outcome, y_{ikt}, and for the auxiliary outcome, x_{ikt}, follow Gaussian distributions centred at the mean standardized effects for the main outcome, θ_{ikt}, and for the auxiliary outcome, η_{ikt}, with respective variances σ^2_{ikt} and τ^2_{ikt}, model 2 is specified as follows:

$$y_{ikt} \sim \mathcal{N}(\theta_{ikt}, \sigma^2_{ikt})$$
$$x_{ikt} \sim \mathcal{N}(\eta_{ikt}, \tau^2_{ikt})$$

$$\begin{pmatrix} \theta_{ik1} \\ \vdots \\ \theta_{ikT} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} m_{ik1} \\ \vdots \\ m_{ikT} \end{pmatrix}, \begin{pmatrix} \sigma^2_{\text{time}} & \cdots & \sigma^2_{\text{time}} \\ \vdots & 2\sigma^2_{\text{time}} & \cdots & 2\sigma^2_{\text{time}} \\ & \vdots & \ddots & \vdots \\ \sigma^2_{\text{time}} & 2\sigma^2_{\text{time}} & \cdots & T\sigma^2_{\text{time}} \end{pmatrix} \right)$$

$$\eta_{ikt} \sim \mathcal{N}\left(\nu_{ik}\theta_{ikt} + \kappa_{ik}, \sigma^2_{\text{aux}}\right)$$

$$m_{ikt} = \begin{cases} \mu_{ib} + \beta_{ib}\, t & \text{if } k = b \\ \mu_{ib} + \delta_{ibk} + \beta_{ik}\, t & \text{if } k \neq b \end{cases}$$

$$\delta_{ibk} \sim \mathcal{N}\left(d_{Ak} - d_{Ab}, \sigma^2_{\text{RE}}\right), \quad d_{AA} = 0$$
$$\beta_{ik} \sim \mathcal{N}\left(B_k, \sigma^2_{\text{drift}}\right)$$
$$\nu_{ik} \sim \mathcal{N}\left(c_k, \sigma^2_{\text{coint}}\right)$$

where we have introduced additional variables η_{ikt}, the mean standardized effect for the auxiliary outcome for treatment k in trial i at time point t; ν_{ik}, the key parameter in the trial-specific cointegrating vector for treatment k in trial i; κ_{ik}, the level of the stationary time series resulting from the linear combination with the cointegrating vector for treatment k in trial i; σ^2_{aux}, the variance of Gaussian white noise, assumed to be common across all arms in all trials; c_k, the mean cointegrating vector parameter associated with treatment k; and σ^2_{coint}, the between-trial variance of the key parameters in the cointegrating vectors. In addition to the assumptions we make in model 1, we also make another exchangeability assumption on the ν_{ik} being independently drawn from a Gaussian distribution centred at c_k. In model 2, the key component that captures the cointegration of $\{\theta_{ikt}\}$ and $\{\eta_{ikt}\}$ is

$$\eta_{ikt} \sim \mathcal{N}\left(\nu_{ik}\theta_{ikt} + \kappa_{ik}, \sigma^2_{\text{aux}}\right).$$

Another way to write this is as follows

$$\begin{aligned} \eta_{ikt} &= \nu_{ik}\theta_{ikt} + \kappa_{ik} + \epsilon_{ikt}, \quad \epsilon_{ikt} \sim \mathcal{N}\left(0, \sigma^2_{\text{aux}}\right) \\ \eta_{ikt} - \nu_{ik}\theta_{ikt} &= \kappa_{ik} + \epsilon_{ikt} \\ &= z_{ikt}, \text{ where } \; z_{ikt} \sim \mathcal{N}\left(\kappa_{ik}, \sigma^2_{\text{aux}}\right). \end{aligned} \tag{5}$$

The form in Equation (5) shows $\{\theta_{ikt}\}$ and $\{\eta_{ikt}\}$ are indeed cointegrated random walks, since $\{z_{ikt}\}$, being a collection of independently and identically distributed random variables indexed by time, is clearly a stationary time series. Rewritten in this way, we also see why these additional variables hold the interpretations that they do.

Despite conferring some of the same benefits as multivariate meta-analysis models, model 2 is technically not a multivariate NMA model. This means that to estimate pooled effects for both outcomes, it is necessary to fit model 2 twice; once with the first outcome as the main outcome and the second outcome as the auxiliary outcome, and once with the second outcome as the main outcome and the first outcome as the auxiliary outcome. Model fitting is again conducted in a Bayesian framework using any software package that supports MCMC simulations. As with other Bayesian methods, prior distributions have to be specified for all unknown parameters in the model. We have again opted to use weakly informative prior distributions, which correspond to a $\mathcal{N}(0, 1)$ prior distribution for the relative treatment effects, d_{Ak}, the pooled drifts, B_k, the trial-specific baseline effects, μ_{ib}, and the trial-specific stationary time series levels, κ_{ik}; a $\mathcal{N}(1, 1)$ prior distribution for the mean cointegrating vector parameters, c_k; and a $Unif(0, 2)$ prior distribution for the unknown standard deviation parameters including σ_{time}, σ_{aux}, σ_{RE}, σ_{drift}, and σ_{coint}.

3.3 | Simulation Study

Although the models presented in this paper are Bayesian models, it is nevertheless desirable for Bayesian models to yield inferences with good frequentist properties [20]. We took such a calibrated Bayes approach here and conducted a simulation study to evaluate frequentist properties of our models. To this end, we devised four simulation scenarios corresponding to a two-by-two factorial design in which the number of included trials in the NMA is either 30 or 50, and the degree of heterogeneity in the relative effects, drifts, and several other parameters is either moderate or high. For each scenario, we generated 100 datasets according to the data-generating mechanism outlined below in Algorithm 1. Then, we fitted model 1 and model 2 to each of the simulated datasets and aggregated the results over all 100 replications. We additionally fitted a third model as a reference model against which to compare the two proposed models. This reference model is the random walk NMA model of da Costa et al. [5], hereafter also referred to as model 0. The random walk NMA model of da Costa et al. was chosen as the reference model because it also uses random walks, albeit without drift, to enable borrowing of information across time points, unlike traditional univariate or bivariate NMA models that can only use data from a single time point. The choice of 100 replications was mainly due to limits in available computational resources. For each replication of each scenario, we used a large number of MCMC samples from three chains (100 000 samples per chain for model 0 after 50 000 burn-in; 200 000 samples per chain for model 1 after 100 000 burn-in; 300 000 samples per chain for model 2 after 150 000 burn-in) to obtain posterior summaries for numerous parameters of interest in each model. The MCMC simulations were implemented using JAGS [15] through the rjags package in R. Convergence was checked by examining the trace plots and R-hat values [21]. In some scenarios, we encountered a few replications with non-convergence of some parameters of interest, which could occur when running a simulation study with complex models resulting in difficult posterior geometry. Because there were few such replications (at most 3 out of 100), we opted to replace these with additional replications for which all parameters of interest have converged. Due to the computational demands of running this simulation study, we took advantage of our access to the Niagara supercomputer [22].

3.3.1 | Data-Generating Mechanism

The data-generating mechanism we used to simulate the datasets for our simulation study follow model 2. First, we specified the values for key parameters, such as relative treatment effects and drifts, and for other important variables, such as number of trials, number of arms, and number of time points. Then, we followed the probabilistic model specified in model 2 to produce the observed standardized effects. For all scenarios, we simulated data from 30 or 50 trials ($ntrial \in \{30, 50\}$), each with between two to four arms ($narms = 4$) with each arm featuring one of six different treatments ($ntrt = 6$) and with outcomes measured at least at one of seven time points ($ntime = 7$). Algorithm 1 outlines the steps in pseudocode for generating the RCT data according to model 2. In Algorithm 1, the number of time points for the auxiliary outcome in a given trial, *anout*, is sampled from $\{1, \ldots, 7\}$ with some desired probability distribution. The elements of the vector of time points with auxiliary outcome measurements, *atimept*, are also sampled from $\{1, \ldots, 7\}$ without replacement with some desired vector of probabilities that could be based on empirical knowledge (e.g., from a previous NMA) of the likelihood of sampling each time point. The number of arms in a trial, *narm*, is sampled from $\{2, 3, 4\}$ with sampling probabilities $\{0.6, 0.2, 0.2\}$ to reflect the higher likelihood of obtaining data from two-arm trials. The *narm* elements of the treatment vector, *trt*, are sampled without replacement from $\{1, \ldots, 6\}$ with sampling probabilities $\{0.5, 0.1, \ldots, 0.1\}$ to reflect the higher likelihood of a trial featuring placebo in the control arm. Line 26 and 31 of Algorithm 1 contain a variable G which is sampled from a gamma distribution with a shape parameter of 20 and rate parameter of 15. The random variable G acts as a multiplier on the base standard error of observed standardized effects to give the actual standard error, and the right-skewed gamma distribution is chosen to reflect the presence of fewer small trials (with high standard errors) than large trials in the collection of RCTs.

The four scenarios for the simulation study are distinguished by two different settings for the number of trials and for the degree of heterogeneity in various parameters. Scenarios are used in this simulation study, since doing a full factorial design would not be feasible in this case with the numerous parameters involved as seen in Table 1. For all four scenarios, we take the following list of vectors to be common across all scenarios

$$
\begin{aligned}
d &= \begin{pmatrix} 0 & -0.25 & -0.37 & -0.55 & -0.73 & -1.22 \end{pmatrix}' \\
B &= \begin{pmatrix} -0.018 & -0.027 & -0.035 & -0.042 & -0.053 & -0.058 \end{pmatrix}' \\
bsl &= \begin{pmatrix} -0.2 & -0.5 & -0.3 & -0.6 & -0.4 & -0.8 \end{pmatrix}' \\
sig.m &= \begin{pmatrix} 0.011 & 0.013 & 0.014 & 0.017 & 0.020 & 0.024 \end{pmatrix}' \\
sig.a &= \begin{pmatrix} 0.012 & 0.015 & 0.016 & 0.018 & 0.023 & 0.026 \end{pmatrix}' \\
cr &= \begin{pmatrix} 1.32 & 1.26 & 1.17 & 1.22 & 1.13 & 1.06 \end{pmatrix}' \\
lv &= \begin{pmatrix} -0.02 & 0.01 & -0.03 & -0.02 & 0.03 & 0.04 \end{pmatrix}'
\end{aligned}
$$

with each vector containing six elements corresponding to the six different treatments. The vector d is the vector of pooled effects versus treatment 1 at time point zero, and a range of values have been chosen representing standardized effects that are below, at, or above the minimal clinically important difference (MCID) of -0.37 based on the median MCID found in previous studies with OA patients [23]. B is the vector of pooled drift parameters, bsl is the vector of baseline treatment effects, $sig.m$ is the vector of base standard errors of observed standardized effects for the main outcome, $sig.a$ is the vector of base standard errors of observed standardized effects for the auxiliary outcome, cr is the vector of means of the distributions of cointegrating vector parameters, and lv is the vector of means of the distributions of levels of the stationary time series resulting from taking linear combinations with cointegrating vectors. In Table 1, although the values of the heterogeneity parameters appear small, their sizes are chosen in relation to the underlying parameters they describe. For instance, since d is a vector of standardized effect sizes, a value

TABLE 1 | Four scenarios for the simulation study representing different degrees of heterogeneity in key parameters (moderate or high) with different numbers of trials included in the network meta-analysis (30 or 50).

Scenario	Degree of heterogeneity	Number of trials	$sd.RE$	$sd.time$	$sd.drift$	$sd.bsl$	$sd.cr$	$sd.lv$	$sd.aux$
1	Moderate	30	0.04	0.01	0.015	0.03	0.02	0.002	0.01
2	High	30	0.20	0.04	0.060	0.12	0.10	0.010	0.04
3	Moderate	50	0.04	0.01	0.015	0.03	0.02	0.002	0.01
4	High	50	0.20	0.04	0.060	0.12	0.10	0.010	0.04

of 0.04 for $sd.RE$ would represent a moderate degree of heterogeneity, whereas a value of 0.20 would represent a high degree of heterogeneity.

