Published in final edited form as:

Biometrics. 2019 September ; 75(3): 978–987. doi:10.1111/biom.13028.

A Bayesian Hierarchical Model Estimating CACE in Meta-analysis of Randomized Clinical Trials with Noncompliance

Jincheng Zhou^{1,*}, James S. Hodges¹, M. Fareed K. Suri², Haitao Chu^{1,**}

¹Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455, U.S.A.

²Department of Neurology, University of Minnesota, Minneapolis, MN 55455, U.S.A.

Summary:

Noncompliance to assigned treatment is a common challenge in analysis and interpretation of randomized clinical trials. The complier average causal effect (CACE) approach provides a useful tool for addressing noncompliance, where CACE is defined as the average difference in potential outcomes for the response in the subpopulation of subjects who comply with their assigned treatments. In this article, we present a Bayesian hierarchical model to estimate the CACE in a meta-analysis of randomized clinical trials where compliance may be heterogeneous between studies. Between-study heterogeneity is taken into account with study-specific random effects. The results are illustrated by a re-analysis of a meta-analysis comparing the effect of epidural analgesia in labor versus no or other analgesia in labor on the outcome cesarean section, where noncompliance varied between studies. Finally, we present simulations evaluating the performance of the proposed approach and illustrate the importance of including appropriate random effects and the impact of over- and under-fitting.

Keywords

Bayesian hierarchical model; CACE; causal effect; meta-analysis; noncompliance; randomized trial

1. Introduction

A randomized clinical trial (RCT) is the gold standard method for investigating the effect of medical interventions (Holland, 1986). Yet in human randomized experiments, noncompliance with assigned treatments is common and may induce an information bias in estimating the average treatment effect (Frangakis and Rubin, 1999). One potential solution to this problem is intention-to-treat (ITT) analysis. It ignores noncompliance, withdrawal, and all post-randomization variables, and maintains prognostic balance based on the original

* jzhou@umn.edu. ** chux0051@umn.edu.

Supplementary Materials

Data analyzed in Section 3, parameter estimates from the fixed effects model (Model I) and the final model (Model IVa), sample R JAGS code implementing Model IVa for the case study, sensitivity analysis on prior distributions, and additional simulation scenarios are available with this paper at the Biometrics website on Wiley Online Library.

random treatment allocation, but the estimated treatment effect is generally conservative (Freedman, 1990). Alternative analytic approaches are possible, including per-protocol analysis, which excludes noncompliant individuals, or as-treated analysis, which groups individuals by the treatment actually received. However, these approaches can break randomization and have a high potential for bias in either direction (Little and Yau, 1998).

Important statistical methods have been developed for analyzing RCT outcomes in the presence of noncompliance. Two frequently used estimands of treatment efficacy have been considered, the average treatment effect (ATE) (Little et al., 2009), and the complier average causal effect (CACE) (Yau and Little, 2001). The ATE is the average efficacy across all patients and bounds of ATE have been derived for binary outcomes (Balke and Pearl, 1997). CACE analysis builds on the principal stratification framework (Frangakis and Rubin, 2002). The CACE is defined as the average difference in potential outcomes for the response in a subpopulation of subjects who would comply with their assigned treatment. The main challenge in CACE modeling is illustrated by considering treatment outcomes in the four latent classes defined by the intersection of treatment assignment and treatment compliance: *compliers* (for whom treatment received is the same as assigned), *never-takers* (who do not receive treatment, regardless of assignment), *always-takers* (who receive treatment regardless of assignment), and *defiers* (who always do the opposite of their treatment assignment).

The core insight of CACE analysis is that it retains the initial randomized assignment, so it overcomes the intrinsic bias associated with per-protocol or as-treated analysis. Under certain assumptions (see Section 2.1 for further discussion), we can derive an unbiased estimate of the difference in outcomes between compliers in the intervention group and those in the control group who would have engaged with treatment. Under some assumptions, CACE has been developed and applied in various settings to account for noncompliance. For example, Imbens and Rubin (1997) estimated CACE with the maximum-likelihood (ML) method using the EM algorithm and a Bayesian approach using the data augmentation algorithm. Little and Yau (1998) incorporated covariates in this framework via the ML-EM method. Cheng (2009) estimated CACE when the outcome is ordinal using the ML method. Methods to estimate causal effects under interference have also been developed recently (Hudgens and Halloran, 2008; Liu and Hudgens, 2014).

However, most of those approaches only consider the setting of a single randomized clinical trial. To our knowledge, little attention has been paid to estimating causal effects from metaanalysis of multiple clinical trials accounting for noncompliance. Glasziou (1992) proposed a meta-analysis for “explanatory treatment effect” adjusting for compliance, based on the linear model formulation of Newcombe (1988). Later, Baker and Kramer (2005) proposed ML estimates of “efficacy” when there is switching of interventions (which is similar to CACE) and briefly discussed its implications for meta-analysis using hypothetical sets of trials. Bannister-Tyrrell et al. (2015) presented fixed-effect meta-analysis adjusting for compliance in trials using an instrumental variable (IV) estimator. However, existing meta-regression methods cannot be used directly to derive CACE adjusting for noncompliance for the following reasons: 1) existing meta-regression can only be used to estimate the impact of fixed study-level covariates on study-specific relative effect size; 2)

noncompliance is observed post-randomization and is considered to be a random variable; 3) noncompliance rates generally differ between treatment arms within a study and may affect the response to treatment. We address this gap by extending the CACE approach for ordinal outcomes from a single clinical trial setting to meta-analysis settings with multiple clinical trials. Compared with ITT analysis or a naive meta-regression approach using the averages of post-randomization variables (*e.g.*, each study's average noncompliance rate) as fixed covariates, this method allows us to appropriately account for the inherent heterogeneity in noncompliance rates between treatment arms and between studies.

This article is organized as follows. Section 2 presents our Bayesian hierarchical modeling approach and explains the model assumptions. Section 3 gives a case study reanalyzing a meta-analysis comparing epidural analgesia versus no or other analgesia in labor on the outcome of cesarean section, where noncompliance varies between studies. Section 4 presents a comprehensive simulation study evaluating the performance of our approach under a variety of conditions. Finally, Section 5 discusses our findings and implications for future developments.

