Methods Research Report

A Bayesian Missing Data Framework for Multiple Continuous Outcome Mixed Treatment Comparisons

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • **www.ahrq.gov**

Methods Research Report

A Bayesian Missing Data Framework for Multiple Continuous Outcome Mixed Treatment Comparisons

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-2007-10064-I

Prepared by:
Minnesota Evidence-based Practice Center
Minneapolis, MN

Investigators:
Hwanhee Hong, Ph.D. candidate
Bradley P. Carlin, Ph.D.
Haitao Chu, M.D., Ph.D.
Tatyana A. Shamliyan, M.D., M.S.
Shiyi Wang, M.D., Ph. D.
Robert L. Kane, M.D.

AHRQ Publication No. 13-EHC004-EF
January 2013

This report is based on research conducted by the Minnesota Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2007-10064-I). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

Suggested citation: Hong H, Carlin BP, Chu H, Shamliyan TA, Wang S, Kane RL. A Bayesian Missing Data Framework for Multiple Continuous Outcome Mixed Treatment Comparisons. Methods Research Report. (Prepared by the Minnesota Evidence-based Practice Center under Contract No. 290-2007-10064-I.) AHRQ Publication No.13-EHC004-EF. Rockville, MD: Agency for Healthcare Research and Quality; January 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality. The reports undergo peer review prior to their release as a final report.

We welcome comments on this Methods Research Project. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.hhs.gov.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Parivash Nourjah, Ph.D.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

A Bayesian Missing Data Framework for Mixed Multiple Treatment Comparisons

Structured Abstract

Objectives. Bayesian statistical approaches to mixed treatment comparisons (MTCs) are becoming more popular due to their flexibility and interpretability. Many randomized clinical trials report multiple outcomes with possible inherent correlations, but there is little previous work in modeling them statistically. We aimed to build on existing hierarchical modeling and missing data methods to obtain novel and improved Bayesian approaches to MTCs for multiple continuous outcomes.

Data sources. We reviewed randomized clinical trials published in English after 1979 that examined physical therapy interventions for community-dwelling adults with knee pain secondary to osteoarthritis (OA). After screening, 84 randomized trials met the inclusion/exclusion criteria, reporting variously on knee pain, disability, quality of life, and functional outcomes.

Methods. After a review of existing hierarchical Bayesian methods for MTCs with a single continuous outcome, we introduce novel Bayesian approaches for multiple continuous outcomes (here, pain and disability) simultaneously, rather than in separate MTC analyses, by generalizing existing models to treat missing data the same as unknown parameters and to incorporate correlation structure between outcomes. We also introduce an arm-based model that is less constrained than existing models. We produce Bayesian treatment ranks based on a sensible scoring system incorporating weights for the multiple outcomes. We also offer simulation studies to check our method's Type I error, power, and the probability of incorrectly selecting the best treatment.

Results. In our OA data analysis, while all the models gave similar goodness of fit, they yielded different best treatments, with aerobic exercise emerging as best according to the older models, but proprioception exercise being preferred by our weighted ranking models. Still, few statistically significant differences between treatments were observed. Our missing data approaches had better power and Type I error than previous Bayesian methods in our simulation study. Ignoring missing data or correlation between outcomes can produce biased MTC estimates leading to high Type I error and low power, especially when the data from missing treatments depend on the observed data.

Conclusions. Our missing data approaches appear preferable for incorporating missing data and correlation structure in MTC modeling, to traditional contrast-based approaches, and thus in obtaining more precise and robust parameter estimates.

Key Messages

- Since researchers often choose study arms based on previous trials, it is important to consider any unobserved treatment arms in an MTC as missing data and subsequently use Bayes' Rule to learn about the treatments' relative relationships. This makes it easier to assign prior distributions on random effects and delivers better statistical inference.
- Our arm-based models are less constrained than previous contrast-based models and can thus yield parameters with more straightforward interpretations, especially in the presence of correlations between outcomes.

Contents

Figures

Introduction

Mixed treatment comparisons (MTCs) are meta-analytic statistical techniques that incorporate the findings from several studies, where in most cases none of the studies compared *all* the treatments at one time, to address the comparative effectiveness and safety of interventions accounting for all sources of data.[1,2] In the MTC data framework, since few head-to-head comparisons are available, we must rely on indirect comparisons, typically each investigated treatment against a control or a standard treatment. The biggest assumption in MTCs is exchangeability among studies; that is, any ordering of the true treatment effects across studies is equally likely a priori. In addition, populations in selected studies should be similar to the *target* population for valid clinical interpretation.[3]

Bayesian hierarchical statistical meta-analysis for MTCs with a single binary outcome has been investigated actively since the 1980s.[4-8] However, compared with the binary outcome setting, there has been comparatively little development in Bayesian MTCs for continuous outcomes: we found only a few published papers discussing a simple Gaussian Bayesian hierarchical model using the standard approach.[9-11]

Our interest in Bayesian MTC methods for multiple continuous outcomes is motivated by a systematic literature review at the Minnesota Evidence-based Practice Center (EPC) that investigated the effectiveness of physical therapies on chronic pain secondary to knee osteoarthritis (OA) for community-dwelling adults.[12] OA treatments aim to reduce or control pain, improve physical function, prevent disability, and enhance quality of life. We recorded means of measured pain, disability, function, and quality of life scores associated with various physical therapy interventions from randomized studies.

As our OA data contain many studies reporting multiple outcomes, and measured on the same subjects, correlations across arms and outcomes are likely, but this case has not been discussed much in the literature.[13,14] For example, similar types of drugs or physical therapies may tend to behave similarly inducing correlated results, and multiple outcomes also can induce correlations (e.g., subjects with severe pain would be more likely to have disability).

Most randomized controlled trials (RCTs) include only two or three treatment arms, including a control group, due to limited resources. This results in extremely sparse data for MTCs when used across all possible treatments. Suppose that we can calculate the missingness rate as the summation of the ratio of the number of missing arms to the total number of treatments across all studies. Then, the missingness rate is 40 to 60 percent when we compare 5 treatments, and the rate could increase up to about 70 percent if 10 treatments are considered. Lu and Ades's approach,[8] a standard MTC model, uses only the observed data. However, we can borrow strength from those missing data after imputing them in a Bayesian hierarchical model that accounts for between-treatment and between-outcome correlations using Markov chain Monte Carlo (MCMC) algorithms. Especially when the missingness does not occur randomly but depends on some observed or unobserved information, ignoring such missing data can cause biased estimators.[15]

In this report we review existing MTC models and propose novel Bayesian missing data approaches to combine multiple continuous outcomes. The main objectives are to (1) impute unobserved arms by considering them as unknown parameters which can be modeled along with the other unknown, (2) incorporate between-treatment or between-outcome correlations, and (3) introduce an arm-based approach that features fewer constraints than standard contrast-based methods. We also rank the treatments with a sensible scoring system incorporating such multiple outcomes. We apply our models to the OA data and interpret our findings. Finally, we include a

simulation study to investigate the performance of our methods in terms of Type I error, power, and the probability of incorrectly selecting the best treatment.

Methods

OA Data

We reviewed publications in English after 1979 that examined physical therapy interventions for community dwelling adults with knee pain secondary to osteoarthritis. A total of 4,266 references were retrieved.[12] After screening out studies that contained no eligible exposure, target population, outcomes, or associative hypothesis tested, 422 references were included in our review. Knee pain, disability, quality of life, and functional outcomes after physical therapy interventions were reported in 193 RCTs; 84 of those met the study inclusion/exclusion criteria given in the next paragraph. Because definitions of physical therapy interventions and outcomes varied dramatically among studies, only a small proportion of comparisons met these criteria.

Inclusion/exclusion criteria involved the following aspects. First, comparators should include no active treatment, usual care (education), sham stimulation (placebo), or other therapy intervention (that is, active-active trials were not excluded). Eligible patient-centered outcomes were knee pain, disability, quality of life, perceived health status, and global assessments of treatment effectiveness. The target population was adults with knee pain secondary to knee osteoarthritis in outpatient settings, including home-based therapy. Chronic OA was defined as meeting diagnostic criteria and having symptoms of OA for >2 months. We excluded populations with knee OA who had knee arthroplasty on the "study limb" within 6 months before the study, osteonecrosis, acute knee injuries, inflammatory arthritis, arthritis secondary to systemic disease, and physical therapy treatment combined with drug treatments. Since all included studies are applied to the same inclusion and exclusion criteria, we assume that all populations are similar to each other.

For the present analysis, we selected the pain and disability outcomes as primary and secondary outcomes, respectively, resulting in the inclusion of 54 RCTs. Table 1 displays the data from these 54 RCTs, comprising aggregated continuous outcomes (sample mean and standard deviation [SD]) measuring the level of pain and disability after physical therapies using various standard scores. The OA data compare eight physical therapies (low intensity diathermy, high intensity diathermy, electrical stimulation, aerobic exercise, aquatic exercise, strength exercise, proprioception exercise, and ultrasound treatment) and three reference therapies (no treatment, placebo, and education). Under proprioception exercise, we also included tai chi and balance exercise. Most studies reported treatment outcomes at a single followup time, but when a study investigated outcomes at multiple followup times, we selected the one most commonly reported for that treatment. To measure the pain outcome, the Western Ontario MacMaster (WOMAC), Visual Analogue Scale (VAS), Arthritis Impact Measurement Scale (AIMS), and other standard scores were used. For the disability outcome, the measurement tools included the WOMAC total, Medical Outcome Study (MOS) 36-Item Short-Form Health Survey (SF-36 physical function), AIMS, Health Assessment Questionnaire (HAQ), and Knee Injury and Osteoarthritis Outcome Score (KOOS). Although these scores do not share the same scale and differ in a few details, in general they do measure outcomes equivalently, and all of their scales cover the same qualitative ranges (from "no pain" to "extreme pain" for pain measurements, and from "no impairment" to "profound impairment" for disability). The scores they yield also tend to be highly correlated when reported for the same subjects.[16-18] Because the scores' different scales make their values incomparable, we rescaled the mean scores to range from 0 to 10, where small values indicate better condition, and called this the rescaled score. We also recalculated the SDs based on the transformation of the mean score, and call this the rescaled SD. We remark that

we have no reason to doubt the appropriation of linear retransformation here, but our methods apply equally well under nonlinear transformations if more appropriate clinically.

Among the 54 studies, 51 measure the pain outcome, 26 measure the disability outcome, and 23 include both outcomes. Figure 1 exhibits the trial network among therapies for each outcome. The size of each node represents the number of studies investigating the therapy, and the thickness of each edge denotes the total number of samples for the relation. The numbers on the edges indicate the numbers of studies investigating the relation. For example, in the pain outcome, there are five studies investigating the relation between no treatment and proprioception exercise, but this line is thinner than the line between education and strength exercise, though it has only three studies. The network features are similar in both outcomes, but we have limited information on the disability outcome, with fewer connections between therapies and smaller total sample sizes overall than for the pain outcome.

Likelihood

In MTCs, we must carefully distinguish between the terms *treatment* and *arm*. The former refers to a drug or device being tested, while the latter is the data on patients randomized to a particular drug or device in a *single* study. We must also distinguish between *reference* and *baseline* treatments. The reference treatment is a standard control treatment (often placebo, or simply no treatment) which can be compared with other active treatments. In our OA data, we select "no treatment" as the reference treatment among three possibilities (no treatment, education, and placebo). The baseline treatment is defined as the treatment assigned to the control arm *in each study*. That is, each study has its own baseline treatment, which is often the same as the reference treatment, but could differ. In this report, we assume there is no inconsistency, defined as discrepancy in treatment effects arising from direct and indirect comparisons.[8]

Suppose we are comparing K treatments from I studies in terms of L outcomes. For the continuous outcome, we assume that the data for a specific outcome from each study follow a normal distribution. That is,

$$\bar{y}_{ikl} \sim N\left(\Delta_{ikl}, \frac{\sigma_{ikl}^2}{n_{ikl}}\right), i = 1, \dots, I, k = 1, \dots, K, l = 1, \dots, L,$$

where $\bar{y}_{ikl}$ is the observed sample mean of the measurements, Δ_{ikl} is the unknown true population mean, σ_{ikl}^2 is the known sample variance, and n_{ikl} is the number of subjects in the k^{th} treatment arm from the i^{th} study with respect to the l^{th} continuous outcome. For the simplicity, we consider $k = 1$ as the reference treatment. Generally, in meta-analysis, we cannot estimate within-study correlations because we have only aggregated data.[19] We assume $\bar{y}_{ikl}$ are independent across arms and outcomes in study i since within-study correlations are not observed in every studies.

Existing Lu and Ades-Style Model

Fixed Effects Model

For meta-analysis, a fixed effects model, assuming no variability between studies, can easily be implemented. Following Lu and Ades,[7, 8] the model can be written as

$$\begin{aligned} \Delta_{ikl} &= \alpha_{iBl} && if\ k = B, \\ \Delta_{ikl} &= \alpha_{iBl} + \eta_{Bkl} && if\ k \neq B, \end{aligned} \tag{1}$$

where B indicates the baseline treatment in each study i. Here, α_{iBl} is the effect of baseline treatment and η_{Bkl} is the mean difference between treatment k and the baseline treatment (B) for outcome l in study i. However, we have to be careful to interpret α_{iBl} when the baseline treatment is not always the same. We define d_{kl} as the mean difference between treatment k and the reference treatment for outcome l, with $d_{1l} = 0$. Thus, η_{Bkl} can be calculated as $d_{kl} - d_{Bl}$, and we infer the treatment effects in terms of d_{kl}; that is, we assign a prior distribution to d_{kl}, rather than η_{Bkl}. We denote this model as the Lu and Ades (LA)-style fixed effects model (LAFE). In this approach, it is hard to interpret the baseline treatment effect α_{iBl} because not all studies have the same baseline treatment.

Random Effects Model

Next, in order to allow variability between studies, we introduce random effects, δ_{iBkl}, replacing the η_{Bkl}. Specifically model (1) is respecified as

$$\begin{aligned} \Delta_{ikl} &= \alpha_{iBl} && if\ k = B, \\ \Delta_{ikl} &= \alpha_{iBl} + \delta_{iBkl} && if\ k \neq B, \end{aligned} \tag{2}$$

where we can assume homogeneous variance across random effects for all arms, i.e.,

$$\delta_{iBkl} \sim \mathrm{N}(d_{kl} - d_{Bl}, \tau_l^2). \tag{3}$$

Here, δ_{iBkl} is 0 when $k = B$, and τ_l is the standard deviation of the random effects for each outcome l. We denote this model as the Lu and Ades-style homogeneous random effects model (LAREhom). For multi-arm trials, Lu and Ades provides a between-arm-contrast correlation of 0.5, as a consequence of homogeneous variance and their consistency equation.[8] The δ_{iBkl} in (3) are replaced by a vector $\boldsymbol{\delta}_{il}$ that follows a multivariate normal distribution with dimension equal to the number of arms in study i minus one, for each outcome l.

Allowing for Missing Data and Correlations Between Outcomes

Contrast-Based Approach

We denote a model that parameterizes *relative* effects (e.g., the η_{Bkl} and δ_{iBkl} in (1) and (2), respectively) as a *contrast-based* (CB) model. Lu and Ades-style models use such a CB approach. Note that the mean effect difference between treatment k and reference treatment in terms of outcome l (d_{kl}) is the parameter of interest in CB models. In MTCs it is common that the number of treatments compared in the i^{th} study is less than the complete collection of K treatments. Since each study contributes to the likelihood for a different set of treatments, using the observed measurements only can complicate estimating the covariance matrix for the $\boldsymbol{\delta}_{il}$ and lead to difficulties in prior assignment and parameter inference. In addition, it is plausible that

researchers select study arms based on the trials conducted previously, what statisticians call"nonignorable missingness." In this case, ignoring the missing treatment arms can potentially lead to biased parameter estimates.[15]

To remedy this, we assume that all studies can in principle contain every treatment as their arms, but in practice much of this information is missing for various reasons. Under this assumption, all studies can always have a common (though possibly missing) baseline treatment, $B = 1$, and the distribution for the random effects δ_{iBkl} in (3) can be replaced with a matrix form as follows:

$$\boldsymbol{\delta}_{il} \sim MVN(\boldsymbol{d}_l,\ \boldsymbol{\Sigma}_l^{Trt}), \tag{4}$$

where $\boldsymbol{\delta}_{il} = (\delta_{i12l}, \dots, \delta_{i1Kl})^T$, $\boldsymbol{d}_l = (d_{2l}, \dots, d_{Kl})^T$, and $\boldsymbol{\Sigma}_l^{Trt}$ is a $(K-1) \times (K-1)$ unstructured covariance matrix for $l = 1, \dots, L$. Note that since δ_{i11l} and d_{1l} are always 0, they are not included in $\boldsymbol{\delta}_{il}$ and $\boldsymbol{d}_l$. Here, $\boldsymbol{\Sigma}_l^{Trt}$ captures all random contrasts' relations among treatments in each outcome l. We refer to this model as a contrast-based random effects model assuming independence between outcomes (CBRE1).

To allow correlations among outcomes, the distribution of $\boldsymbol{\delta}_{il}$ in (4) needs to be respecified to

$$\boldsymbol{\delta}_{ik} \sim MVN(\boldsymbol{d}_k,\ \boldsymbol{\Sigma}_k^{Out}), \tag{5}$$

where $\boldsymbol{\delta}_{ik} = (\delta_{i1k1}, \dots, \delta_{i1kL})^T$, $\boldsymbol{d}_k = (d_{k1}, \dots, d_{kL})^T$, and $\boldsymbol{\Sigma}_k^{Out}$ is a $L \times L$ unstructured covariance matrix for $k = 2, \dots, K$. In this model, we assume independent random contrasts between treatments but incorporate the correlation structure of those contrasts between outcomes through $\boldsymbol{\Sigma}_k^{Out}$. We call this model CBRE2. Alternatively, we can also use the same $\boldsymbol{\Sigma}^{Out}$ for all k, if such an assumption is sensible.

In this approach, we can always have the same length of vector $\boldsymbol{\delta}_{il}$ or $\boldsymbol{\delta}_{ik}$ in each study *i*, and incorporate all sources of uncertainty by considering unobserved arms as missing data to be imputed by our MCMC algorithm using Gibbs-Metropolis sampling. For example, suppose Study 1 compares treatments 1, 2, and 3, giving information about two contrasts, δ_{i12l} and δ_{i13l}, whereas Study 2 compares only treatments 1 and 2, and Study 3 includes only treatments 1 and 3. We can impute the missing contrast δ_{i13l} and δ_{i12l} in Studies 2 and 3 respectively by using the information related to these contrasts observed in Study 1. The reference treatment effect, α_{iBl} in (2), is uninterpretable in this case, since each study will have different baseline treatment, as in the LA models. However, in our CB approach, α_{iBl} becomes meaningful because the baseline treatment is the same ($B = 1$) across all studies.

Although we only introduced the LA homogeneous random effects model, a heterogeneous random effects model can be applied with rigorous construction of covariance matrices to satisfy the positive definiteness condition under the consistency assumption.[20] However, our approach does not lead to this same set of consistency equation; the imputation allows us to independently estimate all possible contrasts in every study.