3.3.2 | Performance Measures

For the parameters of interest, a Bayesian analysis produces posterior distributions, each of which we can summarize with the posterior median and the quantile-based 95% credible interval (CrI). We view this as a procedure for constructing the point and interval estimates of the parameters of interest, and aim to evaluate the frequentist properties of these estimates through the simulation study. For all parameters of interest, we computed seven measures of performance including bias, model standard error (SE), empirical SE, relative percent error in model SE, mean squared error (MSE), mean CrI width, and coverage, along with their Monte Carlo SE [24]. For both model 1 and model 2, we computed these performance measures for time point one values of the relative treatment effects versus treatment 1. The reason for choosing time point one instead of the default time point zero is to allow comparisons with an existing NMA model which does not accommodate estimation of pooled effects at time point zero. For model 0, a single reference time point must be chosen from among the numerous time points at which outcome measurements exist. As time point one is such a time point, the pooled effects under model 0 can be estimated at this reference time point. Thus, we also computed estimates of these seven performance measures for model 0 in reference to time point one values of the pooled effects to facilitate comparisons with the results obtained using the two proposed models.

For a given parameter of interest with true value θ, suppose that $\hat{\theta}_i$ represents the estimate from the ith replication given by the posterior median. Bias is estimated by computing

$$\text{bias} = \frac{1}{n_{\text{sim}}}\sum_{i=1}^{n_{\text{sim}}}\hat{\theta}_i - \theta \ .$$

Bias is computed to determine whether the models produce estimates that target the parameters of interest on average. Model SE refers to the root-mean of the posterior variances over the 100 replications and is computed to reflect the precision of the model. Model SE is computed as

$$\text{model SE} = \sqrt{\frac{1}{n_{\text{sim}}}\sum_{i=1}^{n_{\text{sim}}}\widehat{\text{Var}}(\hat{\theta}_i)} \ .$$

Empirical SE refers to the long-run standard deviation of parameter estimates over the 100 replications and is computed to reflect the precision of the estimator of a parameter of interest. Empirical SE is computed as

$$\text{empirical SE} = \sqrt{\frac{1}{n_{\text{sim}}-1}\sum_{i=1}^{n_{\text{sim}}}(\hat{\theta}_i - \bar{\theta})^2} \ .$$

The reason for computing both model SE and empirical SE is that we wish to determine whether the model SE systematically misses the empirical SE, which is targeted by the model SE. The way this is quantified is through the relative percent error in model SE, estimated using the formula

$$\text{relative percent error in model SE} = 100\left(\frac{\widehat{\text{modelSE}}}{\widehat{\text{empiricalSE}}} - 1\right).$$

MSE takes the true value of θ into account and combines bias and empirical SE into one performance measure. MSE is estimated by computing

$$\text{MSE} = \frac{1}{n_{\text{sim}}}\sum_{i=1}^{n_{\text{sim}}}(\hat{\theta}_i - \theta)^2 \ .$$

Mean CrI width refers to the average difference between the 97.5% and 2.5% quantiles of the posterior distribution over the 100 replications and is computed to determine whether there are differences in the mean CrI widths produced by the various models. Mean CrI width is estimated using the formula

$$\text{mean CrI width} = \frac{1}{n_{\text{sim}}}\sum_{i=1}^{n_{\text{sim}}}(\hat{\theta}_{0.975,i} - \hat{\theta}_{0.025,i}) \ .$$

Finally, coverage is defined as the probability that a CrI contains the true parameter value. This frequentist probability is computed to check agreement with the nominal value of 95%. Coverage is estimated by computing

$$\text{coverage} = \frac{1}{n_{\text{sim}}}\sum_{i=1}^{n_{\text{sim}}}\mathbb{1}(\hat{\theta}_{0.025,i} \le \theta \le \hat{\theta}_{0.975,i}) \ .$$

4 | Results

4.1 | Results for Simulation Study

4.1.1 | Bias

For model 1 and 2, bias estimates for the relative treatment effect parameters (the D parameters for time point one) across all four

ALGORITHM 1 | Data-generating mechanism to simulate RCT data based on the cointegrated Gaussian random walks with drift NMA model.

Require: predefined vectors $d[ntrt]$, $B[ntrt]$, $bsl[ntrt]$, $sig.m[ntrt]$, $sig.a[ntrt]$, $cr[ntrt]$, $lv[ntrt]$ and predefined values $ntrial$, $narms$, $ntime$, $sd.RE$, $sd.time$, $sd.drift$, $sd.bsl$, $sd.cr$, $sd.lv$, $sd.aux$

instantiate the following arrays for storing simulated RCT data: $y[ntrial, narms, ntime]$, $x[ntrial, narms, ntime]$, $SE.y[ntrial, narms, ntime]$, $SE.x[ntrial, narms, ntime]$, and $treat[ntrial, narms]$

for $i = 1, \ldots, ntrial$ **do**

  $anout \leftarrow$ sample one of $\{1, \ldots, ntime\}$ with desired probability for number of time points for auxiliary outcome

  $atimept \leftarrow$ sample $anout$ elements of $\{1, \ldots, ntime\}$ without replacement with desired probability vector

  $narm \leftarrow$ sample one of $\{2, \ldots, narms\}$ with probability $\{0.6, \frac{1-0.6}{narms-2}, \ldots, \frac{1-0.6}{narms-2}\}$

  $trt \leftarrow$ sample $narm$ elements of $\{1, \ldots, ntrt\}$ without replacement with probability vector $\{0.5, \frac{1-0.5}{ntrt-1}, \ldots, \frac{1-0.5}{ntrt-1}\}$

  $\mu \leftarrow$ sample one value from $\mathcal{N}(bsl[trt[1]], sd.bsl)$

  $V \leftarrow$ sample vector of $ntime$ values from $Unif(0, 1)$

  for $k = 1, \ldots, narm$ **do**

    instantiate empty vectors $m_{ik}[ntime]$, $\theta_{ik}[ntime]$, $\eta_{ik}[ntime]$

    $\delta \leftarrow$ sample one value from $\mathcal{N}(d[trt[k]] - d[trt[1]], sd.RE)$

    $\beta \leftarrow$ sample one value from $\mathcal{N}(B[trt[k]], sd.drift)$

    $\nu \leftarrow$ sample one value from $\mathcal{N}(cr[trt[k]], sd.cr)$

    $\kappa \leftarrow$ sample one value from $\mathcal{N}(lv[trt[k]], sd.lv)$

    for $t = 1, \ldots, ntime$ **do**

      if $k == 1$ **then**

        $m_{ik}[t] \leftarrow \mu + \beta t$

      else

        $m_{ik}[t] \leftarrow \mu + \delta + \beta t$

      end if

    end for

    $\theta_{ik} \leftarrow$ sample one vector from $\mathcal{N}\left(m_{ik}, \begin{pmatrix} sd.time^2 & \cdots & sd.time^2 \\ \vdots & 2 \times sd.time^2 \cdots & 2 \times sd.time^2 \\ \vdots & \vdots & \ddots \\ sd.time^2 & 2 \times sd.time^2 \cdots & ntime \times sd.time^2 \end{pmatrix}\right)$

    for $t = 1, \ldots, ntime$ **do**

      $\eta_{ik}[t] \leftarrow$ sample one value from $\mathcal{N}(\nu\theta_{ik}[t] + \kappa,\ sd.aux)$

      if $t \in atimept$ and $V[t]>0.1$ **then**

        $G \leftarrow$ sample one value from $Gamma(20, 15)$

        $y[i, k, t] \leftarrow$ sample one value from $\mathcal{N}\left(\theta_{ik}[t],\ G \times sig.m[trt[k]]\right)$

        $SE.y[i, k, t] \leftarrow G \times sig.m[trt[k]]$

      end if

      if $t \in atimept$ **then**

        $G \leftarrow$ sample one value from $Gamma(20, 15)$

        $x[i, k, t] \leftarrow$ sample one value from $\mathcal{N}\left(\eta_{ik}[t],\ G \times sig.a[trt[k]]\right)$

        $SE.x[i, k, t] \leftarrow G \times sig.a[trt[k]]$

      end if

    end for

    $treat[i, k] \leftarrow trt[k]$

  end for

end for

scenarios are close to zero, as seen in Table 2 and Figure 2. Although for many parameters, the bias estimates in high heterogeneity scenarios are larger than the bias estimates in moderate heterogeneity scenarios, the bias estimates in high heterogeneity scenarios still remain at least an order of magnitude smaller than the corresponding true parameter values. Also, bias estimates are not consistently in one direction across all scenarios. Thus, bias estimates from both model 1 and 2 can be considered small and negligible for the relative treatment effect parameters in all scenarios. In comparison, for model 0 under low heterogeneity scenarios, bias estimates appear to get progressively greater in magnitude with increasing treatment number. This is not completely unexpected, since the true drift parameter increases in magnitude with increasing treatment number and since model 0 implicitly assumes there is no drift in the random walks. Thus, despite borrowing information across time points when estimating relative treatment effects, model 0 still tends to yield relatively more biased estimates of relative treatment effects in the presence of nonzero drifts in the data-generating mechanism. Nevertheless, it should be noted that even for model 0, the relatively larger

TABLE 2 | Bias estimates for relative treatment effects, drifts, and related between-trial heterogeneity parameters in all scenarios.

	Scenario 1 (30; moderate)			Scenario 2 (30; high)		
Parameter	**Model 0**	**Model 1**	**Model 2**	**Model 0**	**Model 1**	**Model 2**
D_2	−0.006 (0.003)	−0.002 (0.002)	−0.001 (0.002)	0.01 (0.01)	0.01 (0.01)	0.02 (0.01)
D_3	−0.012 (0.002)	−0.001 (0.002)	−0.001 (0.002)	−0.02 (0.01)	−0.009 (0.009)	−0.004 (0.008)
D_4	−0.014 (0.002)	0.003 (0.002)	0.003 (0.002)	−0.01 (0.01)	0.013 (0.009)	0.012 (0.009)
D_5	−0.025 (0.003)	0.001 (0.002)	0.002 (0.002)	−0.02 (0.01)	0.01 (0.01)	0.013 (0.009)
D_6	−0.036 (0.003)	−0.001 (0.002)	0.000 (0.002)	−0.032 (0.009)	0.003 (0.008)	0.004 (0.008)
B_1		0.0001 (0.0004)	0.0003 (0.0003)		0.000 (0.001)	0.000 (0.001)
B_2		0.0012 (0.0006)	0.0010 (0.0006)		0.000 (0.003)	0.001 (0.002)
B_3		0.0000 (0.0007)	0.0001 (0.0007)		0.003 (0.002)	0.003 (0.002)
B_4		−0.0012 (0.0006)	−0.0012 (0.0006)		−0.001 (0.002)	0.001 (0.002)
B_5		0.0003 (0.0008)	0.0004 (0.0008)		0.000 (0.002)	0.000 (0.002)
B_6		0.0001 (0.0007)	0.0000 (0.0006)		0.000 (0.003)	−0.001 (0.002)
σ_{RE}	0.004 (0.001)	0.001 (0.001)	0.001 (0.001)	0.065 (0.004)	0.018 (0.004)	0.017 (0.004)
σ_{drift}		−0.0003 (0.0002)	−0.0002 (0.0002)		0.0011 (0.0008)	0.0010 (0.0007)
	Scenario 3 (50; moderate)			**Scenario 4 (50; high)**		
Parameter	**Model 0**	**Model 1**	**Model 2**	**Model 0**	**Model 1**	**Model 2**
D_2	−0.004 (0.002)	0.002 (0.002)	0.002 (0.002)	−0.016 (0.008)	−0.006 (0.007)	−0.007 (0.007)
D_3	−0.012 (0.002)	−0.001 (0.001)	−0.001 (0.001)	0.003 (0.007)	0.013 (0.006)	0.016 (0.007)
D_4	−0.016 (0.002)	0.000 (0.002)	0.001 (0.002)	−0.016 (0.008)	0.006 (0.007)	0.004 (0.007)
D_5	−0.027 (0.002)	0.000 (0.002)	0.000 (0.001)	−0.034 (0.007)	−0.002 (0.006)	−0.001 (0.006)
D_6	−0.034 (0.002)	0.002 (0.002)	0.002 (0.002)	−0.034 (0.008)	−0.001 (0.007)	−0.001 (0.007)
B_1		−0.0003 (0.0003)	−0.0003 (0.0003)		0.001 (0.001)	0.001 (0.001)
B_2		−0.0003 (0.0005)	−0.0004 (0.0005)		−0.002 (0.002)	−0.002 (0.002)
B_3		−0.0006 (0.0005)	−0.0004 (0.0005)		−0.002 (0.002)	−0.002 (0.002)
B_4		−0.0002 (0.0005)	−0.0003 (0.0005)		−0.004 (0.002)	−0.004 (0.002)
B_5		−0.0010 (0.0005)	−0.0009 (0.0005)		0.000 (0.002)	0.000 (0.002)
B_6		0.0000 (0.0006)	−0.0003 (0.0005)		0.000 (0.002)	0.000 (0.002)
σ_{RE}	0.0064 (0.0009)	0.0020 (0.0009)	0.0018 (0.0009)	0.059 (0.003)	0.017 (0.003)	0.018 (0.003)
σ_{drift}		0.0002 (0.0002)	0.0001 (0.0001)		0.0006 (0.0005)	0.0006 (0.0004)