2. Statistical Methods

2.1 Notation and assumptions

We consider I two-armed randomized trials. For the i^{th} trial, let N_i be the number of subjects, where N_{i0} are randomly assigned to the control group and N_{i1} are assigned to active treatment. Let $\mathbf{R}_i$ be the N_i -dimensional vector of randomization assignments for all subjects, with individual element $R_{ij} = r$ corresponding to whether subject j is assigned to treatment ($r = 1$) or control ($r = 0$). Let $\mathbf{T}_i^r$ be the N_i -dimensional vector of *potential* treatments received under the randomization assignments $\mathbf{r}$ with individual element T_{ij}^r , where $T_{ij}^r = t \in \{0, 1\}$ according to whether subject j actually took the active treatment ($t = 1$) or placebo ($t = 0$) under $\mathbf{r}$. Let $\mathbf{Y}_i^{r,t}$ be the vector of potential outcomes under randomization assignment $\mathbf{r}$ and treatment received t , with individual element $Y_{ij}^{r,t}$ for the j^{th} subject in the i^{th} trial. In this study, we consider only ordinal outcomes $Y_{ij}^{r,t} = o \in \{1, \dots, O\}$. Note that the sets of $\{Y_{ij}^{r,t}\}$ and $\{T_{ij}^r\}$ are *potential* outcomes and treatment-received status, where for each (i, j) , only one of the possible values of each set can be observed, so we denote the observed response and received treatment as Y_{ij} and T_{ij} for the j^{th} subject in the i^{th} trial.

Following Imbens and Rubin (1997), we let C_{ij} be the latent compliance class of the j^{th} patient in the i^{th} trial, defined as follows:

$$C_{ij} = \begin{cases} 0, & \text{never-taker, if } (T_{ij}^0, T_{ij}^1) = (0, 0) \\ 1, & \text{complier, if } (T_{ij}^0, T_{ij}^1) = (0, 1) \\ 2, & \text{always-taker, if } (T_{ij}^0, T_{ij}^1) = (1, 1) \\ 3, & \text{defier, if } (T_{ij}^0, T_{ij}^1) = (1, 0). \end{cases}$$

In a two-arm trial, we can only observe one of T_{ij}^1 and T_{ij}^0 , so a subject's compliance status C_{ij} is not an observable variable. It can only be partially identified based on treatment assignment and observed treatment received (see Table 1 columns R_{ij} , T_{ij} and C_{ij}).

To identify treatment effects of interest, for each study we make assumptions identical to those listed in Angrist et al. (1996) and describe their implications below.

Assumption 1: Stable unit treatment value assumption (SUTVA) (Rubin, 1980). The outcome for a subject is unaffected by the particular assignments of treatments to the other individuals. That is, if $r = r'$ then $T_{ij}^r = T_{ij}^{r'}$; and if $r = r'$ and $t = t'$ then $Y_{ij}^{r,t} = Y_{ij}^{r',t'}$.

Assumption 2: Random assignment to randomization groups. For all N_i subjects in the i^{th} trial, the treatment assignment $\mathbf{R}_i$ is random. This assumption implies that the proportion of each compliance class should be the same in the intervention and control groups.

Assumption 3: Exclusion restriction. For subject j in the i^{th} trial, $Y_{ij}^{r,t} = Y_{ij}^{r',t}$ for all r, r' and t , i.e., the randomization assignment affects responses only through its effect on treatment received. With this assumption, *Assumption 1* can be restated: if $t = t'$ then $Y_{ij}^{r,t} = Y_{ij}^{r',t'}$ for any r, r' . Therefore, for always-takers and never-takers, the distribution of outcomes does not depend on randomization group.

Assumption 4: Nonzero average causal effect of $\mathbf{R}_i$ on T_i . For each trial, $E[T_{ij}^1 - T_{ij}^0] \neq 0$. The fraction of subjects who receive each intervention varies by randomization group.

Assumption 5: Monotonicity. $P[T_{ij}^1 \geq T_{ij}^0] = 1$ for each trial. This means that no subject necessarily receives the treatment opposite to the randomized assignment, under assignment to either active treatment and control. This assumption rules out the existence of defiers, so it reduces the number of compliance types and permits a properly identified model (see Table 1).

2.2 Estimation and inference for the causal effect

For discrete outcomes, we extend the notation in Cheng (2009) and Baker (2011), and define the following parameters under the latent compliance model: 1) π_{ia} , the probability of being an always-taker in the i^{th} study; 2) π_{in} , the probability of being a never-taker in the i^{th} study, so the probability of being a complier in the i^{th} study, π_{ic} , is $1 - \pi_{ia} - \pi_{in}$; 3) u_{io} , the probability a complier randomized to the treatment group has outcome o in the i^{th} study; 4) v_{io} , the probability a complier randomized to the control group has outcome o in the i^{th} study; 5) s_{io} , the probability a never-taker has outcome o in the i^{th} study; and 6) b_{io} , the probability an always-taker has outcome o in the i^{th} study, where

$\sum_{o=1}^O u_{io} = \sum_{o=1}^O v_{io} = \sum_{o=1}^O s_{io} = \sum_{o=1}^O b_{io} = 1$. The study-specific probabilities π_{ia} , π_{in} , u_{io} , v_{io} , s_{io} , and b_{io} may be assumed to be draws of their respective random effects, which follow some common distributions, discussed below.

Table 1 shows the distributions of observed n_{ijt} in terms of the parameters for trial i , where $n_{ijt} = \sum_j I(R_{ij} = r, T_{ij} = t)$ denotes the number of individuals in each observed group and $M(n_{ijt}, \mathbf{x}_{io})$ denotes a multinomial distribution with n_{ijt} subjects and multinomial probabilities $\{x_{io}\}$. The observed count for each outcome o in group $\{j: R_{ij} = r, T_{ij} = t\}$ is n_{irtor} , $o = 1, \dots, O$. Although latent compliance classes cannot be fully identified based on randomization group (R_{ij}) and observed treatment-received behaviors (T_{ij}), the randomization assignment and the exclusion restriction assumptions imply that 1) the proportions of always-takers, never-takers and compliers in the control group are equal to those in the treatment group; 2) for never-takers and always-takers, the outcome distribution under control is the same as that under active treatment. They further give $q_{io} = \frac{\pi_{ic}v_{io} + \pi_{in}s_{io}}{1 - \pi_{ia}}$ and $p_{io} = \frac{\pi_{ic}u_{io} + \pi_{ia}b_{io}}{1 - \pi_{in}}$ as probabilities corresponding to n_{i0o} and n_{i1o} , $o \in \{1, \dots, O\}$ (shown in Table 1).