Arm-Based Approach

The CB method estimates the treatment contrasts; say, the mean difference between treatment k and the reference treatment. However, the approach's singular focus on relative treatment effects ultimately leads to many limitations. First, although we may resolve the incomparable baseline treatment problem by imputing such missing arms in our CB models, LA models still need complex model parameterizations for those studies with incomparable baseline treatments. Second, the interpretation of correlations between treatments or outcomes with

respect to relative effects can be difficult. For example, we cannot directly calculate the correlation between treatments via correlation between differences of treatment effects. Furthermore, our CB model restricts the variance of a baseline effect to always be smaller than that of other treatments. That is, the variance of population mean of baseline treatment, Δ_{iBl}, is $\mathrm{Var}(\alpha_{iBl})$, whereas for other treatments we have $\mathrm{Var}(\alpha_{iBl}) + \mathrm{Var}(\delta_{iBkl})$, which is never smaller than $\mathrm{Var}(\alpha_{iBl})$.

As an alternative, we introduce an *arm-based* (AB) approach[10, 21] by respecifying mean structure (2) as

$$\Delta_{ikl} = \mu_{kl} + \nu_{ikl}, \quad (6)$$

where μ_{kl} is the fixed mean effect of treatment k with respect to outcome l and ν_{ikl} is the study-specific random effect. In this approach, we estimate the *absolute* treatment effect size, μ_{kl}, not the relative effect size, d_{kl}.

If we begin by assuming independent random effects between outcomes, then the random effects ν_{ikl} in (6) can be structured as $(\nu_{i1l}, \dots, \nu_{iKl})^T \sim \mathrm{MVN}(\mathbf{0}, \mathbf{\Lambda}_l^{Trt})$ with $\mathbf{\Lambda}_l^{Trt}$ a $K \times K$ unstructured covariance matrix having relations of random effects between treatments, for $l = 1, \dots, L$. We denote this model as ABRE1. Alternatively, we can allow dependence of random effects between outcomes but independence between treatments by defining $(\nu_{ik1}, \dots, \nu_{ikL})^T \sim \mathrm{MVN}(\mathbf{0}, \mathbf{\Lambda}_k^{Out})$ where $\mathbf{\Lambda}_k^{Out}$ is a $L \times L$ unstructured covariance matrix having relations between outcomes, for $k = 1, \dots, K$. We refer to this model as ABRE2. Again, we can also use the same $\mathbf{\Lambda}^{Out}$ for all k when it is reasonable to do so.

The parameters in arm-based models permit more straightforward interpretation, especially in estimating a pure treatment effect. However, these models do require strong assumptions regarding the similarity and exchangeability of all populations, in order to preserve the randomization and permit meaningful clinical inference. Note that in AB models, there is no restriction on variances of random effects because all of our covariance matrices are unstructured. That is, AB models are less constrained, but thus have slightly larger number of parameters than CB models.

Choice of Priors

Lu and Ades assume a noninformative prior on each parameter, in order to let the data dominate the posterior calculation. For α_{iBl} and d_{kl}, a normal distribution with mean 0 and variance 100^2 is used, and a $\mathrm{Uniform}(0.01, 10)$ is assigned for τ in LAREhom. In all CB models, we assume α_{iBl} follows a $\mathrm{N}(a_l, \xi_l^2)$ rather than a $\mathrm{N}(0, 100^2)$ distribution, where a_l is the mean reference treatment effect, with noninformative priors for a_l and ξ_l; namely, $\mathrm{N}(0, 100^2)$ and $\mathrm{Uniform}(0.01, 10)$, respectively. Throughout all CB and AB models, the fixed effects (d_{kl} and μ_{kl}, respectively) follow a $\mathrm{N}(0, 100^2)$ distribution, while the inverse covariance matrices follow a $\mathrm{Wishart}(\mathbf{\Omega}, \gamma)$ having mean $\gamma\mathbf{\Omega}^{-1}$, with the matrix dimension usually chosen for the degrees of freedom parameter γ because it is the smallest value that will still yield a proper prior.[22] We can select $\mathbf{\Omega}$ to be γ times a prior guess for the covariance matrix ($\mathbf{\Omega}_0$). Since we do not know the true covariance matrices, we begin with a vague Wishart prior having $\mathbf{\Omega}_0 = \begin{pmatrix} 5 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 5 \end{pmatrix}$, and later investigate more informative Wishart priors in a sensitivity analysis.

Decisionmaking

Regarding Bayesian model choice, we adopt the Deviance Information Criterion (DIC).[22,23] DIC is a hierarchical models generalization of the Akaike Information Criterion, and is the sum of $\overline{D}$, a measure of goodness of fit, and P_D, a measure of complexity. For all CB and AB models we implement, we insist that only the observed data contribute to the calculation of $\overline{D}$.[24]

We can identify the best treatments based on a reasonable measurement of the effect size.[25] For instance, we can calculate the probability of being the best or second best treatment, which we call the "Best12" probability. Suppose Δ_{kl} is the marginal mean effect of having event l under treatment k, modeled from (2) using the posterior of d_{kl} and posterior mean of μ_{i1l} across studies, instead of δ_{iBkl} and μ_{iBl} in CB models. For AB models, we can obtain Δ_{kl} by plugging in the posterior of μ_{kl} in (6), noting that the prior mean of ν_{ikl} is 0. Denoting the data on outcome l by y_l, then define the "Best12" probability under each outcome as

$$\Pr\{\text{k is the best treatment} \mid y_l\} = \Pr\{\text{rank}(\Delta_{kl}) = 1 \text{ or } 2 \mid y_l\} \quad (7)$$

To integrate these univariate probabilities over all the outcomes and obtain one omnibus measure of "best," we propose an overall, weighted score denoted by S_k. Suppose all measurements have the same directionality, that is, small values indicate better condition in all outcomes, our overall score is defined as

$$S_k = \sum_l w_l \Delta_{kl}, \quad (8)$$

where w_l is the weight for outcome l, and $\sum_l w_l = 1$. This score can be used to obtain overall Best12 probabilities by replacing Δ_{kl} by S_k in (7). The weights can be chosen by physicians or public health professionals based on their preferences (say, for weighting safety versus efficacy).

Simulation Study Settings

In this simulation, we generate 1,000 data pairs $(\bar{y}_{ik1}, \bar{y}_{ik2})$ and fit the LAREhom, CBRE2, and ABRE2 models to investigate how the missingness in our design affects 5 percent two-sided Type I error, power, and the rates of incorrect decisions when the correlation between outcomes is incorporated into the models (CBRE2 and ABRE2) or not (LAREhom). Figure 2 illustrates the design of the simulated complete and partially missing data. For the "complete" data, we generate artificial data from 40 studies having two treatments and two outcomes featuring moderate positive correlation between outcomes, but independence between arms. In panel (b), we drop 20 studies in the first outcome; that is, we mimic our OA data, in which only half the studies report the disability outcome. For simplicity, we assume that every study has sample size 100 and standard deviation of 2 for every arm.

To sample the partially missing data, we compare the results under missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) mechanisms. The MCAR mechanism assumes that the missingness does not depend on the data, so we choose 20 studies randomly and make $\bar{y}_{i11}$ and $\bar{y}_{i21}$ missing for those studies. The MAR mechanism assumes that the missingness depends only on the observed data, but not on the missing data, whereas MNAR missingness can depend on both observed and unobserved data. To generate partially missing data under the MAR and MNAR mechanisms, we first calculate the 'probability of missing' ($p_{i,mis}$) for study i by applying a logit model with the observed or missing data as covariates. Here $\bar{y}_{i12}$ and $\bar{y}_{i22}$ are considered as observed data, and $\bar{y}_{i11}$ and $\bar{y}_{i21}$ are missing data since they are not fully observed in our design. We use the following two logit models:

(9)

$$\text{MAR: logit}(p_{i,mis}) = 2 + \bar{y}_{i12} - \bar{y}_{i22}$$

$$\text{MNAR: logit}(p_{i,mis}) = -4 - \bar{y}_{i11} + \bar{y}_{i22}. \quad (10)$$

The coefficients are selected to result in a mean $p_{i,mis}$ of about 30 to 40 percent. Given $p_{i,mis}$, we generate the missingness indicator vector until 20 studies are selected as missing data.

For the true parameters, $(\mu_{11}^*, \mu_{21}^*, \mu_{12}^*, \mu_{22}^*) = (0, 0, 0, 3)$ is chosen in (6), yielding $d_{21}^* = 0$ and $d_{22}^* = 3$ in the LAREhom and CBRE models. We calculate Type I error in terms of parameter d_{21} in the three models, with the superscript * indicating the truth. To estimate power at two particular alternatives, we select $(\mu_{11}^*, \mu_{21}^*, \mu_{12}^*, \mu_{22}^*) = (0, 1, 0, 3)$ and $(0, 2, 0, 3)$, giving $d_{21}^* = 1$ and 2, respectively, which we notate as "Power1" and "Power2." We also calculate the rate of incorrectly selecting the best treatment, given as $\Pr(\widehat{\mu_{11}} > \widehat{\mu_{21}})$ under Power1 and 2 scenarios because the truth is that $\mu_{11}^* < \mu_{21}^*$. This rate should be around 0.5 under the Type I error setting.

For the random effect parameters, in (6), we generate them from $\begin{pmatrix} \nu_{i11}^{AB} \\ \nu_{i21}^{AB} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{AB}^* \\ \rho_{AB}^* & 1 \end{pmatrix}\right)$ and $\begin{pmatrix} \nu_{i12}^{AB} \\ \nu_{i22}^{AB} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 & 3\rho_{AB}^* \\ 3\rho_{AB}^* & 3 \end{pmatrix}\right)$, which on the CB scale corresponds to $\begin{pmatrix} \nu_{i21}^{CB} \\ \nu_{i22}^{CB} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 3\rho_{AB}^* \\ 3\rho_{AB}^* & 2 \end{pmatrix}\right)$. Here, the superscripts and subscripts on ν_{ikl} and ρ^*, AB and CB, indicate the model used. From the covariance matrix of random effects in the CB model, we can easily calculate the true correlation in the CB model, $\rho_{CB}^* = \frac{3}{2}\rho_{AB}^*$. To ensure a positive definite covariance matrix for the random effects in the CB model, ρ_{AB}^* should therefore be between $-\frac{2}{3}$ and $\frac{2}{3}$. We set $\rho_{AB}^* = 0.6$ and 0.0 which induces $\rho_{CB}^* = 0.9$ and 0.0.

For the OA data analysis, WinBUGS is used to generate two parallel chains of 50,000 MCMC samples after a 50,000-sample burn-in. To check MCMC convergence, we used standard diagnostics, including trace plots and lag 1 sample autocorrelations. The WinBUGS codes are now publicly available at www.biostat.umn.edu/~brad/software.html.

We used the R2WinBUGS package[26] in R to perform our simulation studies, where we call WinBUGS[27] 1,000 times from R, once for each simulated data set. In each case, we obtain 20,000 samples, after a 20,000 sample burn-in, and collect medians of parameters across 1,000 simulated datasets, then estimate Type I error and power.

Table 1. Raw OA data

Study	Country	Duration	Pain Score	Disability Score	N	Therapy	Pain Rescaled Score	Pain Rescaled SD	Disability Rescaled Score	Disability Rescaled SD
Aglamis, 2008[28]	Turkey	2	VAS	SF36	17	No trt.	7.700	2.300	6.360	0.790
			0-10	0-100	17	Ex. aerobic	0.700	1.000	1.280	0.970
Kovar, 1992[29]	US	2	AIMS	AIMS	50	No trt.	4.770	2.120	5.960	2.320
			0-10	0-10	52	Ex. aerobic	3.770	1.730	3.740	2.690
Ettinger, 1997[30]	US	4	Other[a]	Other[b]	149	Education	2.800	1.220	2.250	1.225
			1-6	1-5	144	Ex. aerobic	2.280	1.200	1.800	1.200
					146	Ex. strength	2.420	1.440	1.850	1.225
Sullivan, 1998[31]	US	4	AIMS	AIMS	50	Education	5.500	2.070	6.180	2.750
			0-10	0-10	52	Ex. aerobic	4.590	2.400	6.070	2.950
Patrick, 2001[32]	US	2	HAQ	HAQ	124	No trt.	4.873	2.063	3.757	2.237
			0-3	0-3	125	Ex. aquatic	4.607	2.457	3.110	1.833
Baker, 2001[33]	US	3	WOMAC	SF36	23	Education	3.780	2.345	3.920	2.936
			0-500	0-100	23	Ex. strength	2.560	1.962	3.660	2.869
Kuptniratsaikul, 2002[34]	Thailand	3	AIMS	Other[c]	193	No trt.	5.070	2.530	3.045	1.720
			0-10	0-20	199	Ex. strength	4.060	2.530	2.695	1.805
Callaghan, 2005[35]	UK	1	VAS	AIMS	10	Placebo	6.300	1.900	5.100	1.700
			0-10	0-10	10	Diathermy(L)	5.000	3.200	5.500	3.000
					10	Diathermy(H)	5.500	2.700	5.100	2.300
Laufer, 2005[36]	Israel	1	WOMAC	WOMAC	33	Placebo	4.440	3.510	4.630	3.540
			0-10	0-10	38	Diathermy(L)	4.730	3.480	4.930	3.630
					32	Diathermy(H)	4.030	3.300	4.400	3.440
Yip, 2007[37]	China	3	VAS	HAQ	94	No trt.	4.250	2.367	1.850	1.374
			0-10	0-24	88	Ex. aerobic	3.858	2.201	1.958	1.539
Rooks, 2006[38]	US	2	WOMAC	SF36	23	Education	3.750	2.500	5.980	1.940
			0-30	0-100	22	Ex. aquatic	3.650	0.350	6.600	2.150
Brismee, 2007[39]	US	2	Other[d]	WOMAC	19	No trt.	3.370	1.780	2.990	1.630
			0-10	26-130	22	Ex. prop.	2.410	2.050	2.806	2.327
Garland, 2007[40]	US	2	WOMAC	WOMAC	19	Placebo	4.180	1.659	4.590	1.681
			0-100	0-100	39	Elec. stim.	3.740	2.360	3.960	2.425
Doi, 2008[41]	Japan	2	VAS	SF36	70	No trt.	2.959	2.394	3.660	1.636
			0-10	0-100	72	Ex. strength	2.255	2.068	2.881	1.633
Lund, 2008[42]	Denmark	2	VAS	KOOS	27	No trt.	2.380	1.403	3.860	1.351
			0-10	0-100	27	Ex. aquatic	1.810	1.403	3.700	1.351
					25	Ex. strength	1.560	1.400	3.610	1.350

Table 1. Raw OA data (continued)

Study	Country	Duration	Pain Score	Disability Score	N	Therapy	Pain Rescaled Score	Pain Rescaled SD	Disability Rescaled Score	Disability Rescaled SD
Yip, 2008[43]	Hong Kong	3	VAS 0-10	HAQ 0-100	50	No trt.	3.459	2.355	0.357	0.280
					45	Ex. aerobic	3.523	2.193	0.428	0.368
Özgönenel, 2009[44]	Turkey	1	VAS 0-10	WOMAC 0-96	33	Placebo	4.000	2.600	4.010	1.583
					34	Ultra sound	3.900	2.000	3.469	1.615
Selfe, 2008[45]	US	2	WOMAC 0-50	WOMAC 0-240	20	Placebo	3.178	1.784	3.420	1.654
					20	Elec. stim.	2.834	2.136	2.933	2.034
Péloquin, 1999[46]	Canada	2	AIMS 0-10	AIMS 0-10	68	Education	3.940	2.220	1.930	1.880
					69	Ex. aerobic	3.090	1.540	1.850	2.260
Chaipinyo, 2009[47]	US	2	KOO 0-100	KOOS 0-100	24	Ex. strength	1.800	1.600	1.800	1.300
					24	Ex. prop.	1.300	1.200	1.200	1.000
Lee, 2009[48]	South Korea	2	WOMAC 0-35	SF36 0-100	15	No trt.	1.686	1.057	4.490	1.750
					29	Ex. prop.	1.314	1.143	3.560	2.090
Tascioglu, 2010[49]	Turkey	1	VAS 0-10	WOMAC 0-96	30	Placebo	6.670	1.780	4.618	1.331
					30	Ultra sound	5.250	1.900	4.525	1.717
Fukuda, 2011[50]	Brazil	1	Other[e] 0-10	KOOS 0-100	23	Placebo	6.900	2.000	4.850	1.750
					32	Diathermy(L)	3.800	2.200	3.850	2.030
					31	Diathermy(H)	4.600	2.500	3.680	1.650
Messier, 1997[51]	US	3	Other[f] 1-6		36	Education	2.560	1.800		
					33	Ex. aerobic	2.300	1.954		
Grimmer, 1992[52]	Australia	1	VAS 0-10		20	Placebo	3.500	2.900		
					20	Elec. stim.	2.200	2.800		
Taylor, 1981[53]	US	1	Other[g] -1-3		10	Placebo	6.750	2.375		
					10	Elec. stim.	5.250	1.425		
Borjesson, 1996[54]	Sweden	2	Other[h] 0-10		34	No trt.	3.300	1.500		
					34	Ex. strength	3.000	1.500		
Bautch, 1997[55]	US	2	VAS 0-10		17	Education	2.080	2.090		
					17	Ex. aerobic	2.190	1.670		
Wyatt, 2001[56]	US	2	VAS 0-10		23	Ex. aerobic	3.800	1.600		
					23	Ex. aquatic	2.400	1.600		
Gür, 2002[57]	Turkey	2	Other[i] 0-70		6	No trt.	4.000	0.743		
					9	Ex. strength	1.471	0.643		
Topp, 2002[58]	US	3	WOMAC 0-20		35	No trt.	5.385	1.528		
					32	Ex. strength	5.190	1.657		

Table 1. Raw OA data (continued)

Study	Country	Duration	Pain Score	Disability Score	N	Therapy	Pain Rescaled Score	Pain Rescaled SD	Disability Rescaled Score	Disability Rescaled SD
Talbot, 2003[59]	US	3	Other[l]		21	Education	1.397	1.242		
			0-78		19	Ex. aerobic	1.660	1.463		
Talbot, 2003[60]	US	2	Other[l]		18	Education	1.426	1.026		
			0-78		20	Elec. stim.	2.094	1.712		
Messier, 2004[61]	US	3	WOMAC		78	No trt.	3.095	2.030		
			0-20		82	Education	2.550	1.945		
					80	Ex. aerobic	3.110	2.010		
Keefe, 2004[62]	US	2	AIMS		18	No trt.	4.030	2.080		
			0-10		18	Education	4.000	1.560		
					16	Ex. aerobic	3.190	1.850		
Gaines, 2004[63]	US	2	AIMS		18	Education	5.990	2.400		
			0-10		20	Elec. stim.	5.180	2.110		
Law, 2004[64]	Hong Kong	1	VAS		10	Placebo	4.100	2.600		
			0-10		12	Elec. stim.	0.700	0.700		
Durmus, 2007[65]	Turkey	1	Other[l]		25	Elec. stim.	0.600	0.100		
			0-10		25	Ex. strength	1.040	0.270		
Hay, 2006[66]	UK	3	WOMAC		108	No trt.	4.180	1.950		
			0-20		109	Ex. aerobic	3.755	2.400		
Silva, 2008[67]	Brazil	2	VAS		32	Ex. aerobic	3.840	2.750		
			0-10		32	Ex. aquatic	3.700	1.810		
Jan, 2008[68]	Taiwan	2	WOMAC		34	No trt.	3.550	1.700		
			0-20		34	Ex. strength	2.400	1.750		
Itoh, 2008[69]	Japan	2	VAS		8	No trt.	4.930	2.020		
			0-10		8	Elec. stim.	5.350	0.970		
An, 2008[70]	China	2	WOMAC		14	No trt.	2.764	2.252		
			0-500		14	Ex. aerobic	1.422	2.202		
Tsauo, 2008[71]	Taiwan	2	WOMAC		30	No trt.	1.320	0.760		
			0-500		30	Ex. prop.	1.280	0.740		
Lim, 2008[72]	Australia	2	WOMAC		28	No trt.	3.360	1.540		
			0-100		27	Ex. strength	2.280	1.690		
Pietrosimone, 2009[73]	US	1	VAS		12	No trt.	2.096	1.844		
			0-10		11	Elec. stim.	1.165	1.671		
Lin, 2009[74]	Taiwan	2	WOMAC		36	No trt.	3.650	1.700		
			0-20		36	Ex. strength	2.100	1.500		
					36	Ex. prop.	2.150	1.150		