Note: Monte Carlo standard errors in parentheses. Scenarios labelled with: (number of trials; degree of heterogeneity).

bias estimates are still an order of magnitude smaller than the corresponding true parameter values.

In high heterogeneity scenarios, model 0 does not appear to perform as poorly, in terms of bias of relative treatment effect parameter estimates, when compared with model 1 and 2. In this case, since the high heterogeneity setting also applies to the standard deviation of each random walk step, it is possible for a random walk with no drift to cover the greater distance traversed by a random walk with drift purely by chance, so long as the drift is not too large relative to the standard deviation of the random walk steps. However, as there are treatments with relatively large values of drifts in the simulated data, model 1 and 2 are still more consistent in producing estimates of relative treatment effects with low bias even in high heterogeneity scenarios when compared to model 0.

For model 1 and 2, bias estimates for the drift parameters across all four scenarios are close to zero, as seen in Table 2. Although for many drift parameters, the bias estimates in high heterogeneity scenarios are larger than the bias estimates in moderate heterogeneity scenarios, the bias estimates still remain at least an order of magnitude smaller than the corresponding true parameter values. Additionally, bias estimates are not consistently in one direction across all scenarios. Thus, bias can again be considered low and negligible for all the pooled drift parameters in all scenarios.

For the heterogeneity parameter σ_{RE}, the bias estimates are positive in all scenarios, suggesting that there is a tendency for all models to overestimate the random effects standard deviation to some degree. However, the bias estimates from model 1 and 2 are at least an order of magnitude smaller than the true value of σ_{RE} in all scenarios, while the same observation cannot be made

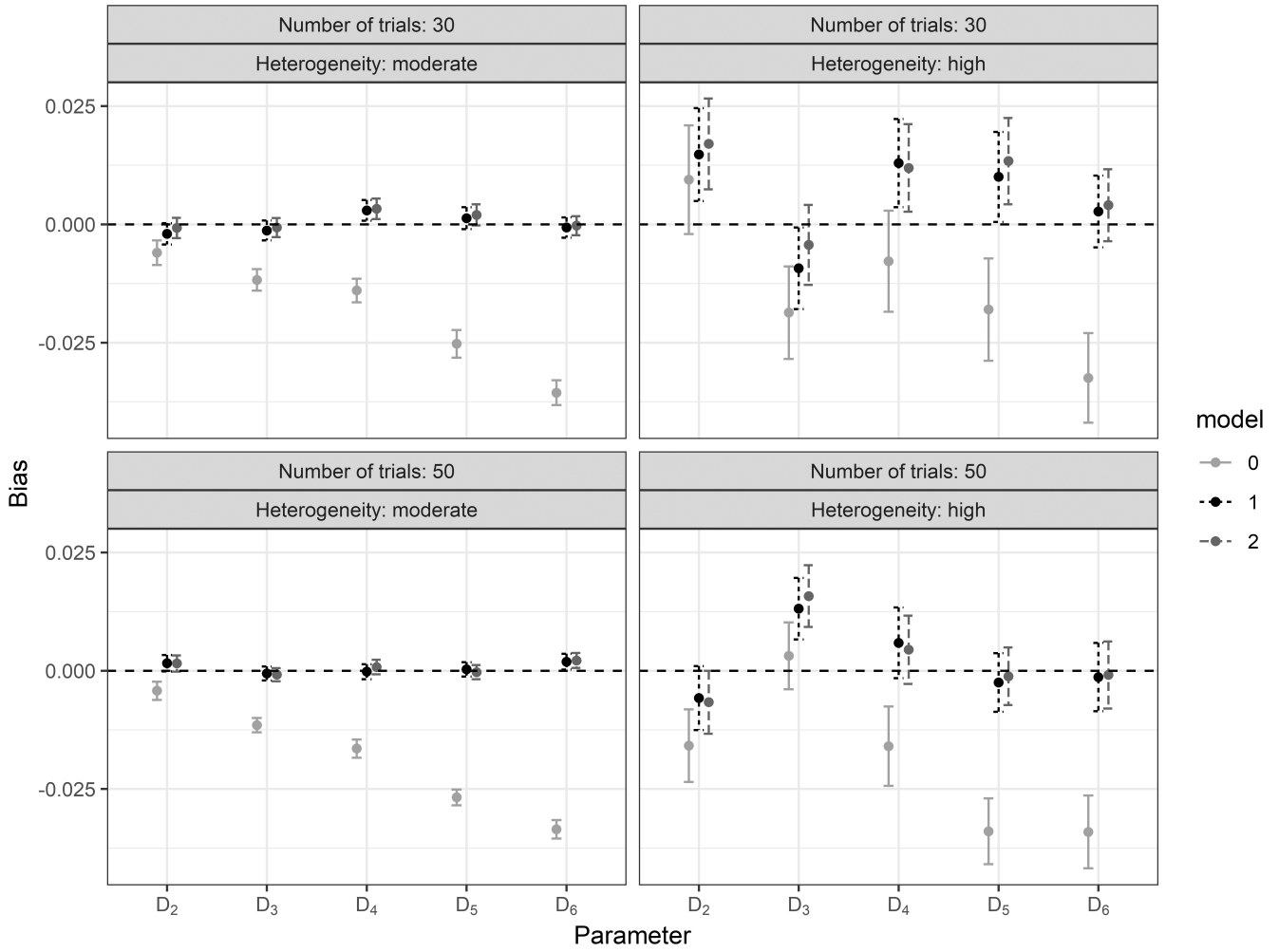

FIGURE 2 | Bias in relative treatment effect parameter estimates from model 0, model 1, and model 2 in all scenarios. Bias shown with error bars extending one Monte Carlo SE above and below estimate.

about model 0. Model 0 likely overestimates the random effects standard deviation to a greater extent to compensate for assuming zero drift in the random walks when, in fact, the true drift parameters are nonzero. For the heterogeneity parameter σ_{drift}, both model 1 and 2 produce bias estimates that are at least an order of magnitude smaller than the true value of σ_{drift} in all scenarios. Thus, bias estimates for both heterogeneity parameters of interest can be considered low when using model 1 and 2.

4.1.2 | Model Standard Error, Empirical Standard Error, and Relative Percent Error in Model Standard Error

Although Tables 3 and 4 display estimates of model SE and empirical SE from all three models in all scenarios, it may not be easy to check the degree of agreement between model SE and empirical SE at a glance. Thus, Table S1 (Supporting Information) is also provided and displays the relative percent error in model SE, with a positive value indicating that model SE overestimates the empirical SE and a negative value indicating that model SE underestimates the empirical SE. While the magnitude of several estimates of relative percent error in model SE may appear high (e.g., model 1 and 2 estimates for D_6 under scenario 2), these estimates are not systematically biased in a particular direction across all scenarios, nor are they systematically greater in magnitude than the relative percent error in model SE estimates for other parameters across all scenarios. Thus, we need not be particularly concerned about the apparent disagreement between model SE and empirical SE for a few parameters under isolated scenarios, especially since only 100 replications are used in this simulation study.

For both model SE and empirical SE, the effect of increasing the degree of heterogeneity from moderate to high while keeping the number of trials the same is to increase the model SE and empirical SE as expected. That is, the model SE estimates and empirical SE estimates in scenario 2 are much greater than the corresponding estimates in scenario 1 for all models, and similarly for scenario 4 versus scenario 3. For both model SE and empirical SE, the effect of increasing the number of trials from 30 to 50 while keeping the degree of heterogeneity the same is to decrease the model SE and empirical SE also as expected.

For the relative treatment effect parameters, the model SE estimates appear to be noticeably reduced in all scenarios when comparing model 1 and 2 to model 0, as seen in Figure 3. In Figure 4,

TABLE 3 | Model standard error estimates for relative treatment effects, drifts, and related between-trial heterogeneity parameters in all scenarios.

	Scenario 1 (30; moderate)			Scenario 2 (30; high)		
Parameter	**Model 0**	**Model 1**	**Model 2**	**Model 0**	**Model 1**	**Model 2**
D_2	0.0255 (0.0007)	0.0220 (0.0005)	0.0211 (0.0005)	0.104 (0.002)	0.092 (0.003)	0.090 (0.002)
D_3	0.0254 (0.0007)	0.0217 (0.0005)	0.0209 (0.0005)	0.104 (0.002)	0.091 (0.002)	0.089 (0.002)
D_4	0.0256 (0.0007)	0.0220 (0.0005)	0.0212 (0.0005)	0.104 (0.002)	0.091 (0.002)	0.090 (0.002)
D_5	0.0264 (0.0006)	0.0230 (0.0005)	0.0221 (0.0004)	0.108 (0.003)	0.094 (0.002)	0.092 (0.002)
D_6	0.0262 (0.0005)	0.0230 (0.0005)	0.0220 (0.0004)	0.102 (0.002)	0.089 (0.002)	0.088 (0.002)
B_1		0.00349 (0.00004)	0.00336 (0.00003)		0.0135 (0.0002)	0.0133 (0.0001)
B_2		0.0067 (0.0002)	0.0065 (0.0001)		0.0253 (0.0008)	0.0247 (0.0007)
B_3		0.0067 (0.0001)	0.0065 (0.0001)		0.0249 (0.0005)	0.0243 (0.0004)
B_4		0.0068 (0.0001)	0.0066 (0.0001)		0.0250 (0.0006)	0.0247 (0.0006)
B_5		0.0073 (0.0002)	0.0070 (0.0002)		0.026 (0.001)	0.026 (0.001)
B_6		0.0073 (0.0001)	0.0070 (0.0001)		0.0246 (0.0005)	0.0241 (0.0005)
σ_{RE}	0.0125 (0.0002)	0.0123 (0.0002)	0.0108 (0.0001)	0.0427 (0.0006)	0.0370 (0.0005)	0.0358 (0.0004)
σ_{drift}		0.00205 (0.00003)	0.00183 (0.00002)		0.00687 (0.00007)	0.00672 (0.00007)
	Scenario 3 (50; moderate)			**Scenario 4 (50; high)**		
Parameter	**Model 0**	**Model 1**	**Model 2**	**Model 0**	**Model 1**	**Model 2**
D_2	0.0190 (0.0003)	0.0160 (0.0003)	0.0154 (0.0003)	0.077 (0.001)	0.068 (0.001)	0.0674 (0.0009)
D_3	0.0190 (0.0003)	0.0160 (0.0003)	0.0154 (0.0002)	0.077 (0.001)	0.068 (0.001)	0.068 (0.001)
D_4	0.0199 (0.0004)	0.0167 (0.0003)	0.0160 (0.0003)	0.077 (0.001)	0.068 (0.001)	0.067 (0.001)
D_5	0.0202 (0.0003)	0.0172 (0.0003)	0.0167 (0.0003)	0.076 (0.001)	0.0673 (0.0009)	0.0669 (0.0009)
D_6	0.0205 (0.0003)	0.0178 (0.0003)	0.0170 (0.0003)	0.077 (0.001)	0.068 (0.001)	0.068 (0.001)
B_1		0.00272 (0.00002)	0.00262 (0.00002)		0.01016 (0.00008)	0.01004 (0.00007)
B_2		0.00504 (0.00007)	0.00486 (0.00007)		0.0187 (0.0002)	0.0185 (0.0002)
B_3		0.00505 (0.00008)	0.00488 (0.00007)		0.0188 (0.0003)	0.0186 (0.0003)
B_4		0.0053 (0.0001)	0.00506 (0.00009)		0.0188 (0.0003)	0.0185 (0.0003)
B_5		0.00553 (0.00007)	0.00535 (0.00007)		0.0186 (0.0002)	0.0183 (0.0002)
B_6		0.00575 (0.00009)	0.00550 (0.00009)		0.0190 (0.0003)	0.0187 (0.0003)
σ_{RE}	0.0089 (0.0001)	0.0091 (0.0001)	0.0078 (0.0001)	0.0305 (0.0003)	0.0267 (0.0002)	0.0261 (0.0002)
σ_{drift}		0.00149 (0.00001)	0.00136 (0.00001)		0.00506 (0.00003)	0.00498 (0.00003)