One causal effect of interest in many studies is the CACE discussed in Section 1. For an ordinal outcome $Y_{ij} = o \in \{1, \dots, O\}$, suppose we can code scores $\{W_1, W_2, \dots, W_O\}$ to reflect the real distances between categories. Then the CACE of the i^{th} trial is defined as

$$\theta_i^{CACE} = E(Y_{ij}^1 - Y_{ij}^0 | C_{ij} = 1) = \sum_o (W_o \times u_{io}) - \sum_o (W_o \times v_{io}). \quad (1)$$

The overall causal effect from the meta-analysis of I trials θ^{CACE} can be defined as $\theta^{CACE} = E(\theta_i^{CACE})$. It is often unclear how to select scores $\{W_1, W_2, \dots, W_O\}$ to reflect distances between categories. Equally spaced scores $\{1, 2, \dots, O\}$ or their linear transforms are sensible in many cases and provide a reasonable compromise when there is no obvious choice (Agresti, 2003). Alternative scoring systems such as midranks are also possible. When uncertain about which scoring choice to use, a sensitivity analysis can consider different sensible choices to see how they affect the results.

2.3 Likelihood and the Bayesian hierarchical model

For the i^{th} trial, let λ_i denote the probability $P(R_{ij} = 1)$, which is usually known and can therefore be treated as fixed. Let $\boldsymbol{\beta} = (\pi_{ia}, \pi_{in}, s_i, \mathbf{b}_i, \mathbf{u}_i, \mathbf{v}_i)$, where $s_i = (s_{i1}, \dots, s_{i(O-1)})$, $\mathbf{b}_i = (b_{i1}, \dots, b_{i(O-1)})$, $\mathbf{u}_i = (u_{i1}, \dots, u_{i(O-1)})$, $\mathbf{v}_i = (v_{i1}, \dots, v_{i(O-1)})$, then the i^{th} trial's contribution to the likelihood is

$$L_i(\beta_i) = \prod_j \prod_o \{ (1 - \lambda_i)(\pi_{ic}v_{io} + \pi_{in}s_{io}) \}^{(1-R_{ij})(1-T_{ij})I(Y_{ij}=o)} \{ (1 - \lambda_i)\pi_{ia}b_{io} \}^{(1-R_{ij})T_{ij}I(Y_{ij}=o)} \\ \{ \lambda_i\pi_{in}s_{io} \}^{R_{ij}(1-T_{ij})I(Y_{ij}=o)} \{ \lambda_i(\pi_{ic}u_{io} + \pi_{ia}b_{io}) \}^{R_{ij}T_{ij}I(Y_{ij}=o)}, \quad (2)$$

where $j = 1, \dots, N_j$, $o = 1, \dots, O$, $\pi_{ic} = 1 - \pi_{ia} - \pi_{in}$, the indicator function $I(Y_{ij} = o) = 1$ if $Y_{ij} = o$ and 0 otherwise, ${}_o s_{io} = {}_o b_{io} = {}_o u_{io} = {}_o v_{io} = 1$, and $0 \leq \pi_{ia}, \pi_{in}, \pi_{ic}, s_{io}, b_{io}, u_{io}, v_{io} \leq 1$.

The likelihood for all trials in a meta-analysis is $\mathcal{L}(\beta) = \prod_i L_i(\beta_i)$. Using the multivariate random effects meta-analysis framework (Jackson et al., 2011), one can specify $\beta_i \sim F(\beta_0, \Sigma_0)$ where F is some distribution function, β_0 is the overall mean parameter and Σ_0 is the variance-covariance matrix. In a Bayesian framework, one would also specify the prior distributions of β_0 and Σ_0 as $f(\beta_0)$ and $f(\Sigma_0)$, respectively. The joint posterior distribution is then proportional to $\prod_i L_i(\beta_i) F(\beta_i | \beta_0, \Sigma_0) f(\beta_0) f(\Sigma_0)$. The posterior distribution can be estimated by Markov chain Monte Carlo (MCMC) methods, specifically the Gibbs and Metropolis-Hastings sampling algorithms (Gelfand and Smith, 1990).

2.4 Accounting for heterogeneity through random effects models

Between-study heterogeneity commonly exists in a meta-analysis because studies usually differ in their subject recruitment methods, measurement techniques, study qualities, *etc.*. To account for potential between-study heterogeneity of the fractions in the compliance classes and of the outcome probabilities, we consider a random-effects model.

To guarantee the desired properties of the probabilities of being in each principal stratum in study i , *i.e.*, $\pi_{in} + \pi_{ia} + \pi_{ic} = 1$, and $0 \leq \pi_{in}, \pi_{ia}, \pi_{ic} \leq 1$, and to allow these probabilities to vary between studies, we assume $\pi_{in} = \frac{\exp(n_i)}{1 + \exp(n_i) + \exp(a_i)}$, $\pi_{ia} = \frac{\exp(a_i)}{1 + \exp(n_i) + \exp(a_i)}$, where $n_i = \alpha_n + \delta_{in}$, $a_i = \alpha_a + \delta_{ia}$, and $(\delta_{in}, \delta_{ia})^T \sim N(0, \Sigma_{ps})$. To allow correlation between n_i and a_i , the variance-covariance matrix is $\Sigma_{ps} = \begin{pmatrix} \sigma_n^2 & \rho\sigma_n\sigma_a \\ \rho\sigma_n\sigma_a & \sigma_a^2 \end{pmatrix}$.

To be consistent with the case study presented in Section 3, we describe the rest of the model for the binary case, where $o \in \{0, 1\}$, *i.e.*, $s_{i0} + s_{i1} = b_{i0} + b_{i1} = u_{i0} + u_{i1} = v_{i0} + v_{i1} = 1$ for the i^{th} trial, so the vector parameters of s_i, b_i, u_i, v_i become scalars $s_{i1}, b_{i1}, u_{i1}, v_{i1}$. Specifically, we assume a normal distribution for each response probability s_{i1}, b_{i1}, u_{i1} , and v_{i1} on the transformed scales, and allow heterogeneity between studies: $g(s_{i1}) = \alpha_s + \delta_{is}$, $g(b_{i1}) = \alpha_b + \delta_{ib}$, $g(u_{i1}) = \alpha_u + \delta_{iu}$, $g(v_{i1}) = \alpha_v + \delta_{iv}$, where $g(\cdot)$ is a link function such as the

logit or probit. We then assume the probabilities of having an outcome are independent across principal strata, so $\delta_{is} \sim N(0, \sigma_s^2)$, $\delta_{ib} \sim N(0, \sigma_b^2)$, $\delta_{iu} \sim N(0, \sigma_u^2)$, $\delta_{iv} \sim N(0, \sigma_v^2)$. To extend this model to more general cases with outcomes taking more than two values, one could fit a generalized logit model, *e.g.*, the baseline-category logit model for nominal responses or the cumulative logit model for ordinal responses (Agresti, 2003).