Table 1. Raw OA data (continued)

Study	Country	Duration	Pain Score	Disability Score	N	Therapy	Pain Rescaled Score	Pain Rescaled SD	Disability Rescaled Score	Disability Rescaled SD
Weng, 2009[75]	Taiwan	2	VAS		66	No trt.	4.400	1.400		
			0-10		66	Ex. strength	3.600	0.700		
					66	Ex. prop.	2.700	1.900		
Farr, 2010[76]	US	2	WOMAC		98	Education	7.200	6.630		
			0-100		100	Ex. aerobic	6.710	6.880		
Bennell, 2010[77]	Australia	2	WOMAC		44	No trt.	3.250	1.650		
			0-20		45	Ex. strength	2.450	1.650		
Swank, 2011[78]	US	2	VAS		36	Education	4.556	0.467		
			0-10		37	Ex. strength	3.667	0.422		
Schilke, 1996[79]	US	2		AIMS	10	No trt.			2.500	0.850
				0-10	10	Ex. strength			2.300	0.840
Deyle, 2000[80]	US	2		WOMAC	41	Placebo			3.893	2.723
				0-2400	42	Ex. aerobic			1.927	1.826
Rejeski, 2002[81]	US	4		SF36	78	Education			6.559	0.899
				0-100	80	Ex. aerobic			6.286	1.038

Abbreviations: WOMAC = Western Ontario MacMaster; VAS = Visual Analogue Scale; AIMS = Arthritis Impact Measurement Scale; SF36 = Medical Outcome Study 36-Item Short-Form health Survey; HAQ = Health Assessment Questionnaire; KOOS = Knee Injury and Osteoarthritis Outcome Score

Note: treatment duration (followup) is the category of weeks spent in therapy (1: 0-5, 2: 6-12, 3: 13-26, and 4: >27); N is the sample size; Pain Score and Disability Scores are the scores used for pain and disability with the original ranges

[a]pain intensity score

[b]self-report of physical disability

[c]functional incapacity score

[d]overall knee pain

[e]numeric pain rating scale

[f]ambulation intensity

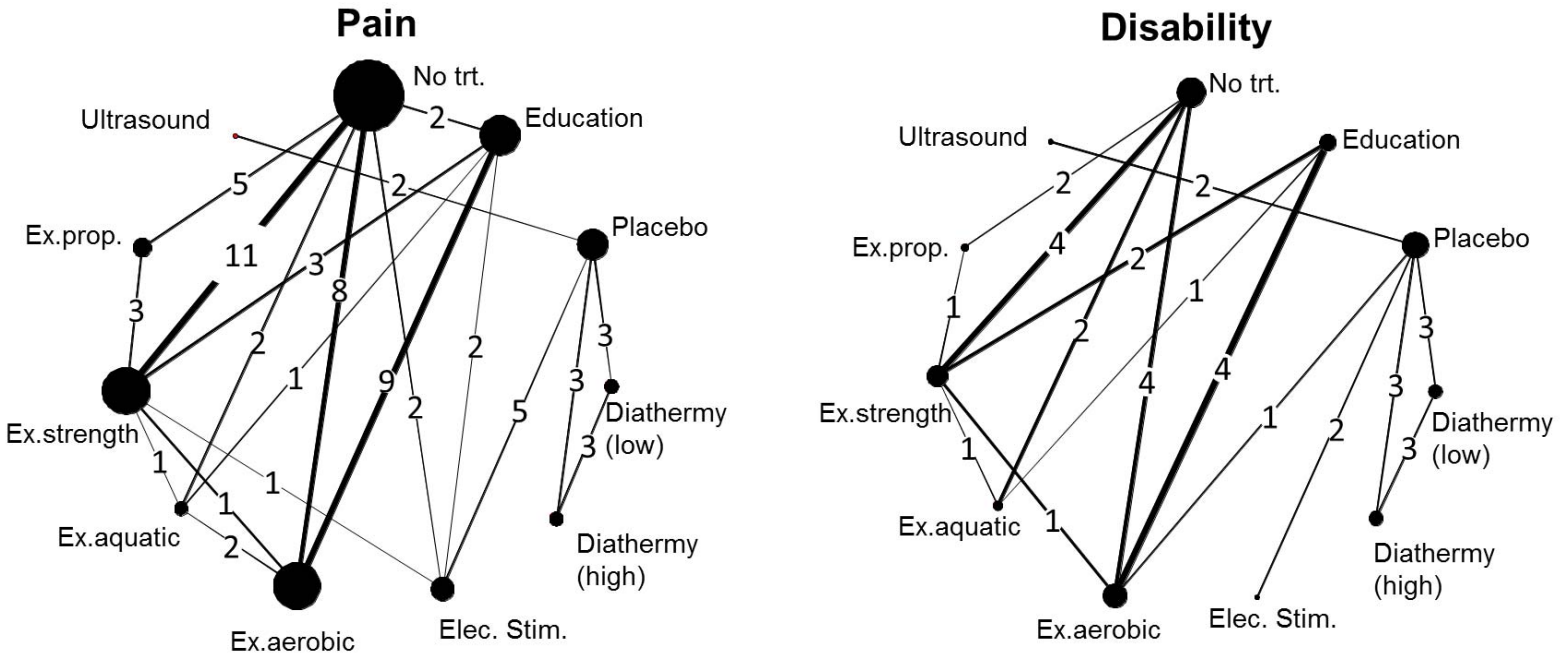
[g]subject pain

[h]Borg scale

[I]not clear

[j]pain rating index

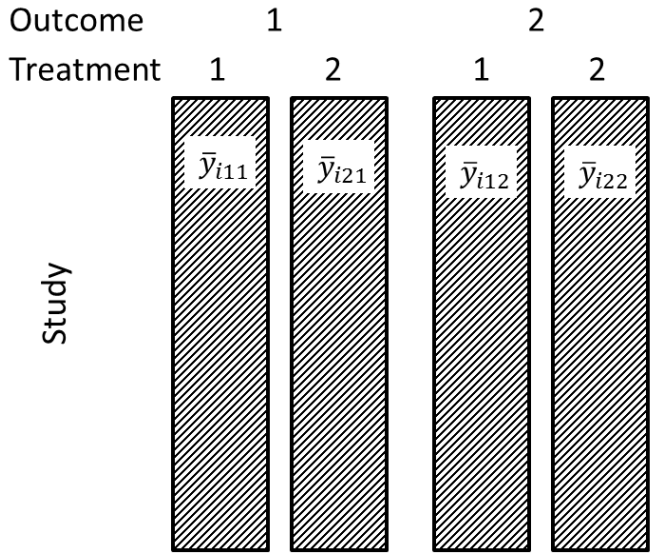
Figure 1. Network graphs of OA data for each outcome; (a) pain and (b) disability



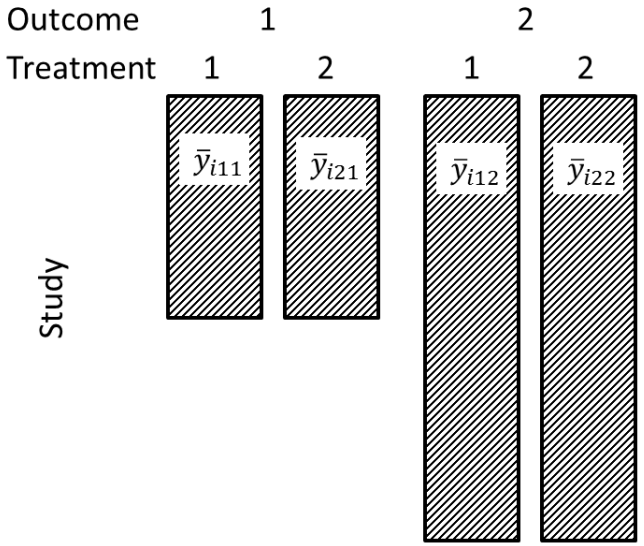
(a) Pain (b) Disability

Note: The size of each node represents the number of studies investigating the therapy, and the thickness of each edge implies the total number of samples for the relation. The number on the line is the number of studies for the relation.

Figure 2. Data structure for simulation; (a) complete data and (b) partially missing data



(a) Complete data



(b) Partially missing data

Results

Results for OA Data

Table 2 compares the fit of six models with our OA data. We apply homogeneous variance across arms in LARE and homogeneous covariance matrices for CBRE2 and ABRE2; that is, $\boldsymbol{\Sigma}_k^{Out}$ and $\boldsymbol{\Lambda}_k^{Out}$ are the same for all k, respectively. All CB and AB models incorporate the missingness into models, and only CBRE2hom and ABRE2hom models allow correlation structure between outcomes. The fixed effects model gives the largest mean deviance score $\overline{D}$ when applied to the OA data, and an unacceptably large DIC score. ABRE1 fits the data best with the smallest $\overline{D}$, but there is no significant difference in fit across random effects models. AB models give slightly higher pD than CB models because they are less constrained and more parameters need to be estimated. Since our data are sparse, heterogeneous variance assumption, a feature of CBRE1 and ABRE1, is not a good choice here. Considering both goodness of fit and complexity, CBRE2hom gives the smallest DIC, though again, the DIC differences between this model and ABRE2hom or LAREhom are not of practical importance (less than five units). The estimated variability on the standard deviation scale is always between 1 and 1.5, with associated 95% credible interval widths around 0.4 based on the median posteriors in LAREhom, CBRE2hom and ABRE2hom models. The median posterior of correlations between two outcomes are 0.494 (95% credible interval 0.18 to 0.71) and 0.377 (0.06 to 0.61) for the CBRE2hom and ABRE2hom models, respectively, revealing the two outcomes to be positively but weakly correlated (data is not shown).

Table 3 displays the results from four models; LAFE, LAREhom, CBRE2hom, and ABRE2hom with respect to the pain outcome. Here, smaller values of d_{k1} and μ_{k1}mean better condition and the "best" treatment based on the Best12 probability is in bold. In the LAREhom model, it is essentially tied with aquatic and proprioception exercises for first place. Our CB and ABRE2hom models suggest that proprioception exercise is the best treatment, followed by strength exercise, but the Best12 probability of proprioception exercise from ABRE2hom is much larger than that from CBRE2hom. However, since standard deviations are somewhat large, there is no significant difference between these two treatments. There are large differences in Best12 probabilities across three random effects models. This might be due to different model assumptions and settings but also to the network in the data structure.

Table 4 shows similar information with respect to the disability outcome. Aerobic exercises perform best based on Best12 probabilities from LAREhom models. Proprioception and aerobic exercises are tied for first place in the CBRE2hom model, and proprioception exercise is the best treatment followed by strength exercise in ABRE2hom. It seems that proprioception and aerobic exercises are helpful to reduce disability across all models, but there is still no strong evidence regarding significant difference among the treatments.

Figure 3 delivers our findings above graphically in terms of mean difference between therapy and no treatment (d_{kl}) with 95% credible intervals across the four models. We indicate the best treatment with respect to each outcome in each model with a triangle character, and the worst treatment with a square. For the pain outcome, strength and proprioception exercises perform significantly better than no active treatment across all models, whereas for the disability outcome, only aerobic exercise is significantly different from no active treatment under the three random effects models. Compared with the pain outcome, the 95% credible sets in disability are wider because only about half as many studies reported this outcome.

Figures 4 and 5 exhibit the posterior probabilities of each treatment taking each possible ranking from 1 (best) to 11 (worst) for both the pain reduction and disability improvement outcomes.[25] Although these graphs cannot reveal significant differences in rankings among treatments or the magnitudes of any treatment differences, they do still give a sense of the uncertainty in the rank for each treatment. Note that in both figures the positive correlation between the two outcomes leads to generally similar treatment ranking probabilities for both outcomes. In Figure 5, proprioception exercise's probability of being the best treatment for pain is roughly 0.8, leaving the remaining 10 treatments to share the remaining 0.2 probability of being the best; this treatment also has the single largest probability of being best for disability improvement (about 0.4). By contrast, the LA model rankings in Figure 4 do not suggest a dominant treatment for either outcome, though aerobic exercise has a nearly 0.4 chance of being best for disability improvement, and placebo is unequivocally worst for pain reduction.

To obtain Best12 probabilities with combined score in Equation (8), we investigate three sets of weights: (w_1, w_2) = (0.5, 0.5), (0.8, 0.2), and (0.2, 0.8). Our CB and ABRE2hom models give proprioception exercise as the global winner for all three sets of weights. Aerobic exercise is the overall winner in the LAREhom model (results not shown). The reason why the weights do not have much effect here is that some treatment effects are so large in one outcome that they dominate the effects from the other outcome, even when we put low weight on the former (e.g., Best12 probability of aerobic exercise in the disability outcome is much larger than that of low intensity diathermy the pain outcome for LAREhom).

Sensitivity Analysis

Our CB and ABRE2hom models yield weakly positive correlation between two outcomes under noninformative Wishart prior on covariance matrix of random effects, assuming zero correlation between outcomes with $\gamma = 2$ degrees of freedom. As a sensitivity analysis, we consider three different more informative Wishart priors: 0.5 between-outcome correlation with $\gamma = 2$ and 4, and 0.9 between-outcome correlation with $\gamma = 4$. Note that a Wishart prior becomes less informative as γ decreases to 0.

Table 5 displays the results of our sensitivity analysis in terms of model fits (pD, $\overline{D}$, and DIC) and posterior estimates of correlation between two outcomes ($\hat{\rho}$). Here, the degree of informativeness in the Wishart hyperprior increases from left to right. The $\hat{\rho}$s in CBRE2hom models are likely to be affected more by the selection of a Wishart prior having $\hat{\rho}$ close to 0.9 when $\rho_0 = 0.9$, $\gamma = 4$ while ABRE2hom gives a bit more robust $\hat{\rho}$ around 0.5 across the three sets of informative priors. In CBRE2hom, pD decreases as we utilize a more informative prior, whereas ABRE2hom gives almost the same pD values across all informative priors. Regarding treatment effect parameters, informative priors do not give dramatic difference in the treatment ranking (proprioception exercise is the best treatment in both outcomes under both CB and ABRE2hom models across all informative prior cases), but provide smaller standard deviation of those parameters.

Results for Simulation Study

Tables 6 and 7 present the results of our simulation under $\rho^*_{AB} = 0.6$ and 0.0, respectively. For CBRE2 and ABRE2 models, we used two different Wishart priors for the covariance matrices; namely, a noninformative $\text{Wishart}\left(\begin{pmatrix}10 & 0\\ 0 & 10\end{pmatrix}, 2\right)$ and a weakly informative

Wishart($4\boldsymbol{R}^*, 4$), respectively, where $\boldsymbol{R}^*$ is the true covariance matrix. We report Pr($\widehat{\mu_{11}} > \widehat{\mu_{21}}$) in parentheses which is interpreted as the probability of an incorrect decision when $d_{21}^* = 1$ or 2, but should be around 0.5 when $d_{21}^* = 0$, along with the simulated Type I error and power. Here, using true covariance matrix in the prior distribution could be a way overly optimistic, but we adopt the truth to investigate how much power could be gained with informative priors.

In Table 6, all models work fairly well when there is no missing data ("complete"). For Type I error, the LAREhom model performs poorly under MAR and MNAR mechanisms with very extreme Pr($\widehat{\mu_{11}} > \widehat{\mu_{21}}$) values, very close to 0 (MAR) or 1 (MNAR). Power1 decreases under the MCAR mechanism as we expected due to the loss of data, but our CBRE2 and ABRE2 models give slightly higher power than LAREhom. The LAREhom model gives extremely high Power1 under MAR, but too low under MNAR. Here, under MNAR the probability of an incorrect decision is 0.377 using LAREhom, while it is only 0.080 using CBRE2 and ABRE2. All models yield very high power when $d_{21}^* = 2$ except the LAREhom model under MNAR mechanism. The fifth and sixth columns show that adopting weakly informative Wishart priors can improve power without severely damaging Type I error.

Table 7 shows that our methods have less benefit when two outcomes are independent. In this case, the LAREhom model does not suffer as much on Type I error under MAR and MNAR mechanisms, and Power1 values are not extreme; it also gives slightly smaller Pr($\widehat{\mu_{11}} > \widehat{\mu_{21}}$) values when $d_{21}^* = 1$ under MNAR than our CBRE2 and ABRE2 models. This is because these methods do not borrow much strength across outcomes since the correlation is close to zero in this setting. Compared with Table 2, CBRE2 and ABRE2 produce somewhat smaller powers under severe missingness mechanisms than when the two outcomes were correlated.

Figure 6 exhibits the density plot of median posteriors of d_{21} from 1,000 simulated partially missing data under each of three models with noninformative Wishart priors, when $\rho_{AB}^* = 0.6$ and d_{21}^* is 0, 1, and 2 under MCAR, MAR, and MNAR mechanisms. When the missingness does not depend on the data (MCAR), the median posteriors of d_{21} are unbiased across all three models, though ABRE2 gives slightly smaller estimator variances, suggesting smaller mean squared error (MSE). On the other hand, the MAR and MNAR mechanisms lead to huge positive or negative biases with the LAREhom model, resulting in large Type I error and extreme Power1 values. This bias depends on the choices of coefficients in Equation (9); for example, if we alter (9) to $\text{logit}(p_{i,mis}) = -4 - 2\bar{y}_{i12} + \bar{y}_{i22}$ for MAR, LAREhom gives 0.087 Power1 while CBRE2 and ABRE2 give 0.37 and 0.311, respectively. No matter which rules drive the missingness, it is obvious that LAREhom models produce larger bias than our models when the missingness does not randomly occur and the two outcomes are correlated.

Figure 7 displays the same density plots as in Figure 6, but under $\rho_{AB}^* = 0.0$. All three models deliver unbiased estimates under MCAR and MAR, but give somewhat biased estimates under MNAR, although the magnitudes of bias are similar across models. Our CBRE2 and ABRE2 models tend to give slightly larger estimator variances. Here, the missingness does not much affect the bias of estimators in LAREhom with two uncorrelated outcomes. Although our methods do not deliver strikingly better features over the existing LAREhom model in this idealized case, our methods do not surrender much in terms of Type I error and power, justifying their uses across both dependent and independent scenarios.

Table 2. Model comparisons for the OA data

	LAFE	LAREhom	CBRE1	CBRE2hom	ABRE1	ABRE2hom
pD	96.9	154.5	162.9	153.9	164.4	158.0
$\overline{D}$	688.3	169.4	168.4	169.0	165.9	167.8
DIC	769.2	323.9	331.3	322.9	330.3	325.8

Note: lower DIC indicates the better model; lower pD indicates smaller effective model size; lower $\overline{D}$ indicates better model fit.