Note: Monte Carlo standard errors in parentheses. Scenarios labelled with: (number of trials; degree of heterogeneity).

the reductions in empirical SE estimates from using model 1 or 2 over model 0 are generally present, but relatively less pronounced than the model SE reductions given the Monte Carlo standard errors around these estimates. Comparing model 2 to model 1, the model SE estimates appear to be reduced in moderate heterogeneity scenarios; while in high heterogeneity scenarios, model SE reductions appear to be small even when present. For empirical SE, the reductions in estimates from using model 2 over model 1 are seen in some parameters albeit less pronounced than the model SE reductions. As seen in Figure S1 (Supporting Information), the estimates of relative percent error in model SE for the relative treatment effect parameters are comparable between model 1 and model 2 in all scenarios, while differing from model 0 estimates in a few cases.

For the drift parameters, small reductions in model SE estimates from using model 2 over model 1 are generally observed in all scenarios. Empirical SE estimates remain comparable between model 1 and model 2, and are only slightly reduced by using model 2 over model 1 in some cases. The estimates of relative percent error in model SE for the drift parameters remain comparable between model 1 and model 2 in all scenarios.

For σ_{RE}, reductions in model SE from using model 2 over model 0 are present in all scenarios, while reductions in model SE from using model 1 over model 0 are only seen in high heterogeneity scenarios. Similar observations can be made about empirical SE estimates, with empirical SE reductions generally being less pronounced than model SE reductions given the Monte Carlo standard errors. The estimates of relative percent error in model SE for σ_{RE} are negative in all scenarios, suggesting that there is a tendency for model SE to underestimate empirical SE regardless of the model used. Comparing model 2 to model 1, model SE estimates for σ_{RE} are reduced by using model 2 over model

TABLE 4 | Empirical standard error estimates for relative treatment effects, drifts, and related between-trial heterogeneity parameters in all scenarios.

	Scenario 1 (30; moderate)			Scenario 2 (30; high)		
Parameter	**Model 0**	**Model 1**	**Model 2**	**Model 0**	**Model 1**	**Model 2**
D_2	0.026 (0.002)	0.023 (0.002)	0.021 (0.002)	0.115 (0.008)	0.098 (0.007)	0.096 (0.007)
D_3	0.023 (0.002)	0.021 (0.001)	0.020 (0.001)	0.098 (0.007)	0.086 (0.006)	0.085 (0.006)
D_4	0.025 (0.002)	0.022 (0.002)	0.022 (0.002)	0.107 (0.008)	0.093 (0.007)	0.092 (0.007)
D_5	0.029 (0.002)	0.023 (0.002)	0.022 (0.002)	0.108 (0.008)	0.095 (0.007)	0.091 (0.006)
D_6	0.026 (0.002)	0.022 (0.002)	0.020 (0.001)	0.095 (0.007)	0.076 (0.005)	0.076 (0.005)
B_1		0.0035 (0.0002)	0.0035 (0.0002)		0.014 (0.001)	0.014 (0.001)
B_2		0.0060 (0.0004)	0.0059 (0.0004)		0.025 (0.002)	0.025 (0.002)
B_3		0.0071 (0.0005)	0.0067 (0.0005)		0.022 (0.002)	0.023 (0.002)
B_4		0.0065 (0.0005)	0.0062 (0.0004)		0.024 (0.002)	0.024 (0.002)
B_5		0.0082 (0.0006)	0.0075 (0.0005)		0.025 (0.002)	0.024 (0.002)
B_6		0.0069 (0.0005)	0.0064 (0.0005)		0.026 (0.002)	0.025 (0.002)
σ_{RE}	0.0127 (0.0009)	0.0132 (0.0009)	0.0111 (0.0008)	0.044 (0.003)	0.041 (0.003)	0.039 (0.003)
σ_{drift}		0.0022 (0.0002)	0.0018 (0.0001)		0.0078 (0.0006)	0.0074 (0.0005)
	Scenario 3 (50; moderate)			**Scenario 4 (50; high)**		
Parameter	**Model 0**	**Model 1**	**Model 2**	**Model 0**	**Model 1**	**Model 2**
D_2	0.019 (0.001)	0.017 (0.001)	0.017 (0.001)	0.077 (0.005)	0.068 (0.005)	0.067 (0.005)
D_3	0.015 (0.001)	0.015 (0.001)	0.014 (0.001)	0.071 (0.005)	0.065 (0.005)	0.065 (0.005)
D_4	0.019 (0.001)	0.016 (0.001)	0.015 (0.001)	0.084 (0.006)	0.075 (0.005)	0.072 (0.005)
D_5	0.017 (0.001)	0.015 (0.001)	0.015 (0.001)	0.070 (0.005)	0.062 (0.004)	0.061 (0.004)
D_6	0.020 (0.001)	0.017 (0.001)	0.016 (0.001)	0.077 (0.005)	0.072 (0.005)	0.071 (0.005)
B_1		0.0027 (0.0002)	0.0027 (0.0002)		0.0106 (0.0008)	0.0104 (0.0007)
B_2		0.0052 (0.0004)	0.0052 (0.0004)		0.021 (0.002)	0.020 (0.001)
B_3		0.0051 (0.0004)	0.0049 (0.0003)		0.019 (0.001)	0.019 (0.001)
B_4		0.0050 (0.0004)	0.0049 (0.0003)		0.019 (0.001)	0.019 (0.001)
B_5		0.0052 (0.0004)	0.0051 (0.0004)		0.018 (0.001)	0.019 (0.001)
B_6		0.0057 (0.0004)	0.0054 (0.0004)		0.018 (0.001)	0.018 (0.001)
σ_{RE}	0.0090 (0.0006)	0.0095 (0.0007)	0.0089 (0.0006)	0.032 (0.002)	0.028 (0.002)	0.029 (0.002)
σ_{drift}		0.0016 (0.0001)	0.0015 (0.0001)		0.0046 (0.0003)	0.0044 (0.0003)

Note: Monte Carlo standard errors in parentheses. Scenarios labelled with: (number of trials; degree of heterogeneity).

1 in all scenarios. Empirical SE reductions from using model 2 over model 1 are also present and relatively more conspicuous in moderate heterogeneity scenarios than in high heterogeneity scenarios. Both model SE and empirical SE estimates are slightly reduced for σ_{drift} in all scenarios by using model 2 over model 1.

4.1.3 | Mean Squared Error

Table S2 (Supporting Information) summarizes the MSE estimates for all parameters of interest using model 0, model 1, and model 2 under all four scenarios. Keeping the number of trials the same, it appears that increasing the degree of heterogeneity has the effect of increasing the MSE as expected. Keeping the degree of heterogeneity the same, it appears that increasing the number of trials included in the NMA also has the expected effect of decreasing the MSE.

Comparing model 1 and 2 to model 0, the MSE estimates are reduced in both moderate heterogeneity and high heterogeneity scenarios for the relative treatment effect parameters, as seen in Figure S2 (Supporting Information). For treatments with large values of drifts in moderate heterogeneity scenarios, MSE reductions in using model 1 or 2 over model 0 are particularly significant. This can mostly be attributed to model 0 estimates of relative treatment effects being relatively more biased than the estimates produced by model 1 and 2 for treatments with large values of drifts in moderate heterogeneity scenarios, as discussed in the bias subsection. In high heterogeneity scenarios, the MSE reductions from using model 1 or 2 over model 0 are due partly to reductions in bias and partly to reductions in empirical SE. For σ_{RE}, the most significant reductions in MSE estimates result from using model 1 or 2 over model 0 in high heterogeneity scenarios. In moderate heterogeneity scenarios, MSE estimates are

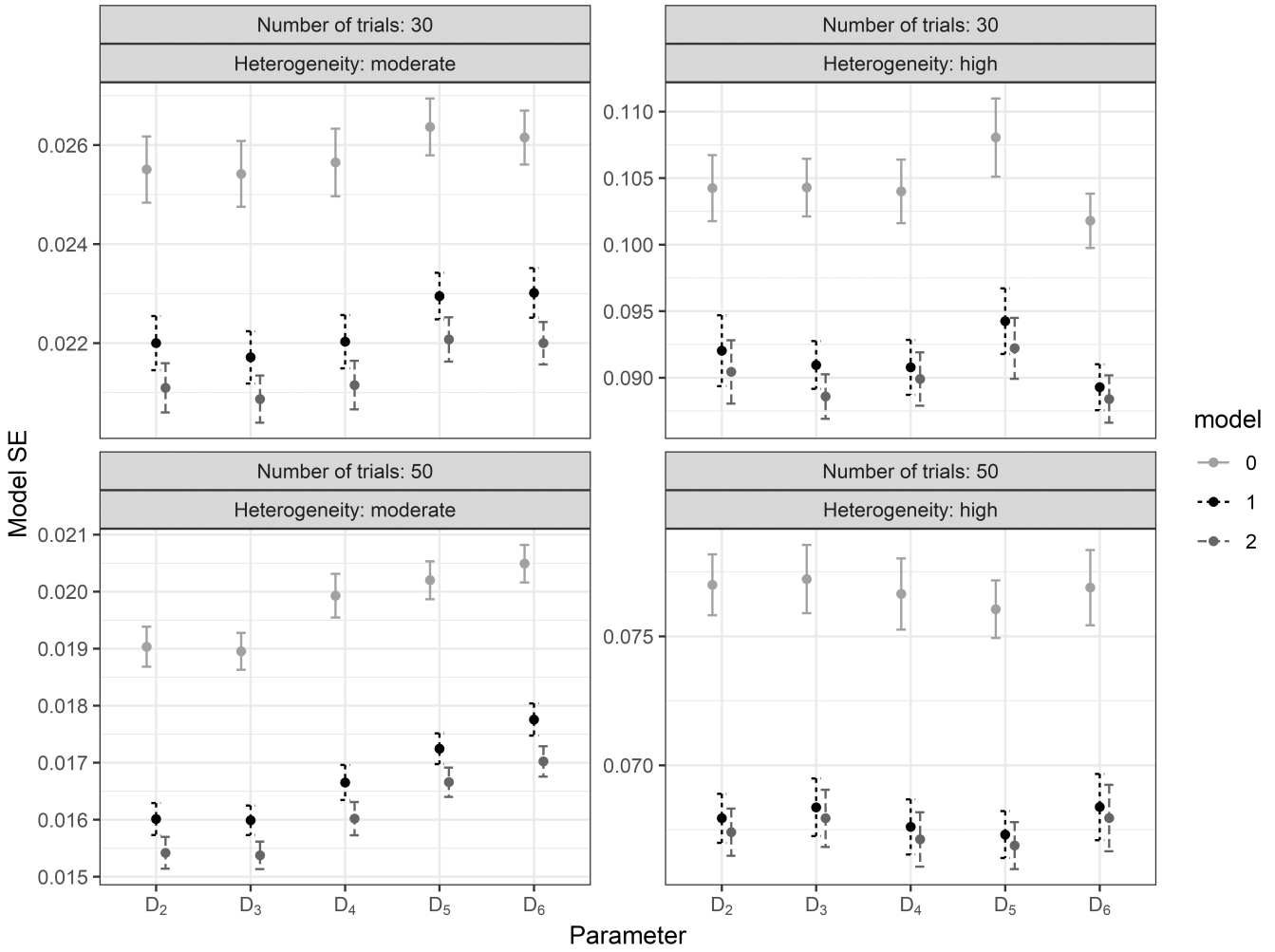

FIGURE 3 | Model SE for relative treatment effect parameters from model 0, model 1, and model 2 in all scenarios. Model SE shown with error bars extending one Monte Carlo SE above and below estimate.

only slightly reduced by using model 1 or 2 over model 0. Comparing model 1 and 2, MSE estimates either remain comparable, or are reduced slightly by using model 2 over model 1 in some cases. For instance, small reductions in MSE estimates are seen in several relative treatment effect and drift parameters under moderate heterogeneity scenarios.