2.5 Implementation

We used a Bayesian hierarchical model (Gelman et al., 2013), fit using JAGS. To avoid overfitting the data with an excess of random effects, we used a forward selection procedure on the random effects. Specifically, at each forward step, we added the random-effect component that provided the largest improvement in the deviance information criterion (DIC) (Spiegelhalter et al., 2002). Other model-selection approaches are easily substituted for these choices, *e.g.*, using a different model-selection criterion (Celeux et al., 2006), or a different search strategy.

Computation used Markov chain Monte Carlo (MCMC) in JAGS, interfacing with R using the package “rjags”. Each fit consisted of three independent chains initiated with parameter values drawn randomly from their prior distributions. Each chain was allowed 10,000 iterations of burn-in, and the subsequent 100,000 iterations were collected as posterior samples. Estimates are presented as the posterior mean with a 95% credibility interval. Convergence of Markov chains was assessed using trace plots, sample autocorrelation, and the Gelman and Rubin convergence statistic (Gelman and Rubin, 1992). We also used the R package “mcmcse” (Dai and Jones, 2017) to calculate the effective sample size and to make sure the chains gave adequately precise posterior estimates.

3. Case Study

3.1 Data and descriptive analysis

We reanalyzed the data from a meta-analysis of the impact of epidural analgesia in labor compared with no or other analgesia, on the outcome of cesarean section (Bannister-Tyrrell et al., 2015). Epidural analgesia in labor is a highly effective method of labor pain relief, but there is a controversy over whether epidural analgesia in labor is associated with an increased risk of cesarean delivery. However, noncompliance is common in RCTs in obstetrics: it is unethical to deny laboring women their preferred mode of labor analgesia because trial entry and randomization occurs before the experience of labor pain. Thus the ITT result is not a good estimate of the consequence of receiving epidural analgesia due to noncompliance, and it motivates us to investigate the causal effect. Data were recorded on randomized treatment assignment (1=Epidural analgesia, 0=No/other analgesia in labor), treatment actually received (1=Epidural analgesia, 0=No/other analgesia in labor), and frequency of cesarean section (1=Yes, 0=No) by compliance with allocated intervention. 10 RCTs had complete data available on the number of cesarean sections in compliant and noncompliant participants, which could be meta-analyzed.

In these 10 trials, 1,684 women were randomly assigned to epidural analgesia, with 130 cesarean deliveries, and 1,732 were assigned to no/other analgesia with 115 cesarean

deliveries. Unlike the ITT and IV analysis used by Bannister-Tyrrell et al. (2015), in which the data from Nikkola et al. (1997) were excluded because it had no cesarean section events, we included this trial because our Bayesian method accommodates trials with zero events.

Figure 1 displays the noncompliance rates of these studies, where $P(T = 1|R = 0)$ is the fraction of participants who were randomized to no/other analgesia in labor but actually received epidural analgesia, and $P(T = 0|R = 1)$ is the fraction of participants who did not receive epidural analgesia among those who were assigned to it. Because several arms had 0 events, the confidence intervals were calculated using an exact method, the Clopper-Pearson exact interval (Clopper and Pearson, 1934). We adopted a bivariate generalized linear mixed effects model (BGLMM) (Chu and Cole, 2006) to model the two overall noncompliance rates. The BGLMM assumes a bivariate normal distribution for the logits of the probabilities in the two groups: $\text{logit}(p_{1i}) = \mu_1 + \eta_{1i}$, $\text{logit}(p_{0i}) = \mu_0 + \eta_{0i}$, $(\eta_{1i}, \eta_{0i})^T \sim \text{MVN}(\mathbf{0}, \Sigma_\eta)$, where $p_{0i} = P(T = 1|R = 0)$ and $p_{1i} = P(T = 0|R = 1)$. As shown in Figure 1, the noncompliance rates varied substantially across studies. The index of heterogeneity I^2 (Higgins and Thompson, 2002) was 98.48% for $P(T = 1|R = 0)$ and 98.07% for $P(T = 0|R = 1)$. The high heterogeneity motivated us to apply Bayesian hierarchical models to estimate CACE accounting for between-study heterogeneity.

In this case study, the SUTVA assumption holds because a participant's outcome did not depend on the other participants' treatments. The exclusion restriction assumption holds because a participant's randomized assignment affected the cesarean section outcome only through the analgesia that was actually received. The other three assumptions are also reasonable based on the randomized clinical trial design.

3.2 Likelihood and priors

As discussed in Section 2.4, for a binary outcome ($o \in \{0, 1\}$) the parameter vector is reduced to $\beta_i = (\pi_{ia}, \pi_{in}, s_{i1}, b_{i1}, u_{i1}, v_{i1})$. From Table 1, the probability of each observed n_{irto} is determined by the above parameters. Therefore, the i^{th} trial's contribution to the likelihood is

$$L_i(\beta_i) = [(1 - \lambda_i)\{\pi_{ic}(1 - v_{i1}) + \pi_{in}(1 - s_{i1})\}]^{n_{i000}} \{(1 - \lambda_i)(\pi_{ic}v_{i1} + \pi_{in}s_{i1})\}^{n_{i001}} \quad (3) \\ \{(1 - \lambda_i)\pi_{ia}(1 - b_{i1})\}^{n_{i010}} \{(1 - \lambda_i)\pi_{ia}b_{i1}\}^{n_{i011}} \{\lambda_i\pi_{in}(1 - s_{i1})\}^{n_{i100}} \\ \{\lambda_i\pi_{in}s_{i1}\}^{n_{i101}} [\lambda_i\{\pi_{ic}(1 - u_{i1}) + \pi_{ia}(1 - b_{i1})\}]^{n_{i110}} \{\lambda_i(\pi_{ic}u_{i1} + \pi_{ia}b_{i1})\}^{n_{i111}}$$

and the likelihood for all trials in the meta-analysis is $\mathcal{L}(\beta) = \prod_i L_i(\beta_i)$.