Table 3. Estimates of treatment effects and Best12 probabilities from four models with outcome pain

Treatment Effects	LAFE (d_{k1})	LAREhom (d_{k1})	CBRE2hom (d_{k1})	ABRE2hom(μ_{k1})
No treatment	0	0	0	3.732 (0.28)
Education	-0.178 (0.10)	-0.530 (0.36)	-0.223 (0.38)	3.674 (0.38)
Placebo	-0.125 (0.33)	0.627 (0.78)	1.217 (0.54)	5.000 (0.46)
Diathermy (low)	-2.014 (0.55)	-0.895 (1.07)	0.224 (0.90)	4.480 (0.86)
Diathermy (high)	-1.651 (0.56)	-0.665 (1.07)	0.439 (0.90)	4.683 (0.86)
Electrical stimulation	-1.201 (0.10)	-0.664 (0.55)	-0.547 (0.46)	2.904 (0.44)
Aerobic exercise	-0.676 (0.10)	-0.982 (0.32)	-0.856 (0.32)	3.192 (0.34)
Aquatic exercise	-0.654 (0.18)	-0.958 (0.50)	-0.617 (0.55)	3.252 (0.62)
Strength exercise	-0.799 (0.08)	-0.935 (0.27)	-1.001 (0.28)	2.632 (0.34)
Proprioception exercise	-0.778 (0.12)	-1.007 (0.42)	-1.057 (0.46)	1.814 (0.57)
Ultrasound	-1.002 (0.50)	-0.152 (1.11)	0.583 (1.00)	4.540 (0.99)
Best12				
No treatment	0.000 (0.00)	0.000 (0.00)	0.000 (0.00)	0.001 (0.03)
Education	0.000 (0.00)	0.020 (0.14)	0.014 (0.12)	0.003 (0.06)
Placebo	0.000 (0.00)	0.000 (0.01)	0.000 (0.00)	0.000 (0.00)
Diathermy (low)	**0.948** (0.22)	**0.371** (0.48)	0.075 (0.26)	0.015 (0.12)
Diathermy (high)	0.755 (0.43)	0.268 (0.44)	0.046 (0.21)	0.007 (0.08)
Electrical stimulation	0.191 (0.39)	0.091 (0.29)	0.146 (0.35)	0.253 (0.43)
Aerobic exercise	0.000 (0.02)	0.274 (0.45)	0.330 (0.47)	0.059 (0.23)
Aquatic exercise	0.002 (0.05)	0.311 (0.46)	0.230 (0.42)	0.135 (0.34)
Strength exercise	0.003 (0.05)	0.208 (0.41)	0.533 (0.50)	0.556 (0.50)
Proprioception exercise	0.004 (0.06)	0.339 (0.47)	**0.575** (0.49)	**0.948** (0.22)
Ultrasound	0.097 (0.30)	0.119 (0.32)	0.051 (0.22)	0.023 (0.15)

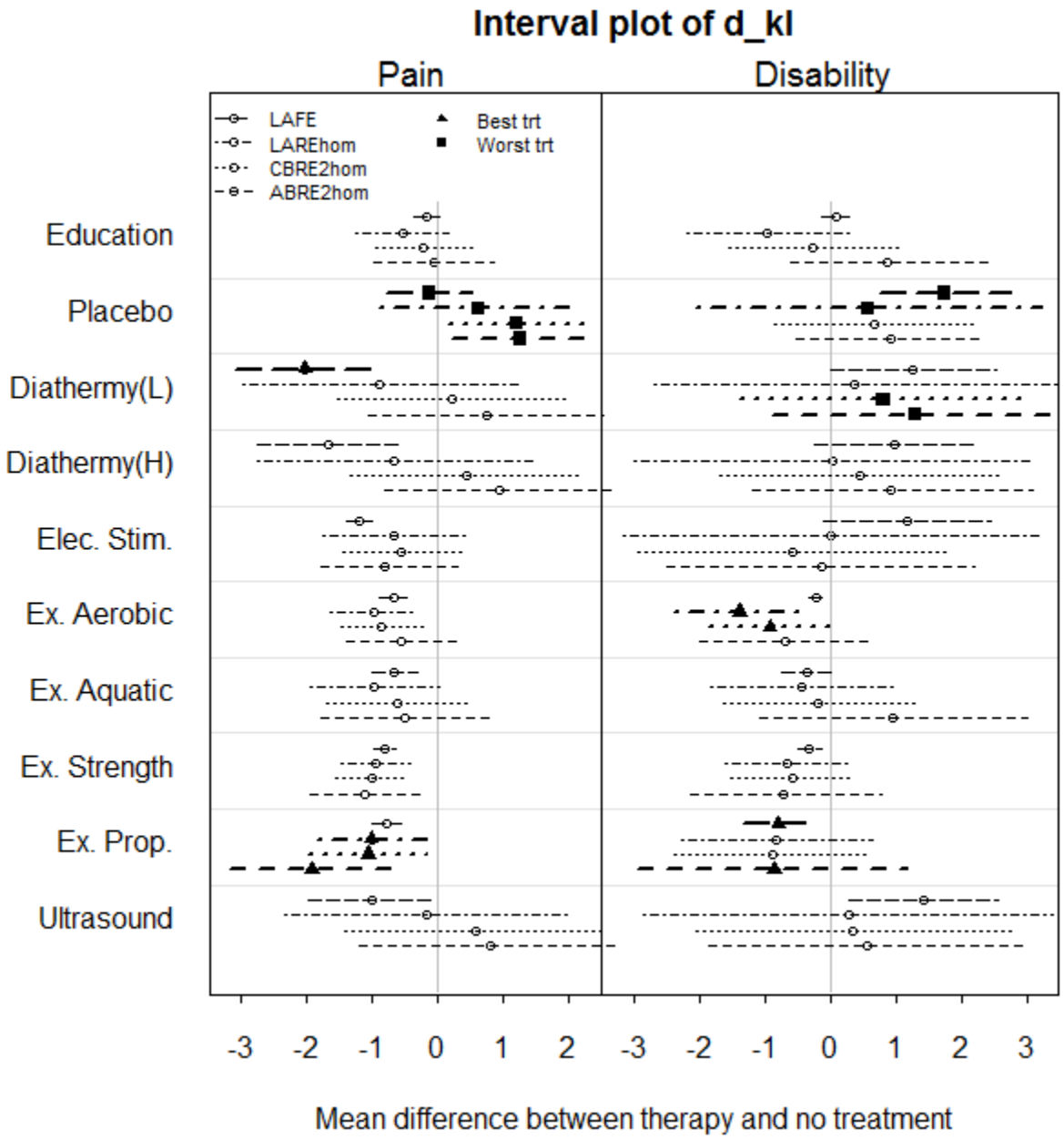
Note: Standard error is in parentheses, and the "best" treatment in terms of Best12. Probability is in bold

Table 4. Estimates of treatment effects and Best12 probabilities from four models with outcome disability

Treatment Effects	LAFE (d_{k2})	LAREhom (d_{k2})	CBRE2hom (d_{k2})	ABRE2hom (μ_{k2})
No treatment	0	0	0	3.425 (0.47)
Education	0.072 (0.11)	-0.959 (0.62)	-0.266 (0.67)	4.300 (0.64)
Placebo	1.741 (0.51)	0.582 (1.33)	0.658 (0.79)	4.359 (0.57)
Diathermy (low)	1.263 (0.65)	0.368 (1.55)	0.809 (1.12)	4.717 (1.00)
Diathermy (high)	0.985 (0.63)	0.035 (1.54)	0.460 (1.10)	4.340 (0.96)
Electrical stimulation	1.174 (0.65)	0.015 (1.61)	-0.567 (1.12)	3.301 (1.01)
Aerobic exercise	-0.214 (0.06)	-1.392 (0.50)	-0.926 (0.48)	2.766 (0.49)
Aquatic exercise	-0.354 (0.19)	-0.452 (0.70)	-0.200 (0.73)	4.380 (0.93)
Strength exercise	-0.321 (0.10)	-0.650 (0.48)	-0.590 (0.48)	2.727 (0.59)
Proprioception exercise	-0.789 (0.27)	-0.819 (0.74)	-0.895 (0.76)	2.558 (0.91)
Ultrasound	1.423 (0.58)	0.274 (1.58)	0.333 (1.21)	3.981 (1.13)
Best12				
No treatment	0.000 (0.00)	0.002 (0.04)	0.002 (0.05)	0.062 (0.24)
Education	0.000 (0.00)	0.317 (0.47)	0.107 (0.31)	0.008 (0.09)
Placebo	0.000 (0.00)	0.006 (0.08)	0.005 (0.07)	0.003 (0.05)
Diathermy (low)	0.003 (0.06)	0.074 (0.26)	0.036 (0.19)	0.018 (0.13)
Diathermy (high)	0.013 (0.12)	0.136 (0.34)	0.071 (0.26)	0.035 (0.18)
Electrical stimulation	0.007 (0.08)	0.158 (0.36)	0.359 (0.48)	0.267 (0.44)
Aerobic exercise	0.047 (0.21)	**0.676** (0.47)	**0.481** (0.50)	0.443 (0.50)
Aquatic exercise	0.550 (0.50)	0.121 (0.33)	0.137 (0.34)	0.031 (0.17)
Strength exercise	0.412 (0.49)	0.110 (0.31)	0.218 (0.41)	0.462 (0.50)
Proprioception exercise	**0.967** (0.18)	0.297 (0.46)	0.470 (0.50)	**0.560** (0.50)
Ultrasound	0.000 (0.02)	0.105 (0.31)	0.114 (0.32)	0.112 (0.32)

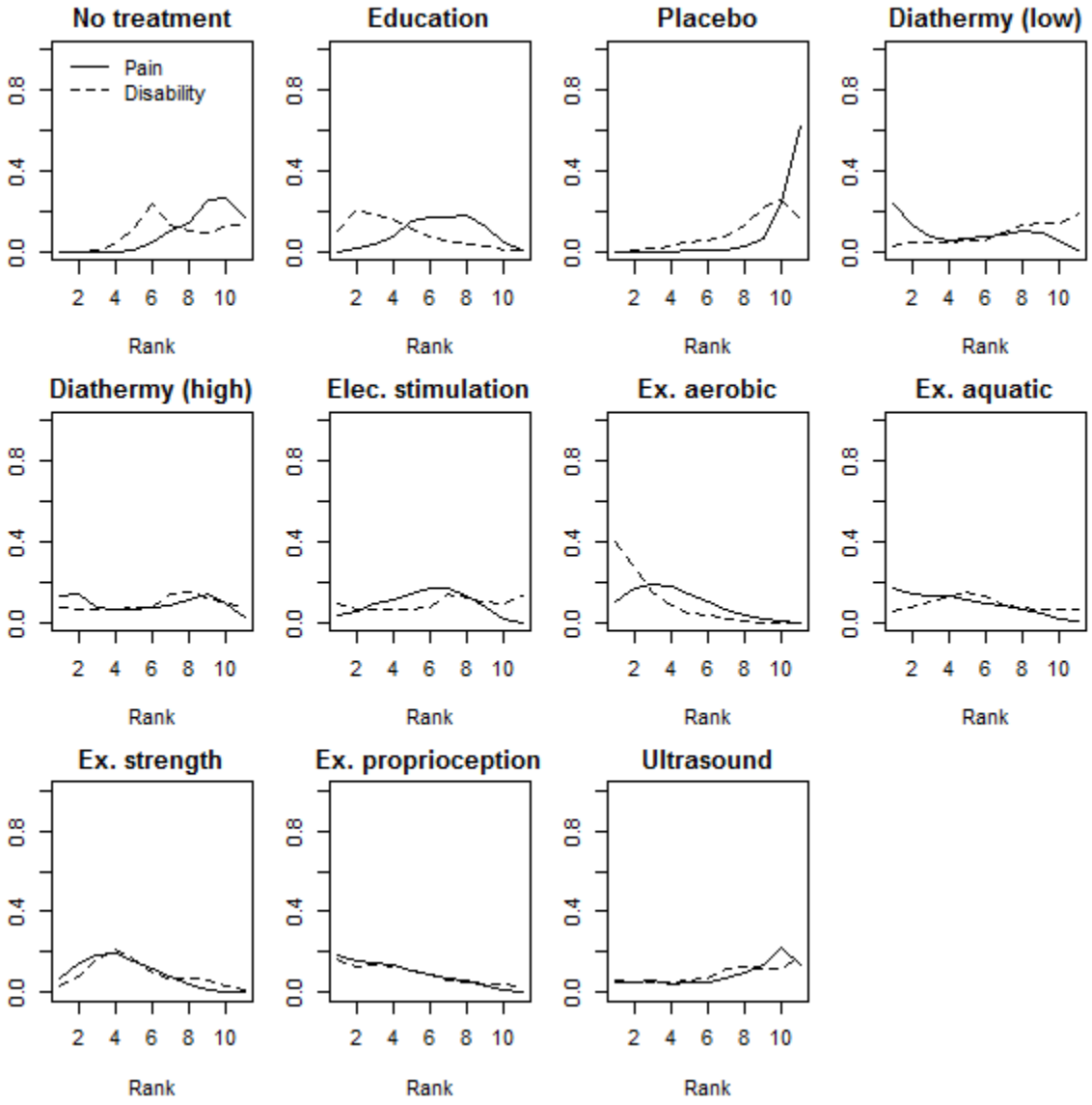
Note: Standard error is in parentheses, and the "best" treatment in terms of Best12. Probability is in bold

Figure 3. OA data interval plot of difference between fixed mean of therapies and no treatment for each outcome



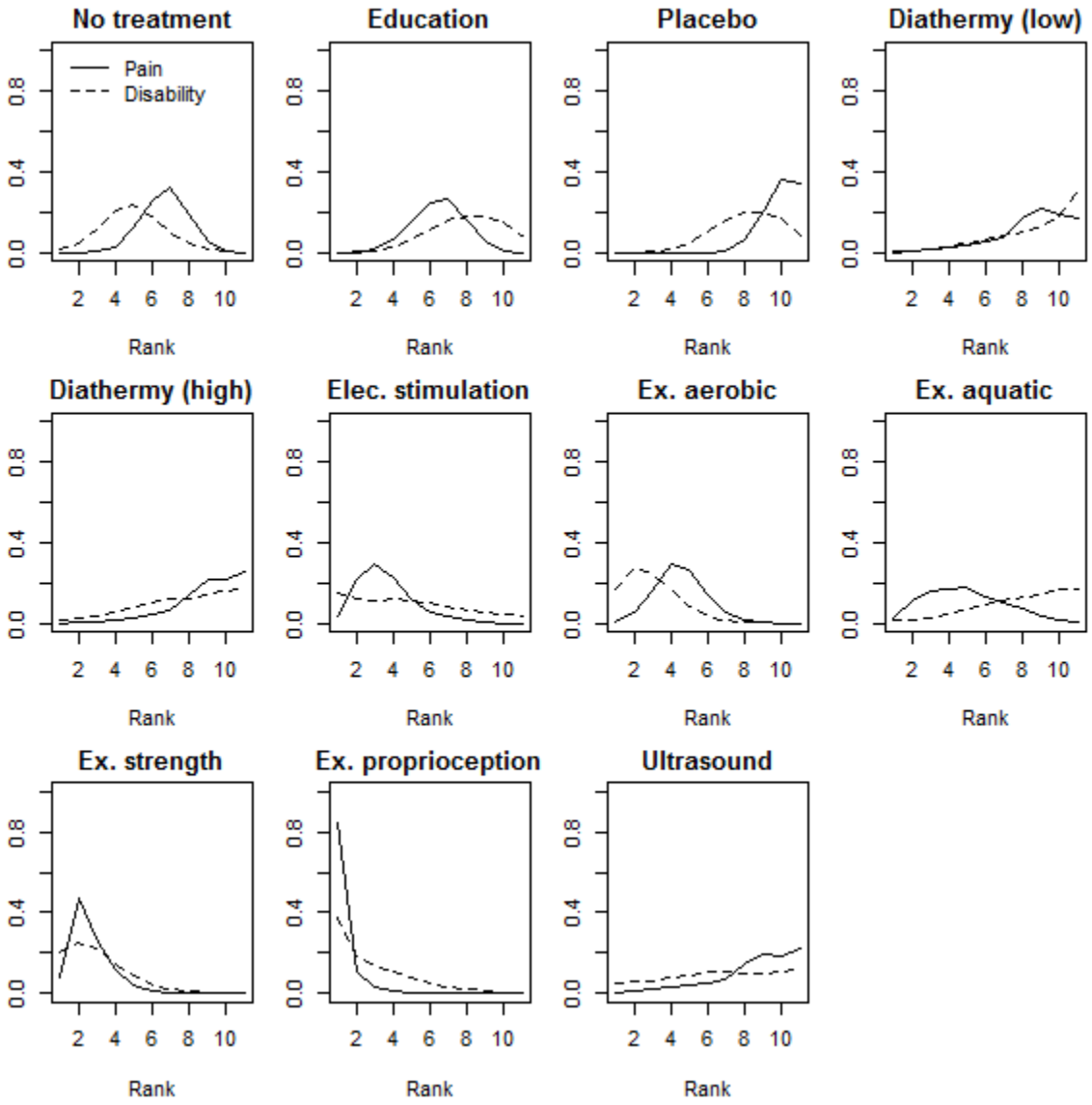
Abbreviations: ABRE2 = arm-based random effects model assuming dependence between outcomes; CBRE2 = contrast-based random effects model assuming dependence between outcomes; hom = homogeneous variance or covariance matrix; LAFE = Lu and Ades-style fixed effects model; LARE = Lu and Ades-style random effects model

Figure 4. Ranking of treatments for reducing pain and improving disability from the homogeneous Lu and Ades-style random effects model (LAREhom)



Note: The vertical axis gives the posterior probability of the indicated treatment taking each of the ranks on the horizontal axis, where 1 is best and 11 is worst

Figure 5. Ranking of treatments for reducing pain and improving disability from the homogeneous arm-based random effects model 2 (ABRE2hom)



Note: The vertical axis gives the posterior probability of the indicated treatment taking each of the ranks on the horizontal axis, where 1 is best and 11 is worst

Table 5. Results from sensitivity analysis

	Noninformative Prior	Informative Prior		
	ρ_0 = 0, γ = 2	ρ_0 = 0.5, γ = 2	ρ_0 = 0.5, γ = 4	ρ_0 = 0.9, γ = 4
CBRE2hom				
pD	153.9	152.5	151.9	147.9
$\bar{D}$	169.0	168.0	168.1	167.9
DIC	322.9	320.5	320.0	315.8
$\hat{\rho}$	0.494 (0.18 - 0.71)	0.670 (0.41 – 0.82)	0.675 (0.43 – 0.82)	0.879 (0.73 – 0.94)
ABRE2hom				
pD	158.0	157.7	157.5	157.2
$\bar{D}$	167.8	167.9	168.3	168.4
DIC	325.8	325.6	325.8	325.6
$\hat{\rho}$	0.377 (0.06 - 0.61),	0.449 (0.45 – 0.66)	0.459 (0.16 – 0.67)	0.518 (0.23 – 0.71)

ρ_0 = prior guess of between-outcome correlation; γ = degrees of freedom in Wishart prior; $\hat{\rho}$ = median posterior of the correlation with 95% credible interval in parentheses