4.1.4 | Mean Credible Interval Width

Table S3 (Supporting Information) summarizes the estimates of mean credible interval width for all parameters of interest using model 0, model 1, and model 2 under all four scenarios. Keeping the number of trials the same, it appears that increasing the degree of heterogeneity has the effect of increasing the mean CrI width as expected. Keeping the degree of heterogeneity the same, it appears that increasing the number of trials included in the NMA has the effect of decreasing the mean CrI width also as expected.

In all scenarios, both model 1 and 2 appear to produce CrI widths that are, on average, narrower than the CrI widths produced by model 0 for the relative treatment effect parameters, as shown in Figure S3 (Supporting Information). Model 2 generally performs slightly better than model 1 in terms of mean CrI width for all parameters of interest, with differences in estimates of mean CrI widths between the two models being more pronounced in moderate heterogeneity scenarios than in high heterogeneity scenarios given the Monte Carlo standard errors. Overall, model 2 appears to perform no worse than model 1, and model 2 has the potential to produce narrower CrI widths especially in moderate heterogeneity scenarios.

4.1.5 | Coverage

Table 5 summarizes the coverage estimates from using model 0, model 1, and model 2 in all four scenarios for all parameters of interest. In moderate heterogeneity scenarios, it is not unexpected that the coverage estimates from model 0 are significantly lower than the nominal value of 95% for treatments 5 and 6, since these two treatments feature large values of drifts that lead to biased relative treatment effect estimates. In high heterogeneity scenarios, it is not unexpected that the coverage estimates from model 0 are significantly lower than the nominal value of 95% for σ_{RE}, since model 0 tends to overestimate σ_{RE} to a great extent to compensate for assuming zero drift in the random walks.

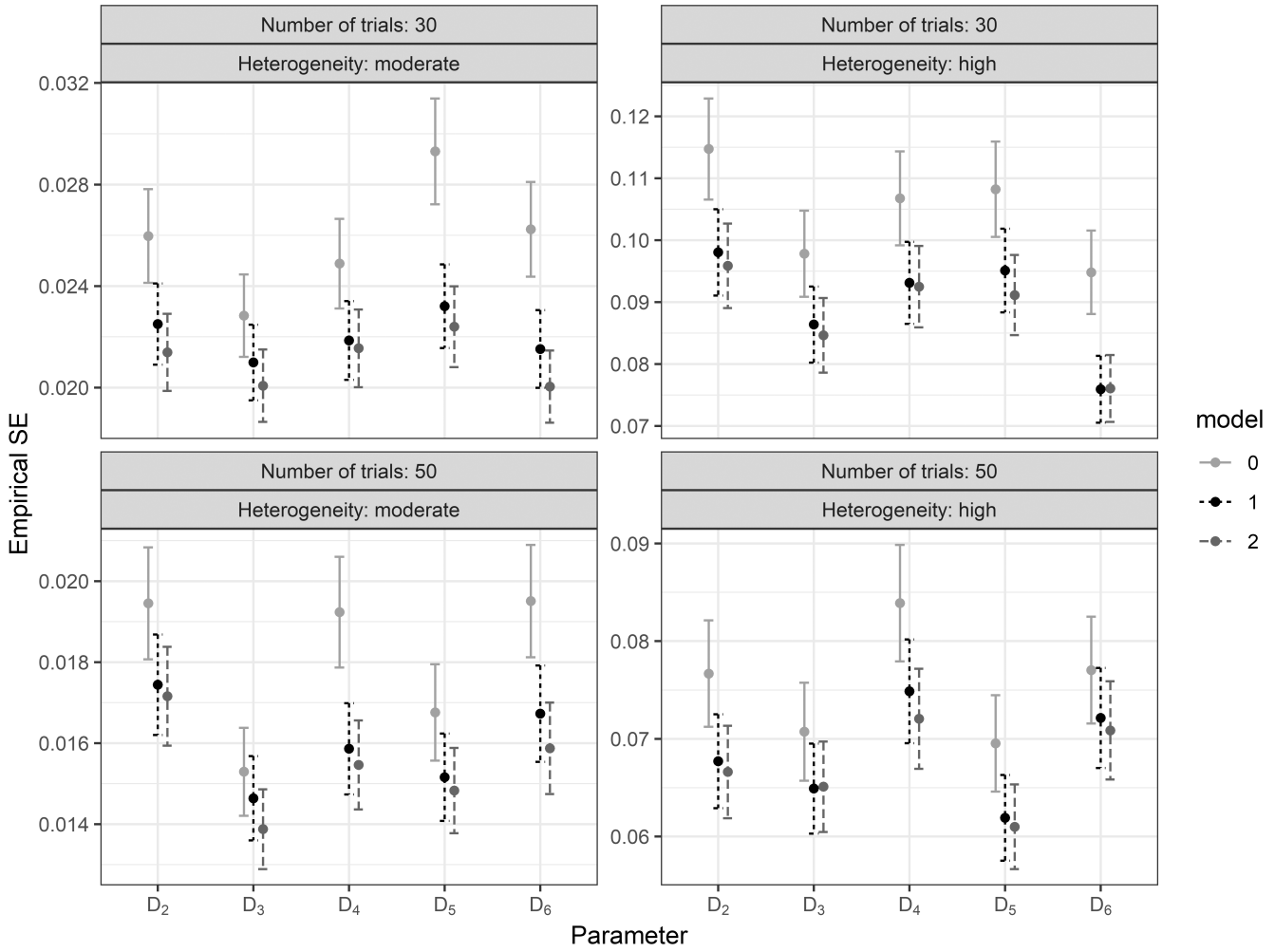

FIGURE 4 | Empirical SE for relative treatment effect parameters from model 0, model 1, and model 2 in all scenarios. Empirical SE shown with error bars extending one Monte Carlo SE above and below estimate.

Focusing on model 1 and 2, one way to check whether the results are in reasonable agreement with the nominal value of 95% given our choice of $n_{\text{sim}} = 100$ is to determine a reasonable range for the proportion of credible intervals containing the true parameter value based on the binomial distribution Bin(100, 0.95). We can compute the 2.5th and 97.5th percentiles of the Bin(100, 0.95) distribution (for instance using the qbinom function in R). This corresponds to the range (0.90, 0.99), in which the majority ($\approx 94\%$) of the coverage estimates for model 1 and 2 in Table 5 fall. This suggests that the coverage estimates for both proposed models across all scenarios are generally reasonable. For the one coverage estimate among the relative treatment effect parameters that falls below this range, the undercoverage can be explained by recognizing that the model SE underestimates the empirical SE in this case. When model SE underestimates the empirical SE, the CrI widths may run too narrow, leading to undercoverage. For σ_{RE}, some coverage estimates from using model 1 and 2 fall just below the lower end of the range (0.90, 0.99). Due to the slightly greater bias that exists in high heterogeneity scenarios (compared to the bias in moderate heterogeneity scenarios), slight undercoverage is expected for σ_{RE} when using model 1 and 2 in high heterogeneity scenarios. Comparing model 1 and 2 coverage estimates, there is little evidence to state that one model outperforms the other in terms of coverage. Figure 5 shows the comparisons of coverage estimates between model 0, model 1, and model 2 for the relative treatment effect parameters in all scenarios.

4.2 | Results for Motivating Example

The results from applying our proposed models to the motivating dataset are presented in Table 6, along with the results obtained from applying an existing random walk NMA model [5] to the same dataset for comparison. As the random walk NMA model of da Costa et al. is restricted to estimating relative treatment effects at the reference time point, which is selected from the set of time points at which outcome measurements exist, Table 6 displays the pooled effects at time point one for the random walk NMA model of da Costa et al. By contrast, for the two proposed models, the relative treatment effects are estimated at time point zero. This small discrepancy in time point is not expected to have a large impact, however, as the random walk NMA model of da Costa et al. implicitly assumes zero drift in the random walks and thus the pooled effects are not expected to be significantly different with a slightly different choice in reference time point.

TABLE 5 | Coverage estimates for relative treatment effects, drifts, and related between-trial heterogeneity parameters in all scenarios.

	Scenario 1 (30; moderate)			Scenario 2 (30; high)		
Parameter	**Model 0**	**Model 1**	**Model 2**	**Model 0**	**Model 1**	**Model 2**
D_2	0.91 (0.03)	0.91 (0.03)	0.95 (0.02)	0.93 (0.03)	0.92 (0.03)	0.91 (0.03)
D_3	0.90 (0.03)	0.94 (0.02)	0.92 (0.03)	0.95 (0.02)	0.94 (0.02)	0.95 (0.02)
D_4	0.90 (0.03)	0.93 (0.03)	0.93 (0.03)	0.93 (0.03)	0.95 (0.02)	0.92 (0.03)
D_5	0.81 (0.04)	0.92 (0.03)	0.92 (0.03)	0.94 (0.02)	0.97 (0.02)	0.97 (0.02)
D_6	0.65 (0.05)	0.94 (0.02)	0.95 (0.02)	0.95 (0.02)	0.99 (0.01)	0.99 (0.01)
B_1		0.95 (0.02)	0.93 (0.03)		0.94 (0.02)	0.94 (0.02)
B_2		0.95 (0.02)	0.96 (0.02)		0.96 (0.02)	0.95 (0.02)
B_3		0.96 (0.02)	0.97 (0.02)		0.98 (0.01)	0.98 (0.01)
B_4		0.95 (0.02)	0.95 (0.02)		0.94 (0.02)	0.93 (0.03)
B_5		0.91 (0.03)	0.94 (0.02)		0.98 (0.01)	0.97 (0.02)
B_6		0.96 (0.02)	0.97 (0.02)		0.92 (0.03)	0.92 (0.03)
σ_{RE}	0.92 (0.03)	0.93 (0.03)	0.97 (0.02)	0.52 (0.05)	0.87 (0.03)	0.89 (0.03)
σ_{drift}		0.92 (0.03)	0.98 (0.01)		0.91 (0.03)	0.93 (0.03)
	Scenario 3 (50; moderate)			**Scenario 4 (50; high)**		
Parameter	**Model 0**	**Model 1**	**Model 2**	**Model 0**	**Model 1**	**Model 2**
D_2	0.94 (0.02)	0.92 (0.03)	0.90 (0.03)	0.94 (0.02)	0.94 (0.02)	0.92 (0.03)
D_3	0.95 (0.02)	0.98 (0.01)	0.97 (0.02)	0.98 (0.01)	0.93 (0.03)	0.93 (0.03)
D_4	0.90 (0.03)	0.98 (0.01)	0.96 (0.02)	0.93 (0.03)	0.88 (0.03)	0.91 (0.03)
D_5	0.73 (0.04)	0.99 (0.01)	0.97 (0.02)	0.95 (0.02)	0.97 (0.02)	0.97 (0.02)
D_6	0.59 (0.05)	0.97 (0.02)	0.95 (0.02)	0.92 (0.03)	0.92 (0.03)	0.94 (0.02)
B_1		0.95 (0.02)	0.94 (0.02)		0.93 (0.03)	0.92 (0.03)
B_2		0.93 (0.03)	0.92 (0.03)		0.90 (0.03)	0.93 (0.03)
B_3		0.95 (0.02)	0.96 (0.02)		0.96 (0.02)	0.97 (0.02)
B_4		0.97 (0.02)	0.95 (0.02)		0.92 (0.03)	0.94 (0.02)
B_5		0.96 (0.02)	0.97 (0.02)		0.97 (0.02)	0.93 (0.03)
B_6		0.96 (0.02)	0.94 (0.02)		0.95 (0.02)	0.96 (0.02)
σ_{RE}	0.83 (0.04)	0.93 (0.03)	0.88 (0.03)	0.37 (0.05)	0.85 (0.04)	0.84 (0.04)
σ_{drift}		0.93 (0.03)	0.90 (0.03)		0.96 (0.02)	0.96 (0.02)

Note: Monte Carlo standard errors in parentheses. Scenarios labelled with: (number of trials; degree of heterogeneity).