Using equally spaced scores ($W_0 = 0$, $W_1 = 1$), the CACE for study i is $\theta_i^{\text{CACE}} = u_{i1} - v_{i1}$, the response rate difference in compliers. Using the model introduced in Section 2.4, the medians of u_{i1} and v_{i1} can be estimated as $M(u_{i1}) = g^{-1}(\alpha_u)$ and $M(v_{i1}) = g^{-1}(\alpha_v)$ respectively, while the means of u_{i1} and v_{i1} can be estimated as

$$E(u_{i1}) = \int_{-\infty}^{+\infty} g^{-1}(\alpha_u + t)\sigma_u^{-1}\phi\left(\frac{t}{\sigma_u}\right)dt \text{ and } E(v_{i1}) = \int_{-\infty}^{+\infty} g^{-1}(\alpha_v + t)\sigma_v^{-1}\phi\left(\frac{t}{\sigma_v}\right)dt \text{ respectively,}$$

where $\phi(\cdot)$ is the standard Gaussian density. Thus

$$\theta^{\text{CACE}} = E(\theta_i^{\text{CACE}}) = E(u_{i1}) - E(v_{i1}) = \int_{-\infty}^{+\infty} g^{-1}(\alpha_u + t) \sigma_u^{-1} \phi\left(\frac{t}{\sigma_u}\right) dt - \int_{-\infty}^{+\infty} g^{-1}(\alpha_v + t) \sigma_v^{-1} \phi\left(\frac{t}{\sigma_v}\right) dt$$

$E(\theta_i^{\text{CACE}})$ involves an integral, the probit-link random effects model implies closed-form

formulas $E(u_{i1}) = \Phi\left(\frac{\alpha_u}{\sqrt{1 + \sigma_u^2}}\right)$ and $E(v_{i1}) = \Phi\left(\frac{\alpha_v}{\sqrt{1 + \sigma_v^2}}\right)$, and the logit-link random effects model

has a well-established approximation, $E(u_{i1}) \approx \text{logit}^{-1}\left(\frac{\alpha_u}{\sqrt{1 + C^2 \sigma_u^2}}\right)$ and

$$E(v_{i1}) \approx \text{logit}^{-1}\left(\frac{\alpha_v}{\sqrt{1 + C^2 \sigma_v^2}}\right), \text{ where } C = \frac{16\sqrt{3}}{15\pi} \text{ (Zeger et al., 1988).}$$

We selected proper but diffuse prior distributions for the hyper-parameters because non-informative prior distributions can lead to inaccurate posterior estimates (Natarajan and McCulloch, 1998). For fixed effects, vague priors were assigned. In particular, α_n and α_a both follow $N(0, 2.5^2)$, so under the simplest situation (a fixed effects model), a 95% prior probability interval for any of the probabilities π_{in} , π_{ia} , π_{ic} would range from about 0.001 to 0.91; and α_s , α_b , α_u , α_v all follow $N(0, 2^2)$, which implies a 95% interval for the probabilities s_{i1} , b_{i1} , u_{i1} , v_{i1} ranging from about 0.01 to 0.98. The hyper-priors for the precision parameters σ_s^{-2} , σ_b^{-2} , σ_u^{-2} , σ_v^{-2} were assumed to be *Gamma*(2, 2), which implies a 95% interval (0.36, 8.36) for the variance parameters, allowing moderate heterogeneity for the response probabilities. The prior for the precision matrix Σ_{ps}^{-1} was Wishart, *i.e.*, $W(I, 3)$, where I is the identity matrix. In the reduced model with one of σ_n^2 , σ_a^2 set to 0, the prior of the other precision parameter was still assumed to be *Gamma*(2, 2), which gave moderate heterogeneity for latent principal strata probabilities.

3.3 Model selection and results

We adopted the forward selection procedure introduced in Section 2.5 to search for the best-fitting model. The models specified in Section 2.4 have 6 potential random effects in total: δ_{in} , δ_{ia} , δ_{is} , δ_{ib} , δ_{iu} and δ_{iv} . In each step of the forward-selection procedure, we added the random-effect component that provided the largest improvement in DIC. Table 2 presents DIC, DIC improvement, and the effective number of parameters (p_D) for each model fit to the 10 studies. Starting with the simplest model with no random effects (Model I), the largest improvement was obtained by allowing the probability of being an always-taker to be a random effect, *i.e.*, adding δ_{ia} (Model II_f). In the second step, adding a random effect for the probability of being a never-taker δ_{in} (Model III_e) decreased DIC the most. This revealed an important characteristic of this meta-analysis: the studies might vary considerably in their recruitment criteria, study procedures, beliefs of the local PIs, *etc.*, resulting in different properties of the latent compliance classes. The next forward step produced a meaningful improvement by including a random effect for the cesarean section rate of a never-taker, δ_{is} (Model IV_a). The resulting improvement was modest compared to adding random effects δ_{ia}

and δ_{in} . It is difficult to say what constitutes an important difference in DIC; we follow Lunn et al. (2012) in considering that a reduction of less than 5 is not a substantial improvement. In the last step, DIC was reduced only by 3.6 compared to Model IVa, so the final model included random effects δ_{ia} , δ_{in} , and δ_{is} .

Figure 2 is a forest plot of θ_i^{CACE} for each study individually, using the Bayesian method with the foregoing priors. The first overall estimate of θ^{CACE} is from the final model (Model IVa). The second estimate is the REML (restricted maximum likelihood) estimate calculated using the individual posterior means and standard errors of the θ_i^{CACE} . We call this method the “two-step” approach because it first analyzes each study separately, then combines the individual estimates to give a pooled estimate of θ^{CACE} . The estimated θ_i^{CACE} varied from negative to positive in individual studies, but the 95% credible intervals all cover zero. Among overall estimates, the two-step approach gave a wider interval than our proposed method, suggesting it may not be efficient because it required the whole set of parameters to be estimated for each study, leading to a larger total number of parameters.

The ITT fixed-effect meta-analysis gave an estimated risk difference of 0.011 with 95% confidence interval $(-0.005, 0.027)$; the overall θ^{CACE} from our final model, 0.022 $(-0.003, 0.048)$, was not significant either, though the point estimate was about twice of that from the ITT analysis in the same direction. Thus we conclude that the potential dilution of the ITT estimated treatment effect notwithstanding, epidural analgesia in labor does not affect the risk of cesarean section in the view of causal interpretation.