Table 6. Simulation results when $\rho^*_{AB} = 0.6$; Type I error, Power1, and Power2 in terms of d_{21}; Pr($\widehat{\mu_{11}} > \widehat{\mu_{21}}$) is in parentheses

		Noninformative Wishart Prior		Weakly Informative Wishart Prior	
	LAREhom	CBRE2	ABRE2	CBRE2	ABRE2
Type I error ($d^*_{21} = 0$)					
Complete	0.042 (0.494)	0.022 (0.494)	0.022 (0.528)	0.027 (0.493)	0.022 (0.482)
MCAR	0.044 (0.487)	0.023 (0.490)	0.014 (0.523)	0.029 (0.482)	0.023 (0.460)
MAR	0.335 (0.050)	0.040 (0.360)	0.041 (0.342)	0.025 (0.523)	0.024 (0.475)
MNAR	0.487 (0.977)	0.003 (0.809)	0.001 (0.829)	0.013 (0.693)	0.010 (0.693)
Power1 ($d^*_{21} = 1$)					
Complete	0.881 (0.000)	0.883 (0.000)	0.890 (0.000)	0.893 (0.000)	0.892 (0.000)
MCAR	0.555 (0.014)	0.625 (0.010)	0.569 (0.012)	0.708 (0.011)	0.667 (0.010)
MAR	0.967 (0.000)	0.575 (0.006)	0.651 (0.004)	0.482 (0.025)	0.580 (0.009)
MNAR	0.057 (0.377)	0.237 (0.084)	0.209 (0.082)	0.430 (0.041)	0.433 (0.032)
Power2 ($d^*_{21} = 2$)					
Complete	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
MCAR	0.985 (0.000)	0.994 (0.000)	0.990 (0.000)	0.995 (0.000)	0.995 (0.000)
MAR	1.000 (0.000)	0.978 (0.000)	0.992 (0.000)	0.961 (0.000)	0.989 (0.000)
MNAR	0.733 (0.002)	0.925 (0.000)	0.937 (0.000)	0.981 (0.000)	0.989 (0.000)

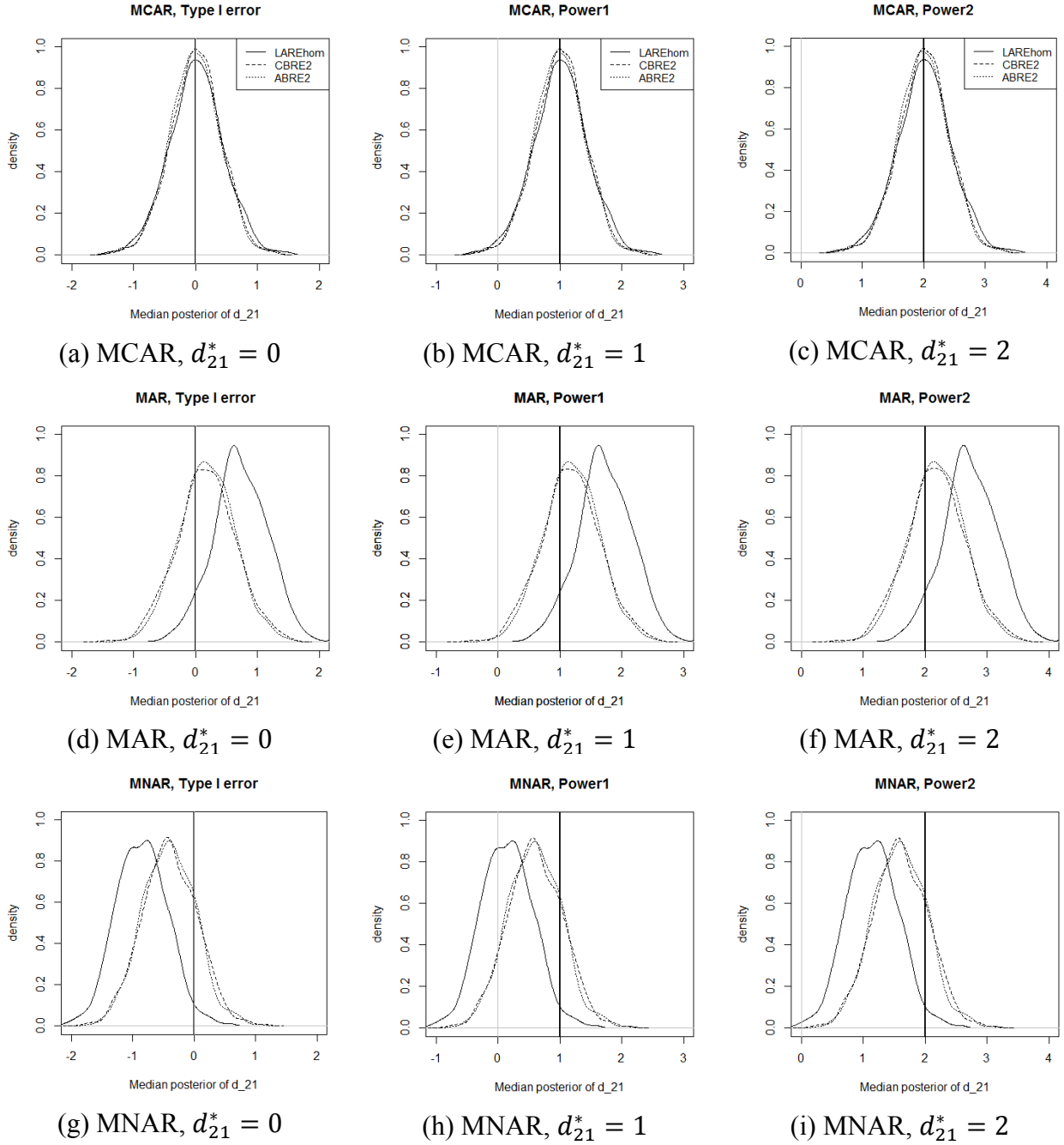
Abbreviations: MAR = missing at random; MCAR = missing completely at random; MNAR = missing not at random

Table 7. Simulation results when $\rho^*_{AB} = 0.0$; Type I error, Power1, and Power2 in terms of d_{21}; Pr($\widehat{\mu_{11}} > \widehat{\mu_{21}}$) is in parentheses

		Noninformative Wishart Prior		Weakly Informative Wishart Prior	
	LAREhom	CBRE2	ABRE2	CBRE2	ABRE2
Type I error ($d^*_{21} = 0$)					
Complete	0.040 (0.484)	0.020 (0.482)	0.022 (0.507)	0.022 (0.483)	0.027 (0.485)
MCAR	0.044 (0.497)	0.023 (0.503)	0.016 (0.528)	0.027 (0.504)	0.023 (0.502)
MAR	0.045 (0.488)	0.021 (0.480)	0.022 (0.503)	0.030 (0.482)	0.037 (0.476)
MNAR	0.089 (0.762)	0.003 (0.780)	0.003 (0.762)	0.004 (0.781)	0.005 (0.740)
Power1 ($d^*_{21} = 1$)					
Complete	0.880 (0.000)	0.885 (0.000)	0.900 (0.000)	0.897 (0.000)	0.918 (0.000)
MCAR	0.531 (0.010)	0.558 (0.011)	0.521 (0.016)	0.607 (0.012)	0.607 (0.012)
MAR	0.547 (0.011)	0.416 (0.024)	0.460 (0.015)	0.463 (0.025)	0.534 (0.011)
MNAR	0.273 (0.057)	0.204 (0.091)	0.226 (0.091)	0.238 (0.094)	0.308 (0.077)
Power2 ($d^*_{21} = 2$)					
Complete	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
MCAR	0.985 (0.000)	0.987 (0.000)	0.987 (0.000)	0.991 (0.000)	0.989 (0.000)
MAR	0.988 (0.000)	0.937 (0.000)	0.972 (0.000)	0.944 (0.000)	0.985 (0.000)
MNAR	0.945 (0.000)	0.864 (0.000)	0.891 (0.000)	0.891 (0.000)	0.933 (0.001)

Abbreviations: MAR = missing at random; MCAR = missing completely at random; MNAR = missing not at random

Figure 6. Density plot of 1,000 median posteriors of d_{21} from simulations when $\rho^*_{AB} = 0.6$ under MCAR (first row), MAR (second row), and MNAR (third row) mechanisms under noninformative Wishart priors; (a), (d), (g) $d^*_{21} = 0$, (b), (e), (h) $d^*_{21} = 1$, and (c), (f), (i) $d^*_{21} = 2$



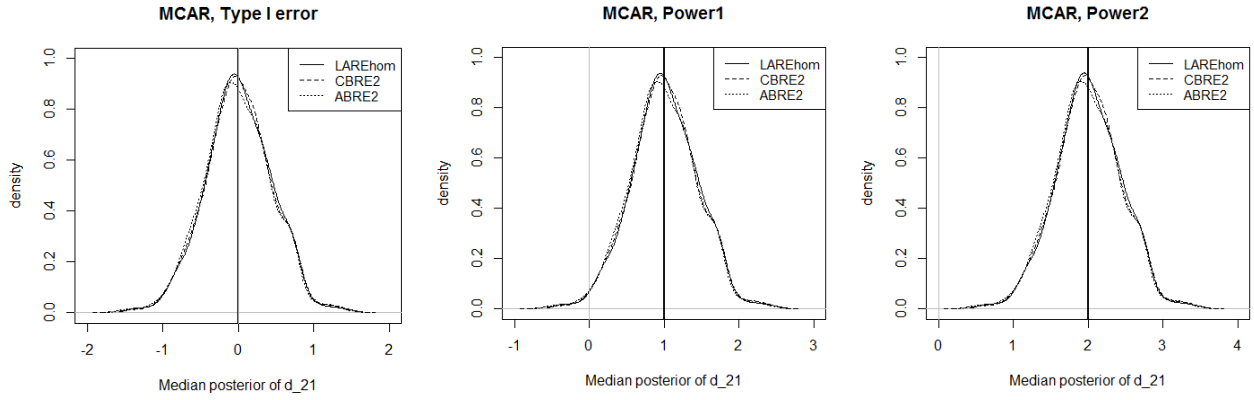
(a) MCAR, $d^*_{21} = 0$ (b) MCAR, $d^*_{21} = 1$ (c) MCAR, $d^*_{21} = 2$

(d) MAR, $d^*_{21} = 0$ (e) MAR, $d^*_{21} = 1$ (f) MAR, $d^*_{21} = 2$

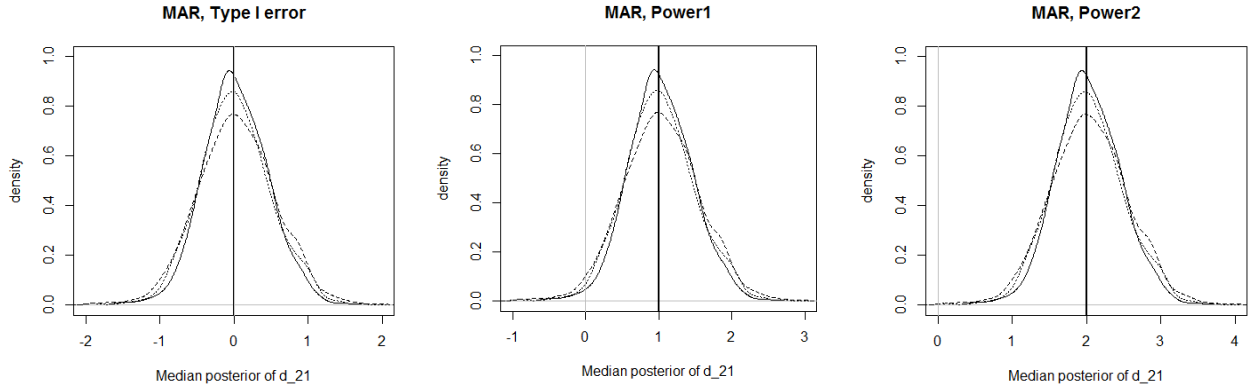
(g) MNAR, $d^*_{21} = 0$ (h) MNAR, $d^*_{21} = 1$ (i) MNAR, $d^*_{21} = 2$

Abbreviations: MAR = missing at random; MCAR = missing completely at random;MNAR = missing not at random

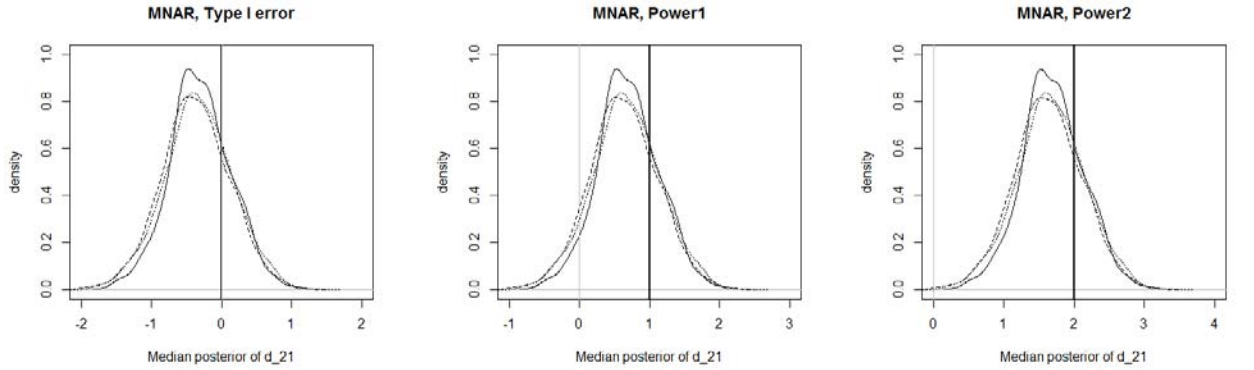
Figure 7. Density plot of 1,000 median posteriors of d_{21} from simulations when $\rho^*_{AB} = 0.0$ under MCAR (first row), MAR (second row), and MNAR (third row) mechanisms under noninformative Wishart priors; (a), (d), (g) $d^*_{21} = 0$, (b), (e), (h) $d^*_{21} = 1$, and (c), (f), (i) $d^*_{21} = 2$



(a) MCAR, $d^*_{21} = 0$ (b) MCAR, $d^*_{21} = 1$ (c) MCAR, $d^*_{21} = 2$



(d) MAR, $d^*_{21} = 0$ (e) MAR, $d^*_{21} = 1$ (f) MAR, $d^*_{21} = 2$



(g) MNAR, $d^*_{21} = 0$ (h) MNAR, $d^*_{21} = 1$ (i) MNAR, $d^*_{21} = 2$

Abbreviations: MAR = missing at random; MCAR = missing completely at random; MNAR = missing not at random

Discussion

The main objective of this report has been to propose new Bayesian MTC approaches for multiple continuous outcomes, and compare them with previous hierarchical modeling methods. We considered unobserved arms to be missing data and handled them by borrowing information from the observed indirect relationships. We also combined multiple outcomes into one model by incorporating the correlation structures between them. Next, we developed arm-based (AB) models that estimate absolute effects of treatments, rather than relative effects. We illustrated our methods using the OA data, and used simulation to show that our models can outperform existing Lu and Ades-style models in terms of Type I error, power, and probability of incorrectly selecting the best treatment under various missing data mechanisms.

We fit six models to the OA data, with LAREhom, CBRE2hom, and ABRE2hom producing slightly smaller DIC values. The fixed effects model performs poorly because it can never fully capture variability across studies. In the random effects models, a homogeneous variance (or covariance matrix) assumption is quite reasonable because our data are so sparse that heterogeneous covariances may not be well estimated. Regarding the pain outcome, low intensity diathermy emerged as the best therapy in LA models, whereas proprioception exercise performed best under CB and ABRE2hom models, followed by strength exercise. However, there were no significant differences between most active therapies, due to the large associated standard deviations (e.g., Figure 3). Note that three studies reported diathermy intervention with only a short length of followup (0 to 5 weeks), so we can only see the short-term effect of diathermy here. By contrast, most studies for proprioception or strength exercises reported a followup period of 6 to 12 weeks. For the disability outcome, aerobic and proprioception exercises perform well across all three random effects models, though again significant differences were rare. Unfortunately, our OA data analysis did not show much impact of our methods compared with the existing methods due to sparseness of the data, although we have shown our methods give less biased estimates through simulation studies.

Our simulation study shows that ignoring missing data and correlations between outcomes can cause biased estimates, resulting in bad hypothesis test performance when missingness of treatment arms depends on the observed (and even missing) data. Although our simulation setting is simple, this problem could be more severe for more complicated data structures. Also, CB models cannot capture the correct correlation in some settings due to their inherent constraints, while AB models can. For example, in our simulation setting, CB models cannot estimate ρ_{CB} if we set $\rho^*_{AB} = 0.9$ because this violates the positive definiteness of the CB covariance matrix. Although our methods perform almost equally when two outcomes are independent ($\rho^*_{AB} = 0$), our methods still outperform the existing LA methods in terms of Type I error, power, and $\Pr(\widehat{\mu_{11}} > \widehat{\mu_{21}})$. Generally, the AB models with weakly informative priors help to yield more reliable estimates resulting in more power.

Regarding the missingness mechanism, we generally assume that the data have MAR missingness. The MCAR assumption might be valid but could be too strong in some cases. For example, in our simulation missingness mechanism (9), the probability of missingness in the first outcome increases as a population has higher and lower second-outcome responses in the first and second treatments, respectively.

Our methods have several limitations. First, since we have only summary statistics for every study, there is the possibility of ecological fallacy. Second, all our models are fitted under the assumption of consistency. Although we do not follow the Lu-and-Ades consistency equation, measuring inconsistency between direct and indirect comparisons in MTCs with incorporating

missingness and multiple outcomes is a topic for a future manuscript. Furthermore, we will try to distinguish the data-driven missingness mechanism by using this inconsistency information. Third, in our CB and AB random effect models, we assumed that either the between-outcome or between-treatment correlations were all zero a priori. However, such assumptions can be loosened by factorizing the random effects into two independent sources. For example, in the AB model, (6) can be rewritten as $\Delta_{ikl} = \mu_{kl} + \nu_{ik} + w_{il}$, where $(\nu_{i1}, \ldots, \nu_{iK})^T \sim \text{MVN}(\mathbf{0}, \boldsymbol{D}^{Trt})$, $(w_{i1}, \ldots, w_{iL})^T \sim \text{MVN}(\mathbf{0}, \boldsymbol{D}^{Out})$, and ν_{ik} and w_{il} are independent. Here, $\boldsymbol{D}^{Trt}$ and $\boldsymbol{D}^{Out}$ are $K \times K$ and $L \times L$ unstructured covariance matrices implying correlation between treatments and outcomes, respectively, where each covariance matrix has an inverse Wishart prior. In this approach, we must select these Wishart priors carefully to ensure identifiability, and this is a subject of ongoing investigation. Fourth, we assumed that the within-study correlations are zero in likelihood. However, Riley et al. discussed when we can estimate within-study correlation and thus produce estimates with smaller standard errors than in the independent setting for bivariate random effect meta-analysis.[82] Finally, we have discussed borrowing strength from the missingness, but this does not mean that our estimates always have narrower 95% credible interval than those from the existing model. If there is not enough observed data, our methods could have a lot of uncertainty, resulting in wider 95% credible intervals.