Although the possibility exists for the proposed models to estimate the relative treatment effects at other time points using Equation (4), we have elected not to do so for this particular dataset, since the function outcome is only available at a single time point for numerous treatments in the network. This results in convergence issues for the drift parameters corresponding to these treatments when fitting the two proposed models. The way around this is to set the pooled drift parameters to the value of the pooled drift of placebo for all treatments with function (main outcome) measurements at only a single time point when applying the Gaussian random walk with drift NMA model. This ensures that the model does not produce unreasonable estimates of the treatment effects relative to placebo due to any assumed differences in drifts. For the cointegrated Gaussian random walks with drift NMA model, it is only necessary to set the pooled drift parameters to the value of the pooled drift of placebo for treatments for which both function and pain outcomes are only available at a single time point in each trial. When pain (auxiliary outcome) measurements are available at more than one time point in some trial for a treatment, even if function measurements are only available at a single time point, the cointegrated Gaussian random walks with drift NMA model is still able to estimate the pooled drift associated with the treatment under consideration. For both proposed models, since estimates of pooled effects at other time points depend on estimates of pooled drifts, and some of the pooled drift parameters are not estimable given the (lack of) available data, we do not recommend estimating pooled effects at time points other than the default time point zero when working with a dataset in which outcome measurements exist at only a single time point in each trial for multiple treatments in the network.

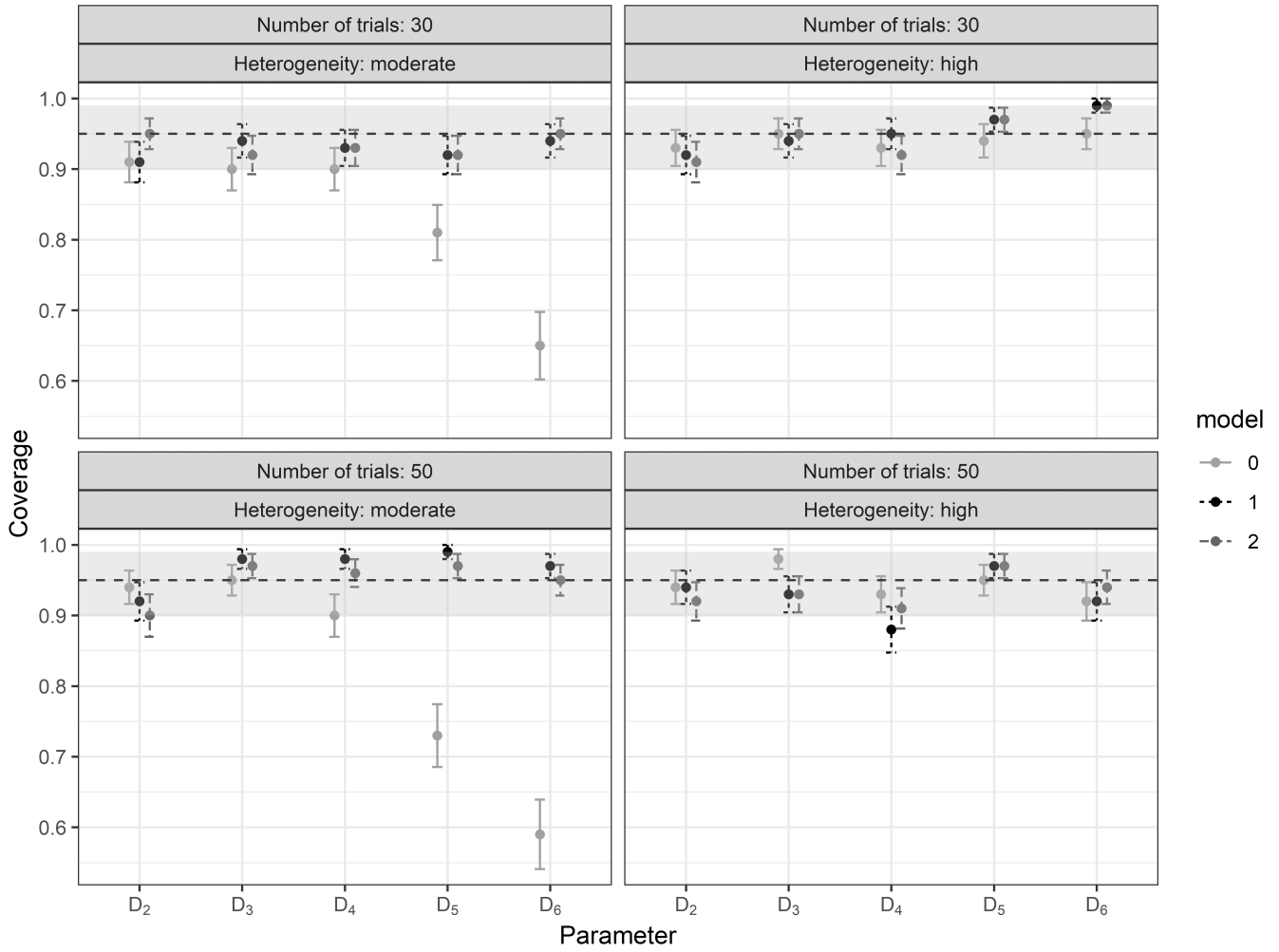

FIGURE 5 | Coverage for relative treatment effect parameters from model 0, model 1, and model 2 in all scenarios. Coverage shown with error bars extending one Monte Carlo SE above and below estimate. Shaded area indicates range between the 2.5th and 97.5th percentiles of a Bin(100, 0.95) distribution.

The treatment effects relative to placebo are shown in a forest plot in Figure 6 for the three NMA models. All three models yield broadly similar estimates of relative treatment effects, as seen in the overlapping credible intervals. However, due to the incorporation of drift terms in both proposed models and the borrowing of strength across outcomes in the cointegration model, we can obtain relative treatment effect estimates that potentially lead to clinically different conclusions in some cases. For instance, both of the proposed models yield pooled effects for celecoxib 100 mg versus placebo with credible intervals that cover the MCID of −0.37, whereas the random walk NMA model of da Costa et al. produces an interval estimate for the same effect that does not cover the MCID. On the other hand, for relative treatment effect estimates produced by the random walk NMA model of da Costa et al. with credible intervals that fall entirely above the MCID (i.e., for rofecoxib 25 mg, diclofenac 150 mg, and naproxen 1000 mg), the two proposed models also yield relative treatment effects with credible intervals that fall entirely above the MCID. Thus, clinically effective treatments with high probability to reach MCID according to the random walk NMA model of da Costa et al. remain clinically effective treatments with high probability to reach MCID using our proposed models.

With respect to credible interval widths, there are numerous examples of relative treatment effects with reductions in credible interval widths from using the cointegration model, such as for lumiracoxib 100 mg and diclofenac 150 mg versus placebo. Credible interval width reduction is more likely to occur when the auxiliary outcome has measurements at more time points across multiple trials than the main outcome does. There are also a few examples in which using the cointegration model leads to increased uncertainty in the relative treatment effect estimates, for instance for ibuprofen 1200 mg versus placebo. In this case, only two trials feature ibuprofen 1200 mg. Both trials are two-arm trials comparing ibuprofen 1200 mg to paracetamol 3000 mg and both trials feature function outcome measurements at only a single time point. There is also only one other treatment that paracetamol 3000 mg is compared to in the network and that treatment is placebo. This comparison only occurs in a single two-arm trial with both function and pain outcome measurements at just a single time point. Thus, where there is a lack of data on the main and auxiliary outcomes for certain treatments in the network, using the cointegration model to borrow information across outcomes could instead lead to increased uncertainty around the pooled effect estimates for the treatments in question. This, however, is

TABLE 6 | Relative treatment effects (versus placebo) on physical function estimated using the random walk network meta-analysis model of da Costa et al., the Gaussian random walk with drift network meta-analysis model, and the cointegrated Gaussian random walks with drift network meta-analysis model.

Treatment	Random walk NMA model of da Costa et al.	Gaussian random walk with drift NMA model	Cointegrated Gaussian random walks with drift NMA model
Paracetamol < 2000 mg	0.036 (−0.387, 0.458)	0.005 (−0.257, 0.270)	0.003 (−0.365, 0.359)
Paracetamol 3000 mg	−0.231 (−0.739, 0.285)	−0.270 (−0.605, 0.088)	−0.209 (−1.152, 1.039)
Paracetamol 3900-4000 mg	−0.141 (−0.224, −0.056)	−0.111 (−0.227, 0.007)	−0.148 (−0.254, −0.039)
Rofecoxib 12.5 mg	−0.323 (−0.408, −0.239)	−0.341 (−0.447, −0.241)	−0.349 (−0.445, −0.250)
Rofecoxib 25 mg	−0.483 (−0.562, −0.403)	−0.538 (−0.637, −0.434)	−0.514 (−0.604, −0.427)
Rofecoxib 50 mg	−0.281 (−0.727, 0.156)	−0.253 (−0.517, 0.009)	−0.265 (−0.663, 0.124)
Lumiracoxib 100 mg	−0.371 (−0.492, −0.250)	−0.371 (−0.533, −0.219)	−0.398 (−0.506, −0.268)
Lumiracoxib 200 mg	−0.386 (−0.482, −0.294)	−0.365 (−0.467, −0.263)	−0.400 (−0.497, −0.307)
Lumiracoxib 400 mg	−0.443 (−0.572, −0.316)	−0.410 (−0.549, −0.275)	−0.425 (−0.549, −0.297)
Etoricoxib 30 mg	−0.433 (−0.556, −0.315)	−0.433 (−0.580, −0.295)	−0.468 (−0.589, −0.335)
Etoricoxib 60 mg	−0.482 (−0.662, −0.295)	−0.471 (−0.696, −0.257)	−0.502 (−0.680, −0.292)
Diclofenac 70 mg	−0.233 (−0.702, 0.219)	−0.226 (−0.531, 0.077)	−0.198 (−0.580, 0.207)
Diclofenac 100 mg	−0.419 (−0.754, −0.113)	−0.395 (−0.599, −0.195)	−0.854 (−1.382, −0.568)
Diclofenac 150 mg	−0.526 (−0.659, −0.386)	−0.587 (−0.730, −0.448)	−0.601 (−0.716, −0.474)
Celecoxib 100 mg	−0.196 (−0.346, −0.045)	−0.379 (−0.518, −0.246)	−0.282 (−0.405, −0.171)
Celecoxib 200 mg	−0.348 (−0.404, −0.292)	−0.377 (−0.455, −0.303)	−0.394 (−0.457, −0.336)
Celecoxib 400 mg	−0.316 (−0.448, −0.183)	−0.468 (−0.635, −0.299)	−0.463 (−0.601, −0.330)
Naproxen 750 mg	−0.024 (−0.285, 0.240)	0.015 (−0.473, 0.495)	−0.009 (−0.397, 0.425)
Naproxen 1000 mg	−0.456 (−0.534, −0.374)	−0.571 (−0.672, −0.469)	−0.577 (−0.655, −0.493)
Ibuprofen 1200 mg	−0.537 (−1.114, 0.043)	−0.512 (−0.910, −0.105)	−0.748 (−1.754, 0.330)
Ibuprofen 2400 mg	−0.384 (−0.502, −0.265)	−0.420 (−0.572, −0.269)	−0.448 (−0.583, −0.310)

Note: Pooled effects shown are standardized effects with 95% credible intervals.

more an issue of data scarcity than it is a problem with the cointegrated Gaussian random walks with drift NMA model itself.