4. Simulation Studies

To evaluate the performance of our modeling approach and to study the impact of misspecification of random effects, we performed four sets of simulations. For ease of presentation and interpretation, we generated data with, at most, random effects only for the latent compliance class probabilities and the response probabilities of treated compliers, $(\delta_{in}, \delta_{iu})$, then fit models with up to three random effects $(\delta_{in}, \delta_{iu}, \delta_{iv})$. Specifically, data were simulated from the following four models: 1) no random effects; 2) random effect for latent compliance class probabilities δ_{in} ; 3) random effect for the response probabilities of compliers in the treatment group δ_{iu} ; and 4) random effects for both δ_{in} and δ_{iu} . The simulations are realistic scenarios that researchers are likely to encounter, such as those in the case studies.

We simulated 2000 meta-analysis datasets in each setting. For each dataset, 20 studies were simulated, each with 350 observations, roughly matching the sample sizes in our case study. In each simulated study, the allocation ratio was 1:1 ($\lambda_i = 0.5$). True values were set to $(\alpha_p, \alpha_a, \alpha_s, \alpha_b, \alpha_u, \alpha_v) = (-0.4, -0.6, 0.5, -0.5, -0.5, 0.5)$, giving true values $\pi_{in} = 0.302$, $\pi_{ia} = 0.247$ and $\theta_i^{CACE} = -0.383$ in the absence of random effects. In the presence of random effects, the variances of δ_{in}, δ_{iu} were set to 0.5^2 so the true θ^{CACE} was approximately -0.364 when δ_{iu} was present.

For each simulated dataset, we fit models with: 1) no random effects; 2) one random effect, for δ_{in} , δ_{iu} , or δ_{iv} ; 3) two random effects, for $[\delta_{in}, \delta_{iu}]$, $[\delta_{in}, \delta_{iv}]$, or $[\delta_{iu}, \delta_{iv}]$; and 4) three random effects, for $[\delta_{in}, \delta_{iu}, \delta_{iv}]$. Each analysis used JAGS with forward selection using DIC. Table 3 summarizes the estimated probability of selecting each candidate model as the “best” model in each simulation setting. DIC identified the true random effects model with probability over 0.95, i.e., our procedure usually selected the correct random effects and thus usually accounted properly for uncertainty in the estimates.

Table 4 shows the estimated mean, bias, 95% credible interval length, and coverage probability for θ^{CACE} under each model. We present results only for θ^{CACE} because of space limitations. In general, the posterior standard deviation becomes larger as more random effects are included. Over-fitting (including a random effect when it is absent) tends to give longer 95% credible intervals. As for the coverage probabilities, 1) under the correct model or when over-fitting occurs, the coverage probabilities are close to or greater than the nominal 0.95; and 2) when under-fitting occurs (omitting a random effect when it is present), failure to include the random effect for the study-specific probability of latent compliance groups (δ_{in}) does not substantially affect the coverage probability, but failure to include the random effect for the response rate of a treated complier (δ_{iu}) reduces coverage for θ^{CACE} notably.

5. Discussion

To estimate compier average causal effects in meta-analysis of RCTs with noncompliance, we proposed a Bayesian hierarchical model that accounts for between-study heterogeneity and applied it to a meta-analysis of epidural analgesia trials. We also performed simulation studies to evaluate the performance of our approach and the impact of misspecification of random effects. To the best of our knowledge, this is the first meta-analysis of RCTs estimating CACE accounting for noncompliance.

In the case study we compared the estimate of θ^{CACE} from the “two-step approach” with the one from our proposed final model. The two-step approach can be a valid alternative to our proposed random-effect model. However, this approach requires the whole set of parameters to be estimated independently for each study. Thus, the total number of effective parameters may be larger, making it potentially not as efficient as the one-step approach. Also, after computing θ_i^{CACE} and $SE(\theta_i^{CACE})$ from posterior samples (which depends on large-sample theory), various methods could be used to account for the heterogeneity besides the REML estimator we used, *e.g.*, the DerSimonian-Laird, Hedges, or Hunter-Schmidt estimators (Viechtbauer, 2005), which may complicate the estimation.

The control procedure in this case study is defined as “nonepidural or no analgesia in labor”, so the control treatment varies between studies, making it possible that the corresponding population of each latent class is not identical across studies. Because θ_i^{CACE} in Figure 2 was estimated from separate analysis only using data from study i , the study-specific θ_i^{CACE} could be interpreted slightly different depending on what the particular control procedure for that study was. However, the different sub-populations of compliers in individual studies are

considered subsets of the general “compliers” population defined for the meta-analysis comparing epidural analgesia in labor with nonepidural or no analgesia. The estimand θ^{CACE} is thus still sufficiently well-defined and its interpretation is not affected.

Note that the CACE is defined under the principal stratification framework assuming SUTVA, random assignment, the exclusion restriction, $E[T_{ij}^1 - T_{ij}^0] \neq 0$, and monotonicity. In our case study, the assumptions are plausible as discussed in Section 3.1. In other cases, some assumptions may be violated; accounting for these violations will increase the complexity of defining and estimating causal effects. Moreover, the CACE analysis typically assumes that compliance status is reported without error, but this might be invalid for self-reported compliance. Therefore, further research is needed on the consequences of relaxing key assumptions and incorporating noncompliance measurement errors. Some extensions have been developed for a single randomized trial, *e.g.*, estimating causal effects in the presence of interference (Liu and Hudgens, 2014), noncompliance measured with error (Imai and Yamamoto, 2010), and incorporating baseline covariates (Roy et al., 2007). Extensions to meta-analysis of RCTs incorporating those issues await further development.

Also, as a reviewer noted, in our epidural analgesia case study, compliance may be strongly determined by external factors such as the severity of pain. Pain severity is another post-randomization variable that is determined after randomization and may relate to both the actual taken treatment and the outcome. Randomization should have made the pain severity balanced between treatment arms. However, pain may influence the behavior of compliance differently in the treatment and control arms: in the epidural analgesia group, severe labor pain may reduce the rate of noncompliance, while in the control group, severe labor pain possibly increases the rate of noncompliance. Other possible post-randomization variables could include the length of labor. However, it is hard to measure pain severity and for our example meta-analysis, we do not have data on that or other post-randomization variables. Methods for handling such post-randomization variables deserve further investigation.