In the standard meta-analysis with a continuous outcome, standardized mean differences (SMDs) are often calculated and used for analysis and inference.[83] However, we avoid using those quantities in our method because it does not fully handle situations with multi-arm trials and uncommon baseline treatments across studies. For example, in a three-arm study, three SMD values can be calculated, but only by reusing the data, violating the Likelihood Principle. Also, it is not reasonable to combine SMD values that can possibly have different control arms (or baseline treatment) across studies.

Our data analysis also has some limitations. First, we assumed that patients in each intervention from each study had similar clinical characteristics, so we did not adjust our models for such baseline covariates, (e.g., age, severity of OA, or comorbidities). Meta-regression[6] is usually applied to see associations between those sample covariates and treatment effects, but it does not detect the relationship well here because we have only aggregated information.[84] To see such relationships correctly, *individual-level* data should be incorporated. Second, we assumed a common covariance matrix across treatments in our CB and ABRE2 models. This might not be a valid assumption because differences in outcome correlations between treatments could exist. Next, we did not control for the effect of varying followup times but instead selected a frequently observed followup time for each treatment when studies reported outcomes from multiple followup times. Although we made an effort to have similar followup times within each treatment, not all studies had precisely the same followup time for a specific treatment. However, a majority of studies investigated only one followup time, and in any case our data were not intended to measure the effect of followup time. Also, the outcomes from different followup times are likely to be correlated because they are typically obtained from the same sample of patients; modeling this feature is beyond the scope of our present report. Lu et al.[85] suggest various models for MTCs at multiple followup times with single binary outcome. We found that the baseline pain scores from the studies not reporting disability scores are slightly smaller than those from the studies that reported both outcomes. This could imply that the missingness depends on the observed, information implying the MAR mechanism.

Our simulation studies can be improved by including more features. For example, we might extend it to have more than two treatments with a more complicated evidence network so that

inconsistency could be measured. In this report, we only considered 50 percent missingness in the first outcome, all studies have the same sample size and assumed standard deviation, and true d_{21}^{*}values were somewhat arbitrarily selected. We could explore various missingness rates and patterns with some heterogeneity between studies for different sample sizes and standard deviations. Also, we need to examine more d_{21}^{*}values rather than just 1 and 2.

Finally, our models can be applied when the MTC data have multiple outcomes (i.e., efficacy and safety outcomes) with possible correlations but not measured at multiple time points. We can reduce our model to handle a single continuous outcome. Also, our CB and AB models can be applied to single or multiple binary outcome settings by using a logit link function rather than a linear link function.[86] We can also extend our approaches to categorical outcomes. We are currently extending our methods to mixed types of outcomes (say, a binary safety outcome paired with a continuous efficacy outcome). Furthermore, we hope to extend our models to incorporate both aggregated and individual-level (i.e., patient-level) data, potentially permitting borrowing of strength from patient-level covariates to investigate how those personal clinical characteristics impact estimated treatment effects.

References

1. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health. 2011 Jun;14(4):417-28. PMID 21669366.
2. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. Value Health. 2011 Jun;14(4):429-37. PMID 21669367.
3. Ades AE, Caldwell DM, Reken S, et al. NICE DSU Technical Support Document 7: Evidence Synthesis of Treatment Efficacy in Decision Making: A Reviewer's Checklist. United Kingdom; 2012. www.nicedsu.org.uk/TSD7%20reviewer%20checklist.final.08.05.12.pdf. Accessed August 23, 2012.
4. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986 Sep;7(3):177-88. PMID 3802833.
5. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. Stat Med. 1995 Dec 30;14(24):2685-99. PMID 8619108.
6. Nixon RM, Bansback N, Brennan A. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. Stat Med. 2007 Mar 15;26(6):1237-54. PMID 16900557.
7. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med. 2004 Oct 30;23(20):3105-24. PMID 15449338.
8. Lu G, Ades A. Assessing evidence inconsistency in Mixed Treatment Comparisons. J Am Stat Assoc. 2006;101:447-59.
9. Jansen JP, Crawford B, Bergman G, et al. Bayesian meta-analysis of multiple treatment comparisons: an introduction to mixed treatment comparisons. Value Health. 2008 Sep-Oct;11(5):956-64. PMID 18489499.
10. Dias S, Welton NJ, Sutton AJ, et al. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. United Kingdom; 2011 (last updated April 2012). www.nicedsu.org.uk/TSD2%20General%20meta%20analysis.final.08.05.12.pdf. Accessed August 23, 2012.
11. Welton NJ, Cooper NJ, Ades AE, et al. Mixed treatment comparison with multiple outcomes reported inconsistently across trials: evaluation of antivirals for treatment of influenza A and B. Stat Med. 2008 Nov 29;27(27):5620-39. PMID 18680174.
12. Shamliyan TA, Wang S-Y, Olson-Kellogg B, et al. Physical Therapy for Knee Pain Secondary to Osteoarthritis. Comparative Effectiveness Review. Prepared by the Minnesota Evidence-based Practice Center under Contract No. 290-1007-10064 I. Agency for Healthcare Research and Quality. Rockville, MD: In press.
13. Ades AE, Mavranezouli I, Dias S, et al. Network meta-analysis with competing risk outcomes. Value Health. 2010 Dec;13(8):976-83. PMID 20825617.
14. Franchini AJ, Dias S, Ades AE, et al. Accounting for correlation in network meta-analysis with multi-arm trials. Research Synthesis Methods. 2012;3(2):142-60.
15. Little RJ, Rubin DB. Statistical Analysis with Missing Data. 2nd Edition. New York: Wiley; 2002.
16. Bullens PHJ, van Loon CJM, de Waal Malefijt MC, et al. Patient satisfaction after total knee arthroplasty. J Arthroplasty. 2001 Sept;16(6):740-7.
17. Roos EM, Toksvig-Larsen S. Knee injury and Osteoarthritis Outcome Score (KOOS) - validation and comparison to the WOMAC in total knee replacement. Health Qual Life Outcomes. 2003;1:17. PMID 12801417.

18. Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. Health Qual Life Outcomes. 2003;1:64. PMID 14613558.

19. Jackson D, Riley R, White IR. Multivariate meta-analysis: Potential and promise. Stat Med. 2011 Jan 26PMID 21268052.

20. Lu G, Ades A. Modeling between-trial variance structure in mixed treatment comparisons. Biostatistics. 2009 Oct;10(4):792-805. PMID 19687150.

21. Salanti G, Higgins JP, Ades AE, et al. Evaluation of networks of randomized trials. Stat Methods Med Res. 2008 Jun;17(3):279-301. PMID 17925316.

22. Carlin BP, Louis TA. Bayesian Methods for Data Analysis. 3rd ed. Boca Raton, FL: Chapman & Hall/CRC; 2009.

23. Spiegelhalter DJ, Best NG, Carlin BP, et al. Bayesian measures of model complexity and fit. J R Stat Soc Series B Stat Methodol. 2002;64(4):583-639.

24. Celeux G, Forbes F, Robert C, et al. Deviance information criteria for missing data models. Bayesian Analysis 1. 2006;1:651-706.

25. Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol. 2011 Feb;64(2):163-71. PMID 20688472.

26. Gelman A. Running WinBugs and OpenBugs from R. www.stat.columbia.edu/~gelman/bugsR/runningbugs.html. 2007; Accessed April 2012.

27. Lunn DJ, Thomas A, Best N, et al. WinBUGS—a Bayesian modeling framework: concepts, structure, and extensibility. Stat Comput 2000;10(4):325-37.

28. Aglamis B, Toraman NF, Yaman H. The effect of a 12-week supervised multicomponent exercise program on knee OA in Turkish women. J Back Musculoskelet Rehabil. 2008;21(2):121-8.

29. Kovar PA, Allegrante JP, MacKenzie CR, et al. Supervised fitness walking in patients with osteoarthritis of the knee. A randomized, controlled trial.Ann Intern Med. 1992 Apr 1;116(7):529-34. PMID 1543305.

30. Ettinger WH Jr., Burns R, Messier SP, et al. A randomized trial comparing aerobic exercise and resistance exercise with a health education program in older adults with knee osteoarthritis. The Fitness Arthritis and Seniors Trial (FAST). JAMA. 1997;277(1):25-31. PMID 8980206.

31. Sullivan T, Allegrante JP, Peterson MG, et al. One-year followup of patients with osteoarthritis of the knee who participated in a program of supervised fitness walking and supportive patient education. Arthritis Care Res. 1998 Aug;11(4):228-33. PMID 9791321.

32. Patrick DL, Ramsey SD, Spencer AC, et al. Economic evaluation of aquatic exercise for persons with osteoarthritis. Med Care. 2001 May;39(5):413-24. PMID 11317090.

33. Baker KR, Nelson ME, Felson DT, et al. The efficacy of home based progressive strength training in older adults with knee osteoarthritis: a randomized controlled trial. J Rheumatol. 2001 Jul;28(7):1655-65. PMID 11469475.

34. Kuptniratsaikul V, Tosayanonda O, Nilganuwong S, et al. The efficacy of a muscle exercise program to improve functional performance of the knee in patients with osteoarthritis. J Med Assoc Thai. 2002 Jan;85(1):33-40. PMID 12075718.

35. Callaghan MJ, Whittaker PE, Grimes S, et al. An evaluation of pulsed shortwave on knee osteoarthritis using radioleucoscintigraphy: a randomised, double blind, controlled trial. Joint Bone Spine. 2005 Mar;72(2):150-5. PMID 15797496.

36. Laufer Y, Zilberman R, Porat R, et al. Effect of pulsed short-wave diathermy on pain and function of subjects with osteoarthritis of the knee: a placebo-controlled double-blind clinical trial. Clin Rehabil. 2005 May;19(3):255-63. PMID 15859526.

37. Yip YB, Sit JW, Fung KK, et al. Impact of an arthritis self-management programme with an added exercise component for osteoarthritic knee sufferers on improving pain, functional outcomes, and use of health care services: An experimental study. Patient Educ Couns. 2007 Jan;65(1):113-21. PMID 17010554.

38. Rooks DS, Huang J, Bierbaum BE, et al. Effect of preoperative exercise on measures of functional status in men and women undergoing total hip and knee arthroplasty. Arthritis Rheum. 2006 Oct 15;55(5):700-8. PMID 17013852.

39. Brismee JM, Paige RL, Chyu MC, et al. Group and home-based tai chi in elderly subjects with knee osteoarthritis: a randomized controlled trial. Clin Rehabil. 2007 Feb;21(2):99-111. PMID 17264104.

40. Garland D, Holt P, Harrington JT, et al. A 3-month, randomized, double-blind, placebo-controlled study to evaluate the safety and efficacy of a highly optimized, capacitively coupled, pulsed electrical stimulator in patients with osteoarthritis of the knee. Osteoarthritis Cartilage. 2007 Jun;15(6):630-7. PMID 17303443.

41. Doi T, Akai M, Fujino K, et al. Effect of home exercise of quadriceps on knee osteoarthritis compared with nonsteroidal antiinflammatory drugs: a randomized controlled trial. Am J Phys Med Rehabil. 2008;87(4):258-69. PMID 141 PMID: 18356618.

42. Lund H, Weile U, Christensen R, et al. A randomized controlled trial of aquatic and land-based exercise in patients with knee osteoarthritis. J Rehabil Med. 2008 Feb;40(2):137-44. PMID 18509579.

43. Yip YB, Sit JW, Wong DY, et al. A 1-year follow-up of an experimental study of a self-management arthritis programme with an added exercise component of clients with osteoarthritis of the knee. Psychol Health Med. 2008;13(4):402-14. PMID 18825579.

44. Özgönenel L, Aytekin E, Durmusoglu G. A double-blind trial of clinical effects of therapeutic ultrasound in knee osteoarthritis. Ultrasound Med Biol. 2009 Jan;35(1):44-9. PMID 18829151.

45. Selfe TK, Bourguignon C, Taylor AG. Effects of noninvasive interactive neurostimulation on symptoms of osteoarthritis of the knee: a randomized, sham-controlled pilot study. J Altern Complement Med. 2008;14(9):1075-81. PMID 98 PMID: 19055333.

46. Péloquin L, Bravo G, Gauthier P, et al. Effects of a cross-training exercise program in persons with osteoarthritis of the knee a randomized controlled trial. J Clin Rheumatol. 1999 Jun;5(3):126-36. PMID 19078371.

47. Chaipinyo K, Karoonsupcharoen O. No difference between home-based strength training and home-based balance training on pain in patients with knee osteoarthritis: a randomised trial. Aust J Physiother. 2009;55(1):25-30. PMID 83 PMID: 19226239.

48. Lee HJ, Park HJ, Chae Y, et al. Tai Chi Qigong for the quality of life of patients with knee osteoarthritis: a pilot, randomized, waiting list controlled trial. Clin Rehabil. 2009;23(6):504-11. PMID 71 PMID: 19389743.

49. Tascioglu F, Kuzgun S, Armagan O, et al. Short-term effectiveness of ultrasound therapy in knee osteoarthritis. J Int Med Res. 2010 Jul-Aug;38(4):1233-42. PMID 20925995.

50. Fukuda TY, Alves da Cunha R, Fukuda VO, et al. Pulsed shortwave treatment in women with knee osteoarthritis: a multicenter, randomized, placebo-controlled clinical trial. Phys Ther. 2011 Jul;91(7):1009-17. PMID 21642511.

51. Messier SP, Thompson CD, Ettinger WH. Effects of long-term aerobic or weight training regimes on gait in an older, osteoarthritic population. J Appl Biomech. 1997;13(2):205-25.

52. Grimmer K. A controlled double blind study comparing the effects of strong burst mode TENS and high rate TENS on painful osteoarthritic knees. Aust J Physiother. 1992;38(1):49-56.

53. Taylor P, Hallett M, Flaherty L. Treatment of osteoarthritis of the knee with transcutaneous electrical nerve stimulation. Pain. 1981 Oct;11(2):233-40. PMID 7033891.

54. Borjesson M, Robertson E, Weidenhielm L, et al. Physiotherapy in knee osteoarthrosis: effect on pain and walking. Physiother Res Int. 1996;1(2):89-97. PMID 9238726.

55. Bautch JC, Malone DG, Vailas AC. Effects of exercise on knee joints with osteoarthritis: a pilot study of biologic markers. Arthritis Care Res. 1997 Feb;10(1):48-55. PMID 9313390.

56. Wyatt FB, Milam S, Manske RC, et al. The effects of aquatic and traditional exercise programs on persons with knee osteoarthritis. J Strength Cond Res. 2001;15(3):337-40. PMID 11710661.

57. Gür H, Cakin N, Akova B, et al. Concentric versus combined concentric-eccentric isokinetic training: effects on functional capacity and symptoms in patients with osteoarthrosis of the knee. Arch Phys Med Rehabil. 2002 Mar;83(3):308-16. PMID 11887109.

58. Topp R, Woolley S, Hornyak J 3rd, et al. The effect of dynamic versus isometric resistance training on pain and functioning among adults with osteoarthritis of the knee. Arch Phys Med Rehabil. 2002 Sep;83(9):1187-95. PMID 12235596.

59. Talbot LA, Gaines JM, Huynh TN, et al. A home-based pedometer-driven walking program to increase physical activity in older adults with osteoarthritis of the knee: a preliminary study. J Am Geriatr Soc. 2003 Mar;51(3):387-92. PMID 12588583.

60. Talbot LA, Gaines JM, Ling SM, et al. A home-based protocol of electrical muscle stimulation for quadriceps muscle strength in older adults with osteoarthritis of the knee. J Rheumatol. 2003 Jul;30(7):1571-8. PMID 12858461.

61. Messier SP, Loeser RF, Miller GD, et al. Exercise and dietary weight loss in overweight and obese older adults with knee osteoarthritis: the Arthritis, Diet, and Activity Promotion Trial. Arthritis Rheum. 2004 May;50(5):1501-10. PMID 15146420.

62. Keefe FJ, Blumenthal J, Baucom D, et al. Effects of spouse-assisted coping skills training and exercise training in patients with osteoarthritic knee pain: a randomized controlled study. Pain. 2004 Aug;110(3):539-49. PMID 15288394.

63. Gaines JM, Metter EJ, Talbot LA. The effect of neuromuscular electrical stimulation on arthritis knee pain in older adults with osteoarthritis of the knee. Appl Nurs Res. 2004 Aug;17(3):201-6. PMID 15343554.

64. Law PP, Cheing GL. Optimal stimulation frequency of transcutaneous electrical nerve stimulation on people with knee osteoarthritis. J Rehabil Med. 2004 Sep;36(5):220-5. PMID 15626162.

65. Durmus D, Alayli G, Canturk F. Effects of quadriceps electrical stimulation program on clinical parameters in the patients with knee osteoarthritis. Clin Rheumatol. 2007 May;26(5):674-8. PMID 16897119.

66. Hay EM, Foster NE, Thomas E, et al. Effectiveness of community physiotherapy and enhanced pharmacy review for knee pain in people aged over 55 presenting to primary care: pragmatic randomised trial. BMJ. 2006 Nov 11;333(7576):995. PMID 17056608.

67. Silva LE, Valim V, Pessanha AP, et al. Hydrotherapy versus conventional land-based exercise for the management of patients with osteoarthritis of the knee: a randomized clinical trial. Physical Therapy. 2008;88(1):12-21. PMID 17986497.

68. Jan MH, Lin JJ, Liau JJ, et al. Investigation of clinical effects of high- and low-resistance training for patients with knee osteoarthritis: a randomized controlled trial. Phys Ther. 2008;88(4):427-36. PMID 153 PMID: 18218827.

69. Itoh K, Hirota S, Katsumi Y, et al. A pilot study on using acupuncture and transcutaneous electrical nerve stimulation (TENS) to treat knee osteoarthritis (OA). Chin Med. 2008;3:2. PMID 18312661.

70. An B, Dai K, Zhu Z, et al. Baduanjin alleviates the symptoms of knee osteoarthritis. J Altern Complement Med. 2008 Mar;14(2):167-74. PMID 18315512.

71. Tsauo JY, Cheng PF, Yang RS. The effects of sensorimotor training on knee proprioception and function for patients with knee osteoarthritis: a preliminary report. Clin Rehabil. 2008;22(5):448-57. PMID: 18441041.

72. Lim BW, Hinman RS, Wrigley TV, et al. Does knee malalignment mediate the effects of quadriceps strengthening on knee adduction moment, pain, and function in medial knee osteoarthritis? A randomized controlled trial. Arthritis Rheum. 2008;59(7):943-51. PMID 122 PMID: 18576289.

73. Pietrosimone BG, Hart JM, Saliba SA, et al. Immediate effects of transcutaneous electrical nerve stimulation and focal knee joint cooling on quadriceps activation. Med Sci Sports Exerc. 2009 Jun;41(6):1175-81. PMID 19461552.

74. Lin DH, Lin CH, Lin YF, et al. Efficacy of 2 non-weight-bearing interventions, proprioception training versus strength training, for patients with knee osteoarthritis: a randomized clinical trial. J Orthop Sports Phys Ther. 2009;39(6):450-7. PMID: 19531879.

75. Weng MC, Lee CL, Chen CH, et al. Effects of different stretching techniques on the outcomes of isokinetic exercise in patients with knee osteoarthritis. Kaohsiung J Med Sci. 2009;25(6):306-15. PMID: 19560995.