5 | Discussion

In this paper, two novel Bayesian models for conducting NMA with one or two continuous outcomes measured at multiple time points have been presented. In the case of a single outcome, the Gaussian random walk with drift NMA model makes use of a Gaussian random walk with drift to model the mean standardized effects over time for each treatment in a RCT. In contrast to the random walk NMA model used by da Costa et al. [5] to synthesize pain or function data from OA trials, the Gaussian random walk with drift NMA model presented in this paper does not make the limiting implicit assumption that the drift is zero in the random walks. By including these drift terms in the model, not only can we account for potential time trends in treatment effects when estimating the relative treatment effects at the reference time point ($t = 0$), but also we can estimate relative treatment effects at other time points by using the estimates of drifts produced by the model in combination with the estimates of relative treatment effects at the reference time point. In fact, we are able to make predictions of relative treatment effects at precisely the same time points as the discrete time points we have chosen for the Gaussian random walk with drift in the model.

Since a Gaussian random walk with drift is a discrete time process, some care is required to select the appropriate time points after the start of administration of treatment at which to extract outcome data so that a Gaussian random walk with drift under this particular choice of time points would make for a reasonable approximation to the behaviour of treatment effects over time. For example, in the case of OA trials, the seven time points that could be chosen to correspond to the discrete time points of the Gaussian random walks with drift are 1 week, 2 weeks, 4 weeks, 6 weeks, 3 months, 6 months, and 12 months after the start of treatment [5]. The time points are chosen to be much more closely spaced initially and to increase in spacing with each additional time point. This is to reflect the rapid decline in pain (or rapid decline in physical disability) in the first few weeks after starting treatment, followed by a plateauing of pain (or physical function) scores so that it would take dramatically longer, in terms of calendar time, to achieve a similar magnitude of improvement as those observed in the first few weeks. Similar curvilinear patterns in time-course of response are seen in other application areas, such as in depression [25] or diabetes [26], and are often discussed and modelled in pharmacokinetics and pharmacodynamics [27]. For other fields, choice of time points would need to be determined by

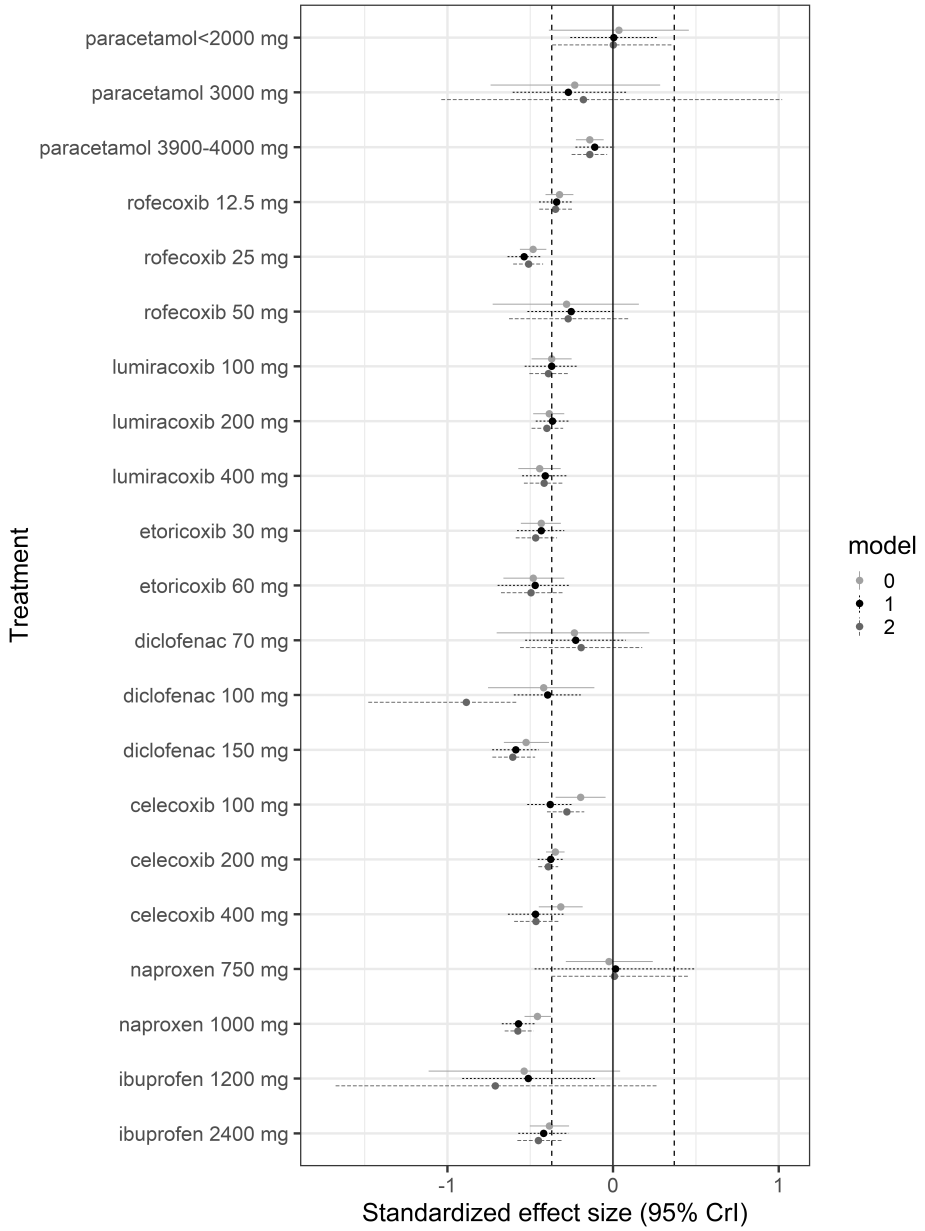

FIGURE 6 | Forest plot of relative treatment effects estimated using the random walk NMA model of da Costa et al. (model 0), the Gaussian random walk with drift NMA model (model 1), and the cointegrated Gaussian random walks with drift NMA model (model 2). Area between vertical dashed lines indicates relative treatment effects below the minimal clinically important difference.

the investigator based on the outcome measured and the expected behaviour of treatment effects over time. Unlike other NMA models for longitudinal data that focus on using flexible continuous functions to model the evolution of outcome over time, such as the fractional polynomial NMA model by Jansen et al. [3], the Gaussian random walk with drift NMA model opts for a different approach that focuses on using Gaussian random walks with drift to model the treatment effects over time, assuming a suitable set of discrete time points has already been identified.

If the RCTs forming the evidence base feature two outcomes that are measured over time and that can be reasonably assumed to be cointegrated random walks, then the cointegrated Gaussian random walks with drift NMA model can be used to synthesize the data. Cointegration of two random walks with drift arises when two random walks share a common stochastic trend, which results in two random walks that move in tandem. For OA trials, pain and physical function outcomes exhibit such a tendency to move together. This tendency can be explained by noting that patients who experience greater pain in their joints can be expected to experience more functional restrictions involving those joints whether due to biomechanical limitations or to patients' voluntary or involuntary restriction of activity [28]; while patients with impairments in physical function can be expected to experience more joint stiffness that can exacerbate pain in the affected joints [29]. By making use of the concept of cointegration of two random walks with drift, we are able to integrate data on a second outcome to aid estimation of parameters of interest for the first outcome without needing to resort to more complex multivariate NMA models. This

allows us to avoid potential obstacles that can sometimes hamper a multivariate NMA modelling approach, such as needing individual participant data to estimate within-study correlations between outcomes [9, 30], or requiring particular matrix decomposition and reparameterization techniques to allow flexibility in specifying a prior distribution for the between-study covariance matrix [9, 19].

For both NMA models presented in this paper, the exchangeability assumptions made for the relative treatment effects, δ_{ibk}, and for the drift terms, β_{ik}, are not the same. For the relative treatment effects, the exchangeability assumption is the usual one seen in the contrast-based approach wherein the trial-specific relative effects come from a common random effects distribution that applies to all trials, regardless of whether a given treatment contrast actually appears in a trial [31]. For the drift terms, however, the exchangeability assumption is not a contrast-based one but rather an arm-based one. That is, the trial-specific drifts are assumed to come from a common distribution centred around a mean drift that is particular to a treatment, rather than a treatment contrast. The choice to make an arm-based exchangeability assumption for the drifts instead of a contrast-based one as we do for the relative treatment effects is driven by the need for the drift term to be present in specifying the mean of the Gaussian random walk with drift, m_{ikt}, even in the baseline case when $k = b$. Although a partial contrast-based approach is still possible in this case, it requires treating the drift for the treatment corresponding to an arbitrary choice of the control arm in a trial differently than the drifts for other treatments in the trial. Thus, to maintain a consistent modelling of the drift terms, we have chosen to use an arm-based approach in this case. While it is true that the arm-based approach requires a stronger exchangeability assumption than the contrast-based approach, and such abandoning of the principle of concurrent control could potentially produce biased estimates, little harm is likely to be done in practice [32]. Others have presented their NMA methods using both contrast-based and arm-based parameterizations and have advocated for the use of arm-based approaches due to ease of parameter interpretation, as well as for situations when absolute measures of effect are of interest [33]. In our case, it is not unreasonable to assume that the time trends of an outcome under a particular treatment, represented by the drift terms, are exchangeable across trials. Because the relative treatment effects still follow contrast-based modelling, and relative drifts are used to obtain the relative treatment effect estimates at other time points, a stronger exchangeability assumption made for the drifts should not be an issue in practice.

Regarding the comparison of performance between the Gaussian random walk with drift NMA model and the cointegrated Gaussian random walks with drift NMA model, this simulation study has shown that both models produce unbiased estimates of relative treatment effects and drifts, with empirical estimates of bias for both proposed models being similarly close to zero and negligible in each scenario. In terms of model SE, the cointegration model has better performance than the single outcome model in scenarios with moderate heterogeneity. A similar observation applies to the empirical SE as well, although the improvements are less pronounced than those observed for model SE given the Monte Carlo standard errors. In high heterogeneity scenarios, the cointegration model does not appear to offer significant improvements over the single outcome model in either model SE or empirical SE, though small reductions in model SE and empirical SE can still occur for some parameters. With respect to relative percent error in model SE, both models have comparable performance in all scenarios. With respect to MSE, the cointegration model slightly outperforms the single outcome model in the sense that MSE estimates either remain comparable or are reduced slightly by using the cointegration model. In terms of mean CrI width, the cointegration model outperforms the single outcome model, with the improvements again being more pronounced in moderate heterogeneity scenarios. Finally, with respect to coverage, both models yield coverage estimates that mostly fall within a reasonable range. Overall, we have shown that the cointegrated Gaussian random walks with drift NMA model performs on par with, if not slightly better than, the Gaussian random walk with drift NMA model. We have also shown that both proposed models outperform, on the basis of various performance measures, the random walk NMA model of da Costa et al. when the data-generating mechanism incorporates cointegrated Gaussian random walks with drift. More research is needed, however, to understand to what extent the performance of these models might be diminished if the true data-generating mechanism is not approximated well by the cointegrated Gaussian random walks with drift model.