Acknowledgements

The authors thank the Co-Editor, the Associate Editor, and two anonymous reviewers for many helpful comments. This research was supported in part by NIH NLM R21012197 (HC, JZ), NLM R21012744 (HC, JH), AHRQ R03HS024743 (HC), NIDDK U01 DK106786 (HC).

References

- Agresti A (2003). Categorical data analysis, volume 482 John Wiley & Sons.
- Angrist JD, Imbens GW, and Rubin DB (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91, 444–455.
- Baker SG (2011). Estimation and inference for the causal effect of receiving treatment on a multinomial outcome: an alternative approach. *Biometrics* 67, 319–323. [PubMed: 20560933]
- Baker SG and Kramer BS (2005). Simple maximum likelihood estimates of efficacy in randomized trials and before-and-after studies, with implications for meta-analysis. *Statistical Methods in Medical Research* 14, 349–67. [PubMed: 16178137]
- Balke A and Pearl J (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association* 92, 1171–1176.

- Bannister-Tyrrell M, Miladinovic B, Roberts CL, and Ford JB (2015). Adjustment for compliance behavior in trials of epidural analgesia in labor using instrumental variable meta-analysis. *Journal of Clinical Epidemiology* 68, 525–533. [PubMed: 25592169]
- Celeux G, Forbes F, Robert CP, Titterton DM, et al. (2006). Deviance information criteria for missing data models. *Bayesian analysis* 1, 651–673.
- Cheng J (2009). Estimation and inference for the causal effect of receiving treatment on a multinomial outcome. *Biometrics* 65, 96–103. [PubMed: 18373714]
- Chu H and Cole SR (2006). Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology* 59, 1331–1332. [PubMed: 17098577]
- Clopper CJ and Pearson ES (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 26, 404–413.
- Dai N and Jones GL (2017). Multivariate initial sequence estimators in markov chain monte carlo. *Journal of Multivariate Analysis* 159, 184–199.
- Frangakis CE and Rubin DB (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* 86, 365–79.
- Frangakis CE and Rubin DB (2002). Principal stratification in causal inference. *Biometrics* 58, 21–29. [PubMed: 11890317]
- Freedman LS (1990). The effect of partial noncompliance on the power of a clinical trial. *Controlled Clinical Trials* 11, 157–168. [PubMed: 2163811]
- Gelfand AE and Smith AFM (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85, 398–409.
- Gelman A and Rubin DB (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 7, 457–472.
- Gelman A, Stern HS, Carlin JB, Dunson DB, Vehtari A, and Rubin DB (2013). *Bayesian data analysis*. Chapman and Hall/CRC.
- Glasziou P (1992). Meta-analysis adjusting for compliance: the example of screening for breast cancer. *Journal of Clinical Epidemiology* 45, 1251–1256. [PubMed: 1432006]
- Higgins J and Thompson SG (2002). Quantifying heterogeneity in a metaanalysis. *Statistics in Medicine* 21, 1539–1558. [PubMed: 12111919]
- Holland PW (1986). Statistics and causal inference. *Journal of the American Statistical Association* 81, 945–960.
- Hudgens MG and Halloran ME (2008). Toward causal inference with interference. *Journal of the American Statistical Association* 103, 832–842. [PubMed: 19081744]
- Imai K and Yamamoto T (2010). Causal inference with differential measurement error: Nonparametric identification and sensitivity analysis. *American Journal of Political Science* 54, 543–560.
- Imbens GW and Rubin DB (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics* pages 305–327.
- Jackson D, Riley R, and White IR (2011). Multivariate meta-analysis: potential and promise. *Statistics in Medicine* 30, 2481–98. [PubMed: 21268052]
- Little RJ, Long Q, and Lin X (2009). A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance. *Biometrics* 65, 640–649. [PubMed: 18510650]
- Little RJ and Yau LHY (1998). Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using rubin's causal model. *Psychological Methods* 3, 147–159.
- Liu L and Hudgens MG (2014). Large sample randomization inference of causal effects in the presence of interference. *Journal of the American Statistical Association* 109, 288–301. [PubMed: 24659836]
- Lunn D, Jackson C, Best N, Spiegelhalter D, and Thomas A (2012). *The BUGS book: A practical introduction to Bayesian analysis*. Chapman and Hall/CRC.
- Natarajan R and McCulloch CE (1998). Gibbs sampling with diffuse proper priors: A valid approach to data-driven inference? *Journal of Computational and Graphical Statistics* 7, 267–277.

- Newcombe R (1988). Explanatory and pragmatic estimates of the treatment effect when deviations from allocated treatment occur. *Statistics in Medicine* 7, 1179–1186. [PubMed: 3201044]
- Nikkola EM, Ekblad UU, Kero PO, Alihanka JJ, and Salonen MA (1997). Intravenous fentanyl PCA during labour. *Canadian Journal of Anaesthesia* 44, 1248–1255. [PubMed: 9429041]
- Roy J, Hogan JW, and Marcus BH (2007). Principal stratification with predictors of compliance for randomized trials with 2 active treatments. *Biostatistics* 9, 277–289. [PubMed: 17681993]
- Rubin DB (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association* 75, 591–593.
- Spiegelhalter DJ, Best NG, Carlin BP, and Van Der Linde A (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64, 583–639.
- Viechtbauer W (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics* 30, 261–293.
- Yau LH and Little RJ (2001). Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association* 96, 1232–1244.
- Zeger S, Liang K, and Albert P (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 44, 1049–1060. [PubMed: 3233245]

Figure 1.

Forest plot of study-specific noncompliance rates in studies of epidural analgesia. The center of each square/circle and the horizontal lines represent the corresponding probability with 95% exact confidence interval. The diamonds indicate overall estimates of noncompliance rates using the bivariate generalized linear mixed effects model (BGLMM). This figure appears in color in the electronic version of this article.

Figure 2.

Forest plot of CACE of epidural analgesia in labor on caesarean section. The center of each square and the horizontal lines represent the posterior medians and 95% equal tail credible intervals of θ_i^{CACE} for each individual study based on separate analyses. A diamond

indicates the pooled estimate of θ^{CACE} and its 95% credible interval or confidence interval. The first overall estimate comes from the selected final model, Model IVa; the second one is from the REML estimator under a random-effect model using the two-step approach; the third one is the overall risk difference (RD) from the fixed-effect ITT analysis. This figure appears in color in the electronic version of this article.