76. Farr JN, Going SB, McKnight PE, et al. Progressive resistance training improves overall physical activity levels in patients with early osteoarthritis of the knee: a randomized controlled trial. Phys Ther. 2010 Mar;90(3):356-66. PMID 20056719.

77. Bennell KL, Hunt MA, Wrigley TV, et al. Hip strengthening reduces symptoms but not knee load in people with medial knee osteoarthritis and varus malalignment: a randomised controlled trial. Osteoarthritis Cartilage. 2010;May;18(5):621-8. Epub 2010 Feb 6.(5):621-8. PMID 15 - 20175973.

78. Swank AM, Kachelman JB, Bibeau W, et al. Prehabilitation before total knee arthroplasty increases strength and function in older adults with severe osteoarthritis. J Strength Cond Res. 2011 Feb;25(2):318-25. PMID 21217530.

79. Schilke JM, Johnson GO, Housh TJ, et al. Effects of muscle-strength training on the functional status of patients with osteoarthritis of the knee joint. Nurs Res. 1996 Mar-Apr;45(2):68-72. PMID 8604366.

80. Deyle GD, Henderson NE, Matekel RL, et al. Effectiveness of manual physical therapy and exercise in osteoarthritis of the knee. A randomized, controlled trial. Ann Intern Med. 2000;132(3):173-81. PMID 10651597.

81. Rejeski WJ, Focht BC, Messier SP, et al. Obese, older adults with knee osteoarthritis: weight loss, exercise, and quality of life. Health Psychol. 2002;21(5):419-26. PMID 12211508.

82. Riley RD, Abrams KR, Lambert PC, et al. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. Stat Med. 2007 Jan 15;26(1):78-97. PMID 16526010.

83. Whitehead A. Meta-Analysis of Controlled Clinical Trials: John Wiley; 2002.

84. Simmonds MC, Higgins JP. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. Stat Med. 2007 Jul 10;26(15):2982-99. PMID 17195960.

85. Lu G, Ades AE, Sutton AJ, et al. Meta-analysis of mixed treatment comparisons at multiple follow-up times. Stat Med. 2007 Sep 10;26(20):3681-99. PMID 17285571.

86. Terasawa T, Trikalinos NA, Djulbegovic B, et al. Comparative efficacy of first-line therapies for advanced-stage chronic lymphocytic leukemia: A multiple-treatment meta-analysis. Cancer Treat Rev. 2012 Jun 13PMID 22703832.

Abbreviations

AB	Arm based
ABRE	Arm-based random effects
AIMS	Arthritis Impact Measurement Scale
CB	Contrast based
CBRE	Contrast-based random effects
DIC	Deviance Information Criterion
EPC	Evidence-based Practice Center
HAQ	Health Assessment Questionnaire
KOOS	Knee Injury and Osteoarthritis Outcome Score
LA	Lu and Ades
LAFE	LA-fixed effects
LAREhom	Lu and Ades-style homogeneous random effects
MAR	Missing at random
MCAR	Missing completely at random
MCMC	Markov chain Monte Carlo
MNAR	Missing not at random
MOS	Medical Outcome Study
MSE	Mean squared error
MTC	Mixed treatment comparison
OA	Osteoarthritis
RCT	Randomized controlled trial
SD	Standard deviation
SF-36 physical function	36-Item Short-Form Health Survey
SMD	Standardized mean differences
VAS	Visual Analogue Scale
WOMAC	Western Ontario MacMaster

Appendix A. WinBUGS Codes

```
#BUGS code for LAREhom

model {

  # Pain
  sw1[1] <- 0
  for (i in 1:N1) {
    mean1[i] <- mu1[s1[i]] + delta1[i]*(1-equals(t1[i],b1[i]))
    y1[i] ~ dnorm(mean1[i], prec1[i])
    se1[i] <- sd1[i]/sqrt(n1[i])
    prec1[i] <- 1/pow(se1[i],2)
    delta1[i] ~ dnorm(md1[i], taud1[i])
    taud1[i] <- tau1 * (1+equals(m1[i],3)/3)
    md1[i] <- d1[t1[i]] - d1[b1[i]] + equals(m1[i],3)*sw1[i]

    fitted1[i] <- mean1[i]
    res1[i] <- y1[i] - mean1[i]
    dev1[i] <- (y1[i] - mean1[i])*(y1[i] - mean1[i])*prec1[i]
  }

  resdev1 <- sum(dev1[])

  for (i in 2:N1) { sw1[i] <- (delta1[i-1] - d1[t1[i-1]] + d1[b1[i-1]])/2 }
  for (j in 1:NS1) { mu1[j] ~ dnorm(0, 0.0001) }

  d1[1] <- 0
  for (k in 2:NT) { d1[k] ~ dnorm(0, 0.0001) }

  sig1 ~ dunif(0.01, 10)
  tau1 <- 1/pow(sig1, 2)

  # Disability
  sw2[1] <- 0
  for (i in 1:N2) {
    mean2[i] <- mu2[s2[i]] + delta2[i]*(1-equals(t2[i],b2[i]))
    y2[i] ~ dnorm(mean2[i], prec2[i])
    se2[i] <- sd2[i]/sqrt(n2[i])
    prec2[i] <- 1/pow(se2[i],2)
    delta2[i] ~ dnorm(md2[i], taud2[i])
    taud2[i] <- tau2 * (1+equals(m2[i],3)/3)
    md2[i] <- d2[t2[i]] - d2[b2[i]] + equals(m2[i],3)*sw2[i]

    fitted2[i] <- mean2[i]
    res2[i] <- y2[i] - mean2[i]
```

```
  dev2[i] <- (y2[i] - mean2[i])*(y2[i] - mean2[i])*prec2[i]
}

resdev2 <- sum(dev2[])

for (i in 2:N2) { sw2[i] <- (delta2[i-1] - d2[t2[i-1]] + d2[b2[i-1]])/2 }
for (j in 1:NS2) { mu2[j] ~ dnorm(0, 0.0001) }

d2[1] <- 0
for (k in 2:NT) { d2[k] ~ dnorm(0, 0.0001) }

sig2 ~ dunif(0.01, 10)
tau2 <- 1/pow(sig2, 2)

resdev <- resdev1 + resdev2

# ranking
for (i in 1:nP1) { mmu1[i] <- mu1[study1[i]] }
for (i in 1:nP2) { mmu2[i] <- mu2[study2[i]] }
mP1 <- mean(mmu1[])        # Take average of mu1[]
SD1 <- sd(mmu1[])
mP2 <- mean(mmu2[])        # Take average of mu1[]
SD2 <- sd(mmu2[])

for (k in 1:NT) {
  T.1[k] <- mP1 + d1[k]
  rk.1[k] <- rank(T.1[], k)
  best1.1[k] <- equals(rk.1[k], 1)
  best2.1[k] <- equals(rk.1[k], 2)
  best12.1[k] <- best1.1[k] + best2.1[k]

  T.2[k] <- mP2 + d2[k]
  rk.2[k] <- rank(T.2[], k)
  best1.2[k] <- equals(rk.2[k], 1)
  best2.2[k] <- equals(rk.2[k], 2)
  best12.2[k] <- best1.2[k] + best2.2[k]

  T.eq[k] <- 0.5*T.1[k] + 0.5*T.2[k]
  rk.eq[k] <- rank(T.eq[], k)
  best1.eq[k] <- equals(rk.eq[k], 1)
  best2.eq[k] <- equals(rk.eq[k], 2)
  best12.eq[k] <- best1.eq[k] + best2.eq[k]

  T.pain[k] <- 0.8*T.1[k] + 0.2*T.2[k]
  rk.pain[k] <- rank(T.pain[], k)
  best1.pain[k] <- equals(rk.pain[k], 1)
```

```
    best2.pain[k] <- equals(rk.pain[k], 2)
    best12.pain[k] <- best1.pain[k] + best2.pain[k]

    T.dis[k] <- 0.2*T.1[k] + 0.8*T.2[k]
    rk.dis[k] <- rank(T.dis[], k)
    best1.dis[k] <- equals(rk.dis[k], 1)
    best2.dis[k] <- equals(rk.dis[k], 2)
    best12.dis[k] <- best1.dis[k] + best2.dis[k]
  }
}

# Data

list(N1=111, NS1=51, N2=57, NS2=26, NT=11,
     study1 = c(1,2,5,7,10, 12,14,15,16,21, 28,31,32,35,36, 40,42,43,44,45,
                  46,47,48,49,51), nP1=25,
      study2 = c(1,2,5,7,10, 12,14,15,16,21, 24), nP2=11)

# Data - pain
```

s1[]	t1[]	y1[]	sd1[]	n1[]	b1[]	m1[]
1	1	7.700	2.300	17	1	1
1	7	0.700	1.000	17	1	2
2	1	4.770	2.120	50	1	1
2	7	3.770	1.730	52	1	2
3	2	2.800	1.220	149	2	1
3	7	2.280	1.200	144	2	2
3	9	2.420	1.440	146	2	3
4	2	5.500	2.070	50	2	1
4	7	4.590	2.400	52	2	2
5	1	4.873	2.063	124	1	1
5	8	4.607	2.457	125	1	2
6	2	3.780	2.345	23	2	1
6	9	2.560	1.962	23	2	2
7	1	5.070	2.530	193	1	1
7	9	4.060	2.530	199	1	2
8	3	6.300	1.900	10	3	1
8	4	5.000	3.200	10	3	2
8	5	5.500	2.700	10	3	3
9	3	4.440	3.510	33	3	1
9	4	4.730	3.480	38	3	2
9	5	4.030	3.300	32	3	3
10	1	4.250	2.367	94	1	1
10	7	3.858	2.201	88	1	2
11	2	3.750	2.500	23	2	1
11	8	3.650	0.350	22	2	2
12	1	3.370	1.780	19	1	1
12	10	2.410	2.050	22	1	2

13	3	4.180	1.659	19	3	1
13	6	3.740	2.360	39	3	2
14	1	2.959	2.394	70	1	1
14	9	2.255	2.068	72	1	2
15	1	2.380	1.403	27	1	1
15	8	1.810	1.403	27	1	2
15	9	1.560	1.400	25	1	3
16	1	3.459	2.355	50	1	1
16	7	3.523	2.193	45	1	2
17	3	4.000	2.600	33	3	1
17	11	3.900	2.000	34	3	2
18	3	3.178	1.784	20	3	1
18	6	2.834	2.136	20	3	2
19	2	3.940	2.220	68	2	1
19	7	3.090	1.540	69	2	2
20	9	1.800	1.600	24	9	1
20	10	1.300	1.200	24	9	2
21	1	1.686	1.057	15	1	1
21	10	1.314	1.143	29	1	2
22	3	6.670	1.780	30	3	1
22	11	5.250	1.900	30	3	2
23	3	6.900	2.000	23	3	1
23	4	3.800	2.200	32	3	2
23	5	4.600	2.500	31	3	3
24	2	2.560	1.800	36	2	1
24	7	2.300	1.954	33	2	2
25	3	3.500	2.900	20	3	1
25	6	2.200	2.800	20	3	2
26	3	6.750	2.375	10	3	1
26	6	5.250	1.425	10	3	2
27	1	3.300	1.500	34	1	1
27	9	3.000	1.500	34	1	2
28	2	2.080	2.090	17	2	1
28	7	2.190	1.670	17	2	2
29	7	3.800	1.600	23	7	1
29	8	2.400	1.600	23	7	2
30	1	4.000	0.743	6	1	1
30	9	1.471	0.643	9	1	2
31	1	5.385	1.528	35	1	1
31	9	5.190	1.657	32	1	2
32	2	1.397	1.242	21	2	1
32	7	1.660	1.463	19	2	2
33	2	1.426	1.026	18	2	1
33	6	2.094	1.712	20	2	2
34	1	3.095	2.030	78	1	1
34	2	2.550	1.945	82	1	2

34	7	3.110	2.010	80	1	3
35	1	4.030	2.080	18	1	1
35	2	4.000	1.560	18	1	2
35	7	3.190	1.850	16	1	3
36	2	5.990	2.400	18	2	1
36	6	5.180	2.110	20	2	2
37	3	4.100	2.600	10	3	1
37	6	0.700	0.700	12	3	2
38	6	0.600	0.100	25	6	1
38	9	1.040	0.270	25	6	2
39	1	4.180	1.950	108	1	1
39	7	3.755	2.400	109	1	2
40	7	3.840	2.750	32	7	1
40	8	3.700	1.810	32	7	2
41	1	3.550	1.700	34	1	1
41	9	2.400	1.750	34	1	2
42	1	4.930	2.020	8	1	1
42	6	5.350	0.970	8	1	2
43	1	2.764	2.252	14	1	1
43	7	1.422	2.202	14	1	2
44	1	1.320	0.760	30	1	1
44	10	1.280	0.740	30	1	2
45	1	3.360	1.540	28	1	1
45	9	2.280	1.690	27	1	2
46	1	2.096	1.844	12	1	1
46	6	1.165	1.671	11	1	2
47	1	3.650	1.700	36	1	1
47	9	2.100	1.500	36	1	2
47	10	2.150	1.150	36	1	3
48	1	4.400	1.400	66	1	1
48	9	3.600	0.700	66	1	2
48	10	2.700	1.900	66	1	3
49	2	7.200	6.630	98	2	1
49	7	6.710	6.880	100	2	2
50	1	3.250	1.650	44	1	1
50	9	2.450	1.650	45	1	2
51	2	4.556	0.467	36	2	1
51	9	3.667	0.422	37	2	2

END

#Data - disability

s2[]	t2[]	y2[]	sd2[]	n2[]	b2[]	m2[]
1	1	6.360	0.790	17	1	1
1	7	1.280	0.970	17	1	2
2	1	5.960	2.320	50	1	1
2	7	3.740	2.690	52	1	2

3	2	2.250	1.225	149	2	1
3	7	1.800	1.200	144	2	2
3	9	1.850	1.225	146	2	3
4	2	6.180	2.750	50	2	1
4	7	6.070	2.950	52	2	2
5	1	3.757	2.237	124	1	1
5	8	3.110	1.833	125	1	2
6	2	3.920	2.936	23	2	1
6	9	3.660	2.869	23	2	2
7	1	3.045	1.720	193	1	1
7	9	2.695	1.805	199	1	2
8	3	5.100	1.700	10	3	1
8	4	5.500	3.000	10	3	2
8	5	5.100	2.300	10	3	3
9	3	4.630	3.540	33	3	1
9	4	4.930	3.630	38	3	2
9	5	4.400	3.440	32	3	3
10	1	1.850	1.374	94	1	1
10	7	1.958	1.539	88	1	2
11	2	5.980	1.940	23	2	1
11	8	6.600	2.150	22	2	2
12	1	2.990	1.630	19	1	1
12	10	2.806	2.327	22	1	2
13	3	4.590	1.681	19	3	1
13	6	3.960	2.425	39	3	2
14	1	3.660	1.636	70	1	1
14	9	2.881	1.633	72	1	2
15	1	3.860	1.351	27	1	1
15	8	3.700	1.351	27	1	2
15	9	3.610	1.350	25	1	3
16	1	0.357	0.280	50	1	1
16	7	0.428	0.368	45	1	2
17	3	4.010	1.583	33	3	1
17	11	3.469	1.615	34	3	2
18	3	3.420	1.654	20	3	1
18	6	2.933	2.034	20	3	2
19	2	1.930	1.880	68	2	1
19	7	1.850	2.260	69	2	2
20	9	1.800	1.300	24	9	1
20	10	1.200	1.000	24	9	2
21	1	4.490	1.750	15	1	1
21	10	3.560	2.090	29	1	2
22	3	4.618	1.331	30	3	1
22	11	4.525	1.717	30	3	2
23	3	4.850	1.750	23	3	1
23	4	3.850	2.030	32	3	2

```
23      5       3.680  1.650  31      3       3
24      1       2.500  0.850  10      1       1
24      9       2.300  0.840  10      1       2
25      3       3.893  2.723  41      3       1
25      7       1.927  1.826  42      3       2
26      2       6.559  0.899  78      2       1
26      7       6.286  1.038  80      2       2
END


# Inits1

list(mu1 = c(0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0,
                0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0),
     d1 = c(NA,0,0,0,0, 0,0,0,0,0, 0), sig1 = 1,
     mu2 = c(0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0),
     d2 = c(NA,0,0,0,0, 0,0,0,0,0, 0), sig2=1)

# Inits2

list(mu1 = c(1,1,1,1,1, 1,1,1,1,1, 1,1,1,1,1, 1,1,1,1,1, 1,1,1,1,1, 1,1,1,1,1,
                1,1,1,1,1, 1,1,1,1,1, 1,1,1,1,1, 1,1,1,1,1, 1),
     d1 = c(NA,1,1,1,1, 1,1,1,1,1, 1), sig1 = 0.5,
     mu2 = c(1,1,1,1,1, 1,1,1,1,1, 1,1,1,1,1, 1,1,1,1,1, 1,1,1,1,1, 1),
     d2 = c(NA,1,1,1,1, 1,1,1,1,1, 1), sig2=0.5)


#BUGS code for CBREhom

model {

  for (j in 1:NS) {
    for (h in 1:NO) {
      delta[j, 1, h] <- 0
    }
  }

  for (i in 1:N) {
    mean[i] <- mu[s[i],l[i]] + delta[s[i], t[i], l[i]]
    y[i] ~ dnorm(mean[i], prec[i])
    se[i] <- sd[i]/sqrt(n[i])
    prec[i] <- 1/pow(se[i],2)
    fitted[i] <- mean[i]
    res[i] <- y[i] - mean[i]
    dev[i] <- (y[i] - mean[i])*(y[i] - mean[i])*prec[i]
  }
```

```
for (j in 1:NS) {
  for (k in 2:NT) {
    delta[j, k, 1:NO] ~ dmnorm(d[k-1, 1:NO], invR[1:NO, 1:NO])
  }
}

for (j in 1:NS) {
  for (h in 1:NO) {
    mu[j,h] ~ dnorm(mP[h], tauP[h])
  }
}

for (h in 1:NO) {
  mP[h] ~ dnorm(0, 0.0001)
  tauP[h] <- 1/pow(sdP[h], 2)
  sdP[h] ~ dunif(0.01, 10)
}

for (k in 1:(NT-1)) { d[k, 1:NO] ~ dmnorm(md[1:NO], invcovd[1:NO, 1:NO]) }
invR[1:NO, 1:NO] ~ dwish(Omega[1:NO, 1:NO], 2)
R[1:NO, 1:NO] <- inverse(invR[ , ])

sig1 <- sqrt(R[1,1])
sig2 <- sqrt(R[2,2])
rho <- R[1,2]/sqrt(R[1,1]*R[2,2])

sumdev <- sum(dev[])

# ranking
dd1[1] <- 0
dd2[1] <- 0
for (k in 2:NT) {
  dd1[k] <- d[k-1,1]
  dd2[k] <- d[k-1, 2]
}

for (k in 1:NT) {
  T.1[k] <- mP[1] + dd1[k]
  rk.1[k] <- rank(T.1[], k)
  best1.1[k] <- equals(rk.1[k], 1)
  best2.1[k] <- equals(rk.1[k], 2)
  best12.1[k] <- best1.1[k] + best2.1[k]

  T.2[k] <- mP[2] + dd2[k]
  rk.2[k] <- rank(T.2[], k)
  best1.2[k] <- equals(rk.2[k], 1)
```

```
    best2.2[k] <- equals(rk.2[k], 2)
    best12.2[k] <- best1.2[k] + best2.2[k]

    T.eq[k] <- 0.5*T.1[k] + 0.5*T.2[k]
    rk.eq[k] <- rank(T.eq[], k)
    best1.eq[k] <- equals(rk.eq[k], 1)
    best2.eq[k] <- equals(rk.eq[k], 2)
    best12.eq[k] <- best1.eq[k] + best2.eq[k]

    T.pain[k] <- 0.8*T.1[k] + 0.2*T.2[k]
    rk.pain[k] <- rank(T.pain[], k)
    best1.pain[k] <- equals(rk.pain[k], 1)
    best2.pain[k] <- equals(rk.pain[k], 2)
    best12.pain[k] <- best1.pain[k] + best2.pain[k]

    T.dis[k] <- 0.2*T.1[k] + 0.8*T.2[k]
    rk.dis[k] <- rank(T.dis[], k)
    best1.dis[k] <- equals(rk.dis[k], 1)
    best2.dis[k] <- equals(rk.dis[k], 2)
    best12.dis[k] <- best1.dis[k] + best2.dis[k]
 }
}


#Data

list(N=168, NT=11, NS=54, NO=2,
md=c(0,0),
invcovd = structure(.Data = c(0.0001,0,
                                        0,0.0001), .Dim = c(2, 2)),
Omega = structure(.Data = c(10,0,0,10), .Dim = c(2, 2))
)
```

s[]	t[]	y[]	sd[]	n[]	l[]
1	1	7.700	2.300	17	1
1	7	0.700	1.000	17	1
1	1	6.360	0.790	17	2
1	7	1.280	0.970	17	2
2	1	4.770	2.120	50	1
2	7	3.770	1.730	52	1
2	1	5.960	2.320	50	2
2	7	3.740	2.690	52	2
3	2	2.800	1.220	149	1
3	7	2.280	1.200	144	1
3	9	2.420	1.440	146	1
3	2	2.250	1.225	149	2
3	7	1.800	1.200	144	2