In simulating the OA trials data, we have used an algorithm that produces measurements for the main outcome at fewer time points than measurements for the auxiliary outcome. Specifically, at each time point with an auxiliary outcome measurement in each trial, the probability of the corresponding main outcome measurement being missing is set to 10%. This is done for the purpose of demonstrating the advantage of using the cointegrated Gaussian random walks with drift NMA model when a second outcome is available and measured at potentially a few more time points than the first outcome. This type of scenario occurs in OA trials if we consider physical function as the main outcome and pain as the auxiliary outcome, since many OA trials report physical function outcome measurements at a subset of the time points at which pain outcome measurements are reported. In the alternative case in which the main outcome is measured at potentially more time points than (or the same set of time points as) the auxiliary outcome, improvements in interval estimates using the cointegration model over the single outcome model are expected to be minimal. Riley et al. discussed the low impact that borrowing of strength across outcomes is likely to have in situations with low percentage of missing outcomes in the context of fitting multivariate meta-analysis models [34], and we feel that a similar limitation applies in the context of fitting the cointegrated Gaussian random walks with drift NMA model. At the same time, we do not anticipate the cointegration model to perform worse than the single outcome model even if the auxiliary outcome has not been measured at more time points than the main outcome, provided that sufficient data from multiple time points are available on either outcome to inform the estimation of relative treatment effects and drifts.

A limitation of the Gaussian random walk with drift NMA model is that, to estimate the pooled drift associated with each treatment, it is necessary for at least one trial featuring each treatment to report the main outcome at more than one time point. If there exists a treatment such that each trial that features this treatment

only reports the main outcome at one time point, then the MCMC chains would fail to converge and the model would not be able to estimate the pooled drift. A similar limitation applies to the cointegrated Gaussian random walks with drift NMA model, with the condition being that it is necessary for at least one trial featuring each treatment to report either the main or the auxiliary outcome at more than one time point in order to estimate the pooled drifts. One possible approach for dealing with this limitation is to include an additional eligibility criterion when screening trials for inclusion in a systematic review. For example, in the case of OA trials, this eligibility criterion could require a trial to have at least one of pain or function outcome be reported at two or more time points of interest, if the investigator plans to use the cointegrated Gaussian random walks with drift NMA model for the analysis.

Another limitation of the cointegrated Gaussian random walks with drift NMA model is that, to obtain estimates for all parameters of interest, it is not sufficient to fit the model only once, unlike a multivariate NMA model. Instead, as is the case for univariate NMA models, the cointegration model must be fitted twice to obtain the relative treatment effect estimates for both outcomes, each of which would serve as the main outcome and the other as the auxiliary outcome in each run of the cointegration model. In our experience, fitting the cointegration model requires more computational time than fitting the single outcome model. However, we consider this a small price to pay in exchange for potentially narrower interval estimates, even if the performance is only marginally improved. Jackson et al. also discussed the potential for borrowing of strength to yield little to no gain in precision for the pooled effect estimates in the context of fitting multivariate meta-analysis models [18]. We believe the cointegrated Gaussian random walks with drift NMA model serves as a compromise between univariate and multivariate NMA models: it allows the investigator to avoid some of the difficulties associated with applying multivariate NMA models while still yielding potential gains in precision that multivariate NMA models can offer over univariate NMA models.

Acknowledgments

This work is part of the lead author's PhD research and was partially supported by the Arthritis Society (Grant Number: YIO-17-0164).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The simulated data that support the findings of this study are openly available in figshare at http://doi.org/10.6084/m9.figshare.22237147.

References

1. G. Lu, A. E. Ades, A. J. Sutton, N. J. Cooper, A. H. Briggs, and D. M. Caldwell, "Meta-Analysis of Mixed Treatment Comparisons at Multiple Follow-Up Times," *Statistics in Medicine* 26, no. 20 (2007): 3681–3699.

2. H. A. Dakin, N. J. Welton, A. E. Ades, S. Collins, M. Orme, and S. Kelly, "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* 30, no. 20 (2011): 2511–2535.

3. J. P. Jansen, M. C. Vieira, and S. Cope, "Network Meta-Analysis of Longitudinal Data Using Fractional Polynomials," *Statistics in Medicine* 34, no. 15 (2015): 2294–2311.

4. D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde, "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 64, no. 4 (2002): 583–639.

5. B. R. da Costa, S. Reichenbach, N. Keller, et al., "Effectiveness of Non-Steroidal Anti-Inflammatory Drugs for the Treatment of Pain in Knee and Hip Osteoarthritis: A Network Meta-Analysis," *Lancet* 390, no. 10090 (2017): e21–e33.

6. H. Pedder, S. Dias, M. Bennetts, M. Boucher, and N. J. Welton, "Modelling Time-Course Relationships With Multiple Treatments: Model-Based Network Meta-Analysis for Continuous Summary Outcomes," *Research Synthesis Methods* 10, no. 2 (2019): 267–286.

7. O. Efthimiou, D. Mavridis, R. D. Riley, A. Cipriani, and G. Salanti, "Joint Synthesis of Multiple Correlated Outcomes in Networks of Interventions," *Biostatistics* 16, no. 1 (2015): 84–97.

8. R. D. Riley, J. R. Thompson, and K. R. Abrams, "An Alternative Model for Bivariate Random-Effects Meta-Analysis When the Within-Study Correlations Are Unknown," *Biostatistics* 9, no. 1 (2008): 172–186.

9. F. A. Achana, N. J. Cooper, S. Bujkiewicz, et al., "Network Meta-Analysis of Multiple Outcome Measures Accounting for Borrowing of Information Across Outcomes," *BMC Medical Research Methodology* 14, no. 1 (2014): 92.

10. E. Waddingham, P. M. Matthews, and D. Ashby, "Exploiting Relationships Between Outcomes in Bayesian Multivariate Network Meta-Analysis With an Application to Relapsing-Remitting Multiple Sclerosis," *Statistics in Medicine* 39, no. 24 (2020): 3329–3346.

11. S. Dagenais, S. Garbedian, and E. K. Wai, "Systematic Review of the Prevalence of Radiographic Primary Hip Osteoarthritis," *Clinical Orthopaedics and Related Research* 467, no. 3 (2009): 623–637.

12. S. Dias, A. J. Sutton, A. E. Ades, and N. J. Welton, "Evidence Synthesis for Decision Making 2: A Generalized Linear Modeling Framework for Pairwise and Network Meta-Analysis of Randomized Controlled Trials," *Medical Decision Making* 33, no. 5 (2013): 607–617.

13. G. Lu and A. E. Ades, "Assessing Evidence Inconsistency in Mixed Treatment Comparisons," *Journal of the American Statistical Association* 101, no. 474 (2006): 447–459.

14. N. J. Cooper, A. J. Sutton, D. Morris, A. E. Ades, and N. J. Welton, "Addressing Between-Study Heterogeneity and Inconsistency in Mixed Treatment Comparisons: Application to Stroke Prevention Treatments in Individuals With Non-Rheumatic Atrial Fibrillation," *Statistics in Medicine* 28, no. 14 (2009): 1861–1881.

15. M. Plummer, "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing* (R Foundation for Statistical Computing, 2003).

16. D. Lunn, D. J. Spiegelhalter, A. Thomas, and N. Best, "The BUGS Project: Evolution, Critique and Future Directions," *Statistics in Medicine* 28, no. 25 (2009): 3049–3067.

17. G. S. Maddala and I. M. Kim, *Unit Roots, Cointegration, and Structural Change*, 1st ed. (Cambridge University Press, 1999).

18. D. Jackson, R. D. Riley, and I. R. White, "Multivariate Meta-Analysis: Potential and Promise," *Statistics in Medicine* 30, no. 20 (2011): 2481–2498.

19. Y. Wei and J. P. Higgins, "Bayesian Multivariate Meta-Analysis With Multiple Outcomes," *Statistics in Medicine* 32, no. 17 (2013): 2911–2934.

20. R. Little, "Calibrated Bayes, for Statistics in General, and Missing Data in Particular," *Statistical Science* 26, no. 2 (2011): 162–174.

21. A. Gelman and D. B. Rubin, "Inference From Iterative Simulation Using Multiple Sequences," *Statistical Science* 7, no. 4 (1992): 457–472.

22. M. Ponce, R. van Zon, S. Northrup, et al., "Deploying a Top-100 Supercomputer for Large Parallel Workloads: The Niagara Supercomputer," in *Proceedings of the Practice and Experience in Advanced Research Computing 2019: Rise of the Machines (Learning)* (Association for Computing Machinery, 2019).

23. S. Wandel, P. Jüni, B. Tendal, et al., "Effects of Glucosamine, Chondroitin, or Placebo in Patients With Osteoarthritis of Hip or Knee: Network Meta-Analysis," *BMJ (Clinical Research Ed.)* 341 (2010): c4675.

24. T. P. Morris, I. R. White, and M. J. Crowther, "Using Simulation Studies to Evaluate Statistical Methods," *Statistics in Medicine* 38, no. 11 (2019): 2074–2102.

25. E. M. Whyte, M. A. Dew, A. Gildengers, et al., "Time Course of Response to Antidepressants in Late-Life Major Depression: Therapeutic Implications," *Drugs & Aging* 21, no. 8 (2004): 531–554.

26. M. Pfohl, S. Pscherer, A. Fritsche, et al., "Predictors of Treatment Response in Type-2 Diabetes Patients Initiating Basal-Supported Oral Therapy With Insulin Glargine 100 U/mL: A Sub-Analysis of the Titration and OPtimisation (TOP) Registry," *Diabetes, Obesity & Metabolism* 21, no. 9 (2019): 2169–2173.

27. M. Rowland and T. N. Tozer, *Clinical Pharmacokinetics and Pharmacodynamics: Concepts and Applications*, 4th ed. (Lippincott William & Wilkins, 2011).

28. H. R. Barthel, J. H. Peniston, M. B. Clark, M. S. Gold, and R. D. Altman, "Correlation of Pain Relief With Physical Function in Hand Osteoarthritis: Randomized Controlled Trial Post Hoc Analysis," *Arthritis Research & Therapy* 12, no. 1 (2010): R7.

29. R. Gooberman-Hill, G. Woolhead, F. MacKichan, S. Ayis, S. Williams, and P. Dieppe, "Assessing Chronic Joint Pain: Lessons From a Focus Group Study," *Arthritis and Rheumatism* 57, no. 4 (2007): 666–671.

30. S. Bujkiewicz, J. R. Thompson, A. J. Sutton, et al., "Multivariate Meta-Analysis of Mixed Outcomes: A Bayesian Approach," *Statistics in Medicine* 32, no. 22 (2013): 3926–3943.

31. S. Dias, N. J. Welton, D. M. Caldwell, and A. E. Ades, "Checking Consistency in Mixed Treatment Comparison Meta-Analysis," *Statistics in Medicine* 29, no. 7–8 (2010): 932–944.

32. S. Senn, "Hans Van Houwelingen and the Art of Summing Up," *Biometrical Journal* 52, no. 1 (2010): 85–94.

33. H. Hong, H. Chu, J. Zhang, and B. P. Carlin, "A Bayesian Missing Data Framework for Generalized Multiple Outcome Mixed Treatment Comparisons," *Research Synthesis Methods* 7, no. 1 (2016): 6–22.

34. R. D. Riley, D. Jackson, G. Salanti, et al., "Multivariate and Network Meta-Analysis of Multiple Outcomes and Multiple Treatments: Rationale, Concepts, and Examples," *BMJ (Clinical Research Ed.)* 358 (2017): j3932.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1.** Supporting Information.