Table 1

Observed groups, latent compliance classes and outcome probabilities of trial i

R_{ij}	T_{ij}	C_{ij}	$Y_{ij} = o \in \{1, \dots, O\}$	Count
0	0	0 (never-taker) or 1 (complier)	$M(n_{i00}, q_{io} = \frac{\pi_{ic} v_{io} + \pi_{in} s_{io}}{1 - \pi_{ia}})$	n_{00o}
0	1	2 (always-taker)	$M(n_{01}, b_{io})$	n_{01o}
1	0	0 (never-taker)	$M(n_{10}, s_{io})$	n_{10o}
1	1	1 (complier) or 2 (always-taker)	$M(n_{i11}, p_{io} = \frac{\pi_{ic} u_{io} + \pi_{ia} b_{io}}{1 - \pi_{in}})$	n_{11o}

Defiers are ruled out by the monotonicity assumption.

Table 2

Selection of random effects using a forward selection procedure on RCTs of epidural analgesia in labor

Random effects models	DIC	DIC improvement	p_D
Model I (None)	917.4	N/A	6.0
Model IIa (δ_{is})	896.8	20.6	21.3
Model IIb (δ_{ib})	918.3	-0.9	16.1
Model IIc (δ_{iu})	915.0	2.4	12.4
Model IId (δ_{iv})	907.4	10.0	14.1
Model IIe (δ_{in})	577.1	340.3	13.6
Model IIIf (δ_{ia})	537.2	380.2	15.7
Model IIIa (δ_{ia} , δ_{is})	514.9	22.3	29.3
Model IIIb (δ_{ia} , δ_{ib})	538.6	-1.4	19.5
Model IIIc (δ_{ia} , δ_{iu})	531.9	5.3	21.1
Model IIId (δ_{ia} , δ_{iv})	526.3	10.9	22.9
Model IIIe (δ_{ia} , δ_{in})	265.7	271.5	21.4
Model IVa (δ_{ia} , δ_{in} , δ_{is})	246.5	25.0	27.3
Model IVb (δ_{ia} , δ_{in} , δ_{ib})	266.5	5.0	24.9
Model IVc (δ_{ia} , δ_{in} , δ_{iu})	262.1	9.4	28.2
Model IVd (δ_{ia} , δ_{in} , δ_{iv})	267.1	4.4	29.3
Model IVe (δ_{ia} , δ_{in} , ρ)	265.7	5.8	22.1
Model Va (δ_{ia} , δ_{in} , δ_{is} , ρ)	246.2	0.3	27.6
Model Vb (δ_{ia} , δ_{in} , δ_{is} , δ_{ib})	247.2	-0.7	30.7
Model Vc (δ_{ia} , δ_{in} , δ_{is} , δ_{iu})	242.9	3.6	34.0
Model Vd (δ_{ia} , δ_{in} , δ_{is} , δ_{iv})	253.0	-6.5	34.4

Bold-face cells are models selected using the forward selection procedure. DIC improvement is the reduction in DIC compared to the previous step's lowest DIC. The study-specific component corresponding to each random effect: δ_{is} never-taker response, δ_{ib} always-taker response, δ_{iu} treated complier response, δ_{iv} control complier response, δ_{in} never-taker probability, δ_{ia} always-taker probability; ρ is the correlation between δ_{ia} and δ_{in} .

Table 3

Estimated probability of selecting each candidate model using DIC^{*}, based on 2000 simulated datasets

True Random Effects Model	Selected Random Effects Model							
	None	δ_m	δ_{iu}	δ_v	δ_m, δ_{iu}	δ_m, δ_v	δ_{iu}, δ_v	$\delta_m, \delta_{iu}, \delta_v$
None	95.05	1.60	1.85	1.45	0	0.05	0	0
δ_m	0	96.25	0	0	1.85	1.90	0	0
δ_{iu}	0	0	96.65	0	1.25	0	2.00	0.10
δ_m, δ_{iu}	0	0	0	0	98.25	0	0	1.75

Bold-face cells give the estimated probability of identifying the correct model. The numbers have been multiplied by 100 for presentation.

* DIC: deviance information criterion.

Table 4Performance of estimates and credible intervals for θ^{CACE} for each model, based on 2000 simulated datasets

True Random Effects Model		Selected Random Effects Model							
		None	δ_{in}	δ_{iu}	δ_{iv}	δ_{in}, δ_{iu}	δ_{in}, δ_{iv}	δ_{iu}, δ_{iv}	$\delta_{in}, \delta_{iu}, \delta_{iv}$
None	Mean	-0.383	-0.383	-0.373	-0.374	-0.372	-0.374	-0.363	-0.363
	Bias	0.000	0.000	0.010	0.009	0.011	0.009	0.020	0.020
	95% CIL *	0.100	0.099	0.171	0.171	0.170	0.172	0.220	0.221
	95% CICp **	0.947	0.948	0.999	0.999	0.999	0.999	1.000	1.000
δ_{in}	Mean	-0.383	-0.382	-0.372	-0.374	-0.372	-0.373	-0.364	-0.363
	Bias	0.000	0.001	0.011	0.009	0.011	0.010	0.019	0.020
	95% CIL	0.101	0.098	0.172	0.173	0.170	0.175	0.222	0.223
	95% CICp	0.946	0.947	0.998	0.999	0.998	0.999	1.000	1.000
δ_{iu}	Mean	-0.365	-0.364	-0.356	-0.355	-0.355	-0.355	-0.346	-0.346
	Bias	-0.001	0.000	0.008	0.009	0.009	0.009	0.018	0.018
	95% CIL	0.100	0.099	0.204	0.171	0.204	0.172	0.247	0.248
	95% CICp	0.755	0.759	0.986	0.944	0.986	0.946	0.995	0.996
δ_{in}, δ_{iu}	Mean	-0.365	-0.365	-0.357	-0.356	-0.356	-0.355	-0.348	-0.347
	Bias	-0.001	-0.001	0.008	0.008	0.008	0.009	0.016	0.017
	95% CIL	0.101	0.099	0.205	0.173	0.204	0.175	0.249	0.250
	95% CICp	0.763	0.753	0.986	0.951	0.987	0.951	0.997	0.996

Bold-face cells are the correct model.

* 95% CIL: 95% equal-tail credible interval length.

** 95% CICp: 95% credible interval coverage probability.