3	9	1.850	1.225	146	2
4	2	5.500	2.070	50	1
4	7	4.590	2.400	52	1
4	2	6.180	2.750	50	2
4	7	6.070	2.950	52	2
5	1	4.873	2.063	124	1
5	8	4.607	2.457	125	1
5	1	3.757	2.237	124	2
5	8	3.110	1.833	125	2
6	2	3.780	2.345	23	1
6	9	2.560	1.962	23	1
6	2	3.920	2.936	23	2
6	9	3.660	2.869	23	2
7	1	5.070	2.530	193	1
7	9	4.060	2.530	199	1
7	1	3.045	1.720	193	2
7	9	2.695	1.805	199	2
8	3	6.300	1.900	10	1
8	4	5.000	3.200	10	1
8	5	5.500	2.700	10	1
8	3	5.100	1.700	10	2
8	4	5.500	3.000	10	2
8	5	5.100	2.300	10	2
9	3	4.440	3.510	33	1
9	4	4.730	3.480	38	1
9	5	4.030	3.300	32	1
9	3	4.630	3.540	33	2
9	4	4.930	3.630	38	2
9	5	4.400	3.440	32	2
10	1	4.250	2.367	94	1
10	7	3.858	2.201	88	1
10	1	1.850	1.374	94	2
10	7	1.958	1.539	88	2
11	2	3.750	2.500	23	1
11	8	3.650	0.350	22	1
11	2	5.980	1.940	23	2
11	8	6.600	2.150	22	2
12	1	3.370	1.780	19	1
12	10	2.410	2.050	22	1
12	1	2.990	1.630	19	2
12	10	2.806	2.327	22	2
13	3	4.180	1.659	19	1
13	6	3.740	2.360	39	1
13	3	4.590	1.681	19	2
13	6	3.960	2.425	39	2
14	1	2.959	2.394	70	1

14	9	2.255	2.068	72	1
14	1	3.660	1.636	70	2
14	9	2.881	1.633	72	2
15	1	2.380	1.403	27	1
15	8	1.810	1.403	27	1
15	9	1.560	1.400	25	1
15	1	3.860	1.351	27	2
15	8	3.700	1.351	27	2
15	9	3.610	1.350	25	2
16	1	3.459	2.355	50	1
16	7	3.523	2.193	45	1
16	1	0.357	0.280	50	2
16	7	0.428	0.368	45	2
17	3	4.000	2.600	33	1
17	11	3.900	2.000	34	1
17	3	4.010	1.583	33	2
17	11	3.469	1.615	34	2
18	3	3.178	1.784	20	1
18	6	2.834	2.136	20	1
18	3	3.420	1.654	20	2
18	6	2.933	2.034	20	2
19	2	3.940	2.220	68	1
19	7	3.090	1.540	69	1
19	2	1.930	1.880	68	2
19	7	1.850	2.260	69	2
20	9	1.800	1.600	24	1
20	10	1.300	1.200	24	1
20	9	1.800	1.300	24	2
20	10	1.200	1.000	24	2
21	1	1.686	1.057	15	1
21	10	1.314	1.143	29	1
21	1	4.490	1.750	15	2
21	10	3.560	2.090	29	2
22	3	6.670	1.780	30	1
22	11	5.250	1.900	30	1
22	3	4.618	1.331	30	2
22	11	4.525	1.717	30	2
23	3	6.900	2.000	23	1
23	4	3.800	2.200	32	1
23	5	4.600	2.500	31	1
23	3	4.850	1.750	23	2
23	4	3.850	2.030	32	2
23	5	3.680	1.650	31	2
24	2	2.560	1.800	36	1
24	7	2.300	1.954	33	1
25	3	3.500	2.900	20	1

25	6	2.200	2.800	20	1
26	3	6.750	2.375	10	1
26	6	5.250	1.425	10	1
27	1	3.300	1.500	34	1
27	9	3.000	1.500	34	1
28	2	2.080	2.090	17	1
28	7	2.190	1.670	17	1
29	7	3.800	1.600	23	1
29	8	2.400	1.600	23	1
30	1	4.000	0.743	6	1
30	9	1.471	0.643	9	1
31	1	5.385	1.528	35	1
31	9	5.190	1.657	32	1
32	2	1.397	1.242	21	1
32	7	1.660	1.463	19	1
33	2	1.426	1.026	18	1
33	6	2.094	1.712	20	1
34	1	3.095	2.030	78	1
34	2	2.550	1.945	82	1
34	7	3.110	2.010	80	1
35	1	4.030	2.080	18	1
35	2	4.000	1.560	18	1
35	7	3.190	1.850	16	1
36	2	5.990	2.400	18	1
36	6	5.180	2.110	20	1
37	3	4.100	2.600	10	1
37	6	0.700	0.700	12	1
38	6	0.600	0.100	25	1
38	9	1.040	0.270	25	1
39	1	4.180	1.950	108	1
39	7	3.755	2.400	109	1
40	7	3.840	2.750	32	1
40	8	3.700	1.810	32	1
41	1	3.550	1.700	34	1
41	9	2.400	1.750	34	1
42	1	4.930	2.020	8	1
42	6	5.350	0.970	8	1
43	1	2.764	2.252	14	1
43	7	1.422	2.202	14	1
44	1	1.320	0.760	30	1
44	10	1.280	0.740	30	1
45	1	3.360	1.540	28	1
45	9	2.280	1.690	27	1
46	1	2.096	1.844	12	1
46	6	1.165	1.671	11	1
47	1	3.650	1.700	36	1

```
47      9       2.100  1.500  36      1
47      10      2.150  1.150  36      1
48      1       4.400  1.400  66      1
48      9       3.600  0.700  66      1
48      10      2.700  1.900  66      1
49      2       7.200  6.630  98      1
49      7       6.710  6.880  100     1
50      1       3.250  1.650  44      1
50      9       2.450  1.650  45      1
51      2       4.556  0.467  36      1
51      9       3.667  0.422  37      1
52      1       2.500  0.850  10      2
52      9       2.300  0.840  10      2
53      3       3.893  2.723  41      2
53      7       1.927  1.826  42      2
54      2       6.559  0.899  78      2
54      7       6.286  1.038  80      2
END


#Inits1

list(
d=structure(.Data = c(0,0,0,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0,0), .Dim=c(10,2)),
mP = c(0,0), sdP = c(1,1),
invR=structure(.Data=c(1,0,0,1), .Dim=c(2,2)),
mu=structure(.Data=c(0,0,0,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0,0,
                     0,0,0,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0,0,
                     0,0,0,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0,0,
                     0,0,0,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0,0,
                     0,0,0,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0,0,
                     0,0,0,0,0,0,0,0), .Dim=c(54,2))
)

#Inits2

list(
d=structure(.Data = c(1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1), .Dim=c(10,2)),
mP = c(1,1), sdP = c(0.5,0.5),
invR=structure(.Data=c(2,0,0,2), .Dim=c(2,2)),
mu=structure(.Data=c(1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,
                     1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,
                     1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,
                     1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,
                     1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,
                     1,1,1,1,1,1,1,1), .Dim=c(54,2))
)
```

```
#BUGS code for ABREhom

model {

  for (i in 1:N) {
    y[i] ~ dnorm(mean[s[i],t[i],l[i]], prec[i])
    prec[i] <- n[i]/pow(sd[i], 2)
    mean[s[i],t[i],l[i]] <- mu[t[i], l[i]] + v[s[i], t[i], l[i]]
    dev[i] <- (y[i]-mean[s[i],t[i],l[i]])*(y[i]-mean[s[i],t[i],l[i]])*prec[i]
    res[i] <- y[i] - mean[s[i],t[i],l[i]]
    fitted[i] <- mean[s[i],t[i],l[i]]
  }
  resdev <- sum(dev[])

  for (j in 1:NS) {
    for (k in 1:NT) {
      v[j, k, 1:NO] ~ dmnorm(meano[1:NO], invRo[1:NO,1:NO])
     }
  }


  invRo[1:NO,1:NO] ~ dwish(Omegao[1:NO, 1:NO], NO)
  Ro[1:NO, 1:NO] <- inverse(invRo[ , ])

  corro <- Ro[1,2]/sqrt(Ro[1,1]*Ro[2,2])
  for (h in 1:NO) { sig[h] <- sqrt(Ro[h,h]) }

  for (k in 1:NT) {
    for (h in 1:NO) {
      mu[k,h] ~ dnorm(0, 0.0001)
    }
  }

  # To compare with CB models
  for (h in 1:NO) {
    for (k in 1:NT) {
      d[k,h] <- mu[k,h] - mu[1,h]
    }
  }

  #rank
  for (k in 1:NT) {
    T.1[k] <- mu[k,1]
    rk.1[k] <- rank(T.1[], k)
    best1.1[k] <- equals(rk.1[k],1)
```

```
    best2.1[k] <- equals(rk.1[k],2)
    best12.1[k] <- best1.1[k] + best2.1[k]

    T.2[k] <- mu[k,2]
    rk.2[k] <- rank(T.2[], k)
    best1.2[k] <- equals(rk.2[k],1)
    best2.2[k] <- equals(rk.2[k],2)
    best12.2[k] <- best1.2[k] + best2.2[k]

    T.eq[k] <- 0.5*T.1[k] + 0.5*T.2[k]
    rk.eq[k] <- rank(T.eq[], k)
    best1.eq[k] <- equals(rk.eq[k],1)
    best2.eq[k] <- equals(rk.eq[k], 2)
    best12.eq[k] <- best1.eq[k] + best2.eq[k]

    T.pain[k] <- 0.8*T.1[k] + 0.2*T.2[k]
    rk.pain[k] <- rank(T.pain[], k)
    best1.pain[k] <- equals(rk.pain[k], 1)
    best2.pain[k] <- equals(rk.pain[k], 2)
    best12.pain[k] <- best1.pain[k] + best2.pain[k]

    T.dis[k] <- 0.2*T.1[k] + 0.8*T.2[k]
    rk.dis[k] <- rank(T.dis[], k)
    best1.dis[k] <- equals(rk.dis[k], 1)
    best2.dis[k] <- equals(rk.dis[k], 2)
    best12.dis[k] <- best1.dis[k] + best2.dis[k]
  }
}


#Data

list(
N=168, NT=11, NS=54, NO=2,
meano= c(0,0),
Omegao = structure(.Data = c(10,0,0,10), .Dim = c(2,2))
)
```

s[]	t[]	y[]	sd[]	n[]	l[]
1	1	7.700	2.300	17	1
1	7	0.700	1.000	17	1
1	1	6.360	0.790	17	2
1	7	1.280	0.970	17	2
2	1	4.770	2.120	50	1
2	7	3.770	1.730	52	1
2	1	5.960	2.320	50	2
2	7	3.740	2.690	52	2

3	2	2.800	1.220	149	1
3	7	2.280	1.200	144	1
3	9	2.420	1.440	146	1
3	2	2.250	1.225	149	2
3	7	1.800	1.200	144	2
3	9	1.850	1.225	146	2
4	2	5.500	2.070	50	1
4	7	4.590	2.400	52	1
4	2	6.180	2.750	50	2
4	7	6.070	2.950	52	2
5	1	4.873	2.063	124	1
5	8	4.607	2.457	125	1
5	1	3.757	2.237	124	2
5	8	3.110	1.833	125	2
6	2	3.780	2.345	23	1
6	9	2.560	1.962	23	1
6	2	3.920	2.936	23	2
6	9	3.660	2.869	23	2
7	1	5.070	2.530	193	1
7	9	4.060	2.530	199	1
7	1	3.045	1.720	193	2
7	9	2.695	1.805	199	2
8	3	6.300	1.900	10	1
8	4	5.000	3.200	10	1
8	5	5.500	2.700	10	1
8	3	5.100	1.700	10	2
8	4	5.500	3.000	10	2
8	5	5.100	2.300	10	2
9	3	4.440	3.510	33	1
9	4	4.730	3.480	38	1
9	5	4.030	3.300	32	1
9	3	4.630	3.540	33	2
9	4	4.930	3.630	38	2
9	5	4.400	3.440	32	2
10	1	4.250	2.367	94	1
10	7	3.858	2.201	88	1
10	1	1.850	1.374	94	2
10	7	1.958	1.539	88	2
11	2	3.750	2.500	23	1
11	8	3.650	0.350	22	1
11	2	5.980	1.940	23	2
11	8	6.600	2.150	22	2
12	1	3.370	1.780	19	1
12	10	2.410	2.050	22	1
12	1	2.990	1.630	19	2
12	10	2.806	2.327	22	2

13	3	4.180	1.659	19	1
13	6	3.740	2.360	39	1
13	3	4.590	1.681	19	2
13	6	3.960	2.425	39	2
14	1	2.959	2.394	70	1
14	9	2.255	2.068	72	1
14	1	3.660	1.636	70	2
14	9	2.881	1.633	72	2
15	1	2.380	1.403	27	1
15	8	1.810	1.403	27	1
15	9	1.560	1.400	25	1
15	1	3.860	1.351	27	2
15	8	3.700	1.351	27	2
15	9	3.610	1.350	25	2
16	1	3.459	2.355	50	1
16	7	3.523	2.193	45	1
16	1	0.357	0.280	50	2
16	7	0.428	0.368	45	2
17	3	4.000	2.600	33	1
17	11	3.900	2.000	34	1
17	3	4.010	1.583	33	2
17	11	3.469	1.615	34	2
18	3	3.178	1.784	20	1
18	6	2.834	2.136	20	1
18	3	3.420	1.654	20	2
18	6	2.933	2.034	20	2
19	2	3.940	2.220	68	1
19	7	3.090	1.540	69	1
19	2	1.930	1.880	68	2
19	7	1.850	2.260	69	2
20	9	1.800	1.600	24	1
20	10	1.300	1.200	24	1
20	9	1.800	1.300	24	2
20	10	1.200	1.000	24	2
21	1	1.686	1.057	15	1
21	10	1.314	1.143	29	1
21	1	4.490	1.750	15	2
21	10	3.560	2.090	29	2
22	3	6.670	1.780	30	1
22	11	5.250	1.900	30	1
22	3	4.618	1.331	30	2
22	11	4.525	1.717	30	2
23	3	6.900	2.000	23	1
23	4	3.800	2.200	32	1
23	5	4.600	2.500	31	1
23	3	4.850	1.750	23	2

23	4	3.850	2.030	32	2
23	5	3.680	1.650	31	2
24	2	2.560	1.800	36	1
24	7	2.300	1.954	33	1
25	3	3.500	2.900	20	1
25	6	2.200	2.800	20	1
26	3	6.750	2.375	10	1
26	6	5.250	1.425	10	1
27	1	3.300	1.500	34	1
27	9	3.000	1.500	34	1
28	2	2.080	2.090	17	1
28	7	2.190	1.670	17	1
29	7	3.800	1.600	23	1
29	8	2.400	1.600	23	1
30	1	4.000	0.743	6	1
30	9	1.471	0.643	9	1
31	1	5.385	1.528	35	1
31	9	5.190	1.657	32	1
32	2	1.397	1.242	21	1
32	7	1.660	1.463	19	1
33	2	1.426	1.026	18	1
33	6	2.094	1.712	20	1
34	1	3.095	2.030	78	1
34	2	2.550	1.945	82	1
34	7	3.110	2.010	80	1
35	1	4.030	2.080	18	1
35	2	4.000	1.560	18	1
35	7	3.190	1.850	16	1
36	2	5.990	2.400	18	1
36	6	5.180	2.110	20	1
37	3	4.100	2.600	10	1
37	6	0.700	0.700	12	1
38	6	0.600	0.100	25	1
38	9	1.040	0.270	25	1
39	1	4.180	1.950	108	1
39	7	3.755	2.400	109	1
40	7	3.840	2.750	32	1
40	8	3.700	1.810	32	1
41	1	3.550	1.700	34	1
41	9	2.400	1.750	34	1
42	1	4.930	2.020	8	1
42	6	5.350	0.970	8	1
43	1	2.764	2.252	14	1
43	7	1.422	2.202	14	1
44	1	1.320	0.760	30	1
44	10	1.280	0.740	30	1

```
45      1       3.360  1.540  28      1
45      9       2.280  1.690  27      1
46      1       2.096  1.844  12      1
46      6       1.165  1.671  11      1
47      1       3.650  1.700  36      1
47      9       2.100  1.500  36      1
47      10      2.150  1.150  36      1
48      1       4.400  1.400  66      1
48      9       3.600  0.700  66      1
48      10      2.700  1.900  66      1
49      2       7.200  6.630  98      1
49      7       6.710  6.880  100     1
50      1       3.250  1.650  44      1
50      9       2.450  1.650  45      1
51      2       4.556  0.467  36      1
51      9       3.667  0.422  37      1
52      1       2.500  0.850  10      2
52      9       2.300  0.840  10      2
53      3       3.893  2.723  41      2
53      7       1.927  1.826  42      2
54      2       6.559  0.899  78      2
54      7       6.286  1.038  80      2
END

#Inits1

list(
mu = structure(.Data = c(0,0,0,0,0, 0,0,0,0,0, 0, 0,0,0,0,0, 0,0,0,0,0, 0), .Dim=c(11,2)),
invRo = structure(.Data=c(1,0,0,1), .Dim=c(2,2))
)

#Inits2

list(
mu = structure(.Data = c(1,1,1,1,1, 1,1,1,1,1, 1, 1,1,1,1,1, 1,1,1,1,1, 1), .Dim=c(11,2)),
invRo = structure(.Data=c(2,0,0,2), .Dim=c(2,2))
)